Card-Shuffling via Convolutions of Projections on Combinatorial Hopf Algebras

C. Y. Amy Pang

Abstract

Recently, Diaconis, Ram and I created Markov chains out of the coproduct-then-product operator on combinatorial Hopf algebras. These chains model the breaking and recombining of combinatorial objects. Our motivating example was the riffle-shuffling of a deck of cards, for which this Hopf algebra connection allowed explicit computation of all the eigenfunctions. The present note replaces in this construction the coproduct-then-product map with convolutions of projections to the graded subspaces, effectively allowing us to dictate the distribution of sizes of the pieces in the breaking step of the previous chains. An important example is removing one "vertex" and reattaching it, in analogy with top-to-random shuffling. This larger family of Markov chains all admit analysis by Hopf-algebraic techniques. There are simple combinatorial expressions for their stationary distributions and for their eigenvalues and multiplicities and, in some cases, the eigenfunctions are also calculable.

Récemment, avec Diaconis et Ram, nous avons construit des chaines de Markov à partir de l'opérateur "coproduit-puis-produit" défini sur un algèbre de Hopf combinatoire. Ces chaines modélisent la déconstruction et la construction d'objets combinatoires. La motivation était le "mélange à l'américaine", une méthode populaire pour mélanger un jeu de cartes, pour lequel les liens avec les algèbres de Hopf combinatoires nous a permis de calculer explicitement toutes les fonctions propres. Ici, on généralise cette construction en remplaçant l'opérateur "coproduit-puis-produit" par les convolutions de projections sur les composantes graduées de l'algèbre. Ceci nous permet de stipuler les tailles des pièces dans la décomposition des objets combinatoires. Un exemple important est la suppression et l'insertion d'un "sommet", par analogie avec la bibliothèque de Tsetlin. On constate que toutes ces chaines peuvent être analysées par des techniques provenant de la théorie des algèbres de Hopf combinatoires. On prouve des expressions combinatoires simples pour les distributions stationnaires ainsi que pour les valeurs propres et leurs multiplicités. Dans certains cas, il est possible de calculer les fonctions propres associées.

1   Introduction

Background: Markov chains from Hopf-powers

Possibly the most popular model of card-shuffling is the Gilbert-Shannon-Reeds (GSR) *riffle-shuffle*: cut the deck into two piles according to the (symmetric) binomial distribution, then drop one-by-

one the bottom card from either pile, chosen with probability proportional to the current pile size. This second step is equivalent to all interleavings of the two piles (counted with multiplicity) being equally likely.

Amongst the plethora of results concerning this shuffle, the most notable must be the work of (Bayer and Diaconis 1992), who determined that $\frac{3}{2} \log n$ shuffles are necessary and sufficient to randomise a deck of n cards. Central to their argument is the generalisation of the GSR model to *a-handed shuffles*: cut the deck into a piles according to a (symmetric) multinomial distribution, then drop the bottom cards from a pile chosen with probability proportional to pile size as before. Performing the GSR shuffle t times is then the same as a 2^t-handed shuffle, so analysing long-term behaviour of 2-handed riffle-shuffles is equivalent to letting the number of hands tend to infinity.

(Diaconis, Pang, and Ram 2014) observed that the transition probabilities of the riffle-shuffle are, up to scaling, the coefficients of the coproduct-then-product map $m\Delta$ on the shuffle algebra. Furthermore, the coefficients of the *ath Hopf-power map* $m^{[a]}\Delta^{[a]}$ on the shuffle algebra give the transition probabilities of the a-handed riffle-shuffle. We then defined Markov chains on the bases of other combinatorial Hopf algebras by setting their transition probabilities to be such Hopf-power coefficients, with a little modificiation (via the Doob transform). These *Hopf-power Markov chains* model the breaking then recombining of the combinatorial objects indexing the bases of the algebras. The thesis of (Pang 2014) greatly extends the Hopf-power Markov chain framework; this encompasses a restriction-then-induction chain on representations of the symmetric groups, and a tree-pruning model - see Examples 3.5 and 5.3 below.

Probabilistic conclusions from Hopf-algebraic techniques

The benefit of this viewpoint is two-fold. Firstly, as recorded by (Pang 2014, Th. 4.7.1), maps between Hopf algebras which "respect the bases" induce projections of the related chains. (Pang 2013) applied this to a map from the shuffle algebra to the algebra of quasisymmetric functions, to conclude that the positions of descents under riffle-shuffling of a deck of distinct cards is a Markov statistic. (A *descent* occurs where a card has greater value than the card immediately below it.) This means that the probability of a shuffle producing descents in prescribed positions depends only on the positions of descents before the shuffle, and not on the exact deck order. (This fact also follows from the descent set being a "shuffle-compatible statistic", which (Gessel 2010) attributes to Stanley.) (Pang 2014, Sec. 4.7) constructs many Markov statistics for inverse riffle-shuffling out of commutation quotients of the free associative algebra.

The second way in which the Hopf formulation aids in studying these Markov chains is that, in many cases, there are algorithms to compute a basis of eigenvectors for the Hopf-power maps and hence the transition matrices. This gives interesting information about the long term behaviour of the chain. As an example, (Pang 2014, Prop. 6.1.3 and Prop. 6.1.5) state that, if a deck of n distinct cards was originally in ascending order, then, after t iterations of the a-handed shuffle, the expected number of descents is $(1 - a^{-t})\frac{n-1}{2}$, and the expected number of peaks is $(1 - a^{-2t})\frac{n-2}{3}$. (A *peak* is a triple of adjacent cards with the middle one having greatest value.) Although the algorithms do not provide all eigenvectors of all Hopf-power Markov chains, their stationary distributions are always computable.

A new extension: Markov chains from convolutions of projections

As (Pang 2014, Th. 4.4.1) shows, the breaking step of a Hopf-power Markov chain always involves a symmetric multinomial distribution. However, it is sometimes more natural to consider other distributions. For example, the restriction-then-induction chain mentioned above is difficult to

express in terms of partitions, because the restriction of a symmetric group representation to a multinomially-chosen Young subgroup involves Littlewood-Richardson coefficients. Simpler is the Markov chain which removes a random corner box and re-inserts it in a random position.

The discovery in this extended abstract is that, by replacing $m^{[a]}\Delta^{[a]}$ in the definition of a Hopf-power Markov chain with a *non-negative convolution of projections* (Definition 2.3), one can change the piece sizes in the breaking step to have any desired distribution. For example, the "remove and re-insert a box" chain above comes from the map $m(\mathrm{Proj}_1 \otimes \iota)\Delta$, using exactly the same Doob transform. (ι is the identity map and Proj_1 is projection to the subspace of degree 1.) On the shuffle algebra, $m(\mathrm{Proj}_1 \otimes \iota)\Delta$ defines the much studied *top-to-random shuffle*: take the top card off the deck, then re-insert it at a uniformly chosen position. The other non-negative convolutions of projections recover the shuffles of (Diaconis, Fill, and Pitman 1992), where the deck is cut in some specified distribution, and then the cards dropped one by one from the bottom of piles chosen with probability proportional to pile size. Their Corollaries 5.1 and 5.2 are a formula for the composition of such shuffles, and an upper bound for the mixing time. The inverses of such shuffles are examples of *pop shuffles* of (Bidigare, Hanlon, and Rockmore 1999), so the eigenvalues and multiplicities can be calculated with their hyperplane walk theory.

This new class of Markov chains admit analysis by the same techniques as for Hopf-power Markov chains. Maps between Hopf algebras "respecting the bases" again induce projections of their associated chains (Theorem 4.1 below). Consequently, the descent set is a Markov statistic under all these shuffling schemes. Existing literature on convolutions of projections provides the eigenvalues and multiplicities of these transition matrices. In some cases, there are eigenbasis algorithms resembling those for the Hopf-power chains. Once again, the stationary distributions of all these chains are accessible - they are precisely the same as those for the Hopf-power chains.

One notable shuffle outside this framework is *random-to-random*: uniformly choose a card to remove from the deck, and re-insert it in a uniform position. Its defining linear map is an interesting operator on other combinatorial Hopf algebras, and it would be great to find a probability interpretation.

This extended abstract is organised as follows: Section 2 gives the conditions on the two main characters in this story, the state space basis of a combinatorial Hopf algebra and the non-negative convolution of projections map. Section 3 explains how to construct the Markov chain via the Doob transform. Section 4 states the three main theorems: how morphisms of Hopf algebras lead to Markov statistics; the eigenvalues of the transition matrices and their multiplicities; and the common stationary distributions. Section 5 shows one scenario where explicit eigenbasis formulae are available, and gives probability applications both for shuffling and for a chain on trees.

2 Combinatorial Hopf Algebras and the Convolution Product

The starting point of our Markov chain construction is a combinatorial Hopf algebra, which encodes the breaking and combining rules for our family of combinatorial objects. An instructive example is the shuffle algebra, whose associated Markov chains describe various models of shuffling.

Example 2.1. The shuffle algebra has a basis $\mathcal{B}$ of words, which we will think of as decks of cards. For example, the word *accb* will denote the deck with card a on top, followed by two copies of card c, and card b on the bottom. This algebra is graded by the lengths of the words, or the number of

cards in the deck. The product of two words is the sum of all their interleavings (with multiplicity), and the coproduct of a word is the sum of all its deconcatenations. For example,

$$m(ac \otimes cb) = 2accb + acbc + cacb + cabc + cbac;$$

$$\Delta(accb) = 1 \otimes accb + a \otimes ccb + ac \otimes cb + acc \otimes b + accb \otimes 1.$$

The exposition of (Grinberg and Reiner 2014) gives background on combinatorial Hopf algebras, and the opening of (Foissy 2012) contains an extensive list of examples with references. Like many recent treatments, these focus on generalisations of the symmetric functions, which, though extremely important, are not so integral to the present Markov chain application. The thesis of (Klausner 2011, Sec. 4) is closer to the viewpoint herein.

There is no rigorous definition of a combinatorial Hopf algebra. The intuition is that such an algebra $\mathcal{H}$ should have a basis $\mathcal{B}$ indexed by a family of combinatorial objects, graded by their sizes. (Assume throughout that the ground field is $\mathbb{Q}$ or $\mathbb{R}$, to facilitate the probability applications.) Write $\mathcal{H}_n$ for the subspace of $\mathcal{H}$ of degree n, so $\mathcal{H} = \bigoplus \mathcal{H}_n$. Since the empty object is the only object of size 0, the vector space $\mathcal{H}$ is *connected, i.e.* $\dim \mathcal{H}_0 = 1$. Now equip $\mathcal{H}$ with a multiplication map $m : \mathcal{H}_i \otimes \mathcal{H}_j \to \mathcal{H}_{i+j}$: for $w, z \in \mathcal{B}$, set $m(w \otimes z)$ to be the sum (possibly weighted) of all possible results from "combining" w and z. Similarly, the coproduct map $\Delta : \mathcal{H}_n \to \bigoplus_{i=0}^{n} \mathcal{H}_i \otimes \mathcal{H}_{n-i}$ takes $x \in \mathcal{B}$ to the sum (possibly weighted) of $w \otimes z$ over all pairs (w, z) obtainable by "breaking" x. There are various axioms that these operations must satisfy.

The above combinatorial interpretation of the product and coproduct motivates that the structure constants be non-negative (conditions i and ii below). This is crucial for the transition probabilities in Definition 3.1 to be non-negative. As we will discover in Section 3, another important hypothesis for the construction of the associated Markov chains is that $\Delta(x) \neq 1 \otimes x + x \otimes 1$ for any $x \in \mathcal{B}$ of degree greater than 1. Intuitively, it mandates that "every combinatorial object of size greater than 1 can be broken into strictly smaller pieces". Combining these criteria leads to the following definition:

Definition 2.2. (Pang 2014, Def. 4.3.3) A basis $\mathcal{B} = \amalg_n \mathcal{B}_n$ of a graded connected Hopf algebra $\mathcal{H} = \bigoplus_n \mathcal{H}_n$ is a *state space basis* if:

i) for all $w, z \in \mathcal{B}$, the expansion of $m(w \otimes z)$ in the $\mathcal{B}$ basis has all coefficients non-negative;

ii) for all $x \in \mathcal{B}$, the expansion of $\Delta(x)$ in the $\mathcal{B} \otimes \mathcal{B}$ basis has all coefficients non-negative;

iii) for $n > 1$, the basis $\mathcal{B}_n$ contains no *primitive* elements. That is, $\Delta(x) \neq 1 \otimes x + x \otimes 1$ for all $x \in \mathcal{B}_n$ with $n > 1$.

Note that $\mathcal{H}$ may contain primitive elements of high degree, so long as they are not in the basis $\mathcal{B}$.

Having fixed a combinatorial Hopf algebra and a degree n, the next step is to choose the distribution of sizes of pieces in the breaking step of the Markov chain. As Step 1 of the interpretation in Section 3 will show, these distributions are in bijection, up to scaling, with the class of non-negative convolutions of projections, as defined below. For example, the operator $\mathrm{Proj}_{d_1} * \mathrm{Proj}_{d_2}$ will constrain the first piece to be of size exactly d_1, and the second piece to be of size exactly d_2.

Given maps $f, g : \mathcal{H} \to \mathcal{H}$, their *convolution product* $f * g : \mathcal{H} \to \mathcal{H}$ is the composition

$$f * g := m(f \otimes g)\Delta.$$

Since the product m on $\mathcal{H}$ is associative and the coproduct is coassociative, the convolution product is associative, and it is useful to view

$$f_1 * \cdots * f_a \quad \text{as} \quad m^{[a]}(f_1 \otimes \cdots \otimes f_a)\Delta^{[a]}.$$

Here $m^{[a]} : \mathcal{H}^{\otimes a} \to \mathcal{H}$ and $\Delta^{[a]} : \mathcal{H} \to \mathcal{H}^{\otimes a}$ are the a-fold product and a-fold coproduct respectively, describing the combining of a objects and breaking one object into a pieces (some of which may be trivial). So $m^{[2]} = m$ and $\Delta^{[2]} = \Delta$. (A precise definition, by induction, is as follows: $m^{[1]} := \iota$, $m^{[a]} := m(m^{[a-1]} \otimes \iota)$; $\Delta^{[1]} := \iota$, $\Delta^{[a]} := (\iota \otimes \cdots \otimes \iota \otimes \Delta)\Delta^{[a-1]}$.)

The present Markov chain application uses only the case $f_i = \text{Proj}_{d_i}$, the projection to the subspace of degree d_i. In other words, $\text{Proj}_d : \mathcal{H} \to \mathcal{H}_d$ is the linear map satisfying $\text{Proj}_d(x) = x$ for $x \in \mathcal{H}_d$, and $\text{Proj}_d(\mathcal{H}_i) = 0$ if $i \neq d$. It will be convenient to allow the case $d = 0$.

(Patras 1994) studied linear combinations of these convolutions of projections. He called them *descent operators*, since his Theorem II.7 shows that, on a commutative or cocommutative Hopf algebra, these operators form an algebra under composition isomorphic to the descent algebra of the symmetric group. This connection is central to the eigenbasis algorithms for the present Markov chains, which come from the following subset of these operators:

Definition 2.3. Let $\mathcal{H}$ be a graded connected Hopf algebra, and fix an integer $a \geq 2$. A map $\mathbf{T} : \mathcal{H} \to \mathcal{H}$ of the form

$$\mathbf{T} := \sum_{d_1, \ldots, d_a} \alpha_{(d_1, \ldots, d_a)} \text{Proj}_{d_1} * \cdots * \text{Proj}_{d_a}$$

is a *non-negative convolution of projections on $\mathcal{H}_n$* if

  i) for all *weak-compositions* $D := (d_1, \ldots, d_a)$ of n (that is, $d_i \geq 0$ and $\sum d_i = n$), the coefficients α_D are non-negative;

  ii) for some weak-composition D of n where each $d_i \neq n$, the coefficient α_D is positive.

The second axiom ensures that $\mathbf{T} : \mathcal{H}_n \to \mathcal{H}_n$ is not merely multiplication by a constant. Note that the map $\mathbf{T}$ does not uniquely determine the coefficients α_D, because of the possibility of parts of size 0. Different choices of α_D lead to different interpretations of the same Markov chain. One final remark on this definition: it is fine for infinitely many α_D to be non-zero, as the image $\mathbf{T}(x)$ of any particular $x \in \mathcal{H}$ is always a finite sum. This is because there are only finitely many weak-compositions of a given integer into exactly a parts.

One key example of a non-negative convolution of projections is the ath Hopf-power map $m^{[a]}\Delta^{[a]}$. This is the ath convolution power of the identity map, so it corresponds to setting $\alpha_D \equiv 1$ for all weak-compositions D. As the three-step interpretation of Section 3 will explain, the associated Markov chains have a symmetric multinomial breaking step. This is the case previously studied in (Diaconis, Pang, and Ram 2014) and in (Pang 2014). Another important specialisation comes from $\alpha_{(1, n-1)} = 1$, $\alpha_D = 0$ if $D \neq (1, n-1)$ for any n, so $\mathbf{T} = \text{Proj}_1 * \iota$. This map produces Markov chains which break off a singleton and reattach it, analogous to the top-to-random shuffle of the introduction.

More examples are at the end of the next section.

3 Building The Markov Chain

The following fact is the main motivation for the definition of a Markov chain for each non-negative convolution of projections: the probability that a riffle-shuffle takes a deck x of n cards to a deck

y is the coefficient of y in $\frac{1}{2^n} m\Delta(x)$. In other words, the *transition matrix* of the riffle-shuffling of n cards is $\left[\frac{1}{2^n} m\Delta\right]^T_{\mathcal{B}_n}$, the transpose of the matrix of the linear operator $\frac{1}{2^n} m\Delta$ with respect to the basis $\mathcal{B}_n$ of words. A similar direct calculation shows that the top-to-random shuffle of n cards has transition matrix $\left[\frac{1}{n} \operatorname{Proj}_1 *\iota\right]^T_{\mathcal{B}_n}$.

These observations suggest defining the transition matrix to be $\left[\frac{1}{\beta}\mathbf{T}\right]^T_{\mathcal{B}_n}$ on other combinatorial Hopf algebras, for other non-negative convolution of projections $\mathbf{T}$ and some appropriate number β. However, such a matrix represents transition probabilities only when each of its rows sums to 1. In other words, this naive generalisation fails if the rows of $[\mathbf{T}]^T_{\mathcal{B}_n}$ do not sum to the same number. One of the major findings of (Diaconis, Pang, and Ram 2014, Th. 3.4) is that, when $\mathbf{T}$ is a Hopf-power map $\mathbf{T} = m^{[a]}\Delta^{[a]}$, and $\mathcal{B}$ is a state space basis, then it is possible to define a rescaling $\check{\mathcal{B}}$ of $\mathcal{B}$ so the row sums of $[\mathbf{T}]^T_{\check{\mathcal{B}}_n}$ are equal. (Pang 2014, Sec. 4.3) gives a much slicker and more general description of this rescaling, in terms of the Doob transform. This allows a generalisation to linear maps on $\mathcal{H}$ that are not the Hopf-power. Indeed, (Pang 2014, Th. 3.1) describes the choice of rescalings $\check{\mathcal{B}}$ of $\mathcal{B}$ that are available for arbitrary linear maps.

It happens that the standard rescaling for the Hopf-power maps also applies to non-negative convolutions of projections; indeed, the rescaling necessary to construct a $\mathbf{T}$-Markov chain depends only on the underlying Hopf algebra, and not on the map $\mathbf{T}$. This resulting Markov chain is:

Definition 3.1. Let $\mathcal{H} = \oplus_{n\geq 0}\mathcal{H}_n$ be a graded connected Hopf algebra with state space basis $\mathcal{B}$. For $x \in \mathcal{B}_n$, let $\eta(x)$ denote the sum of the coefficients (in the $\mathcal{B}^{\otimes n}$ basis) of $\operatorname{Proj}_1^{\otimes n}\Delta^{[n]}(x)$, and let

$$\check{\mathcal{B}}_n := \left\{\left.\frac{x}{\eta(x)}\right| x \in \mathcal{B}_n\right\}.$$

Let $\mathbf{T} = \sum \alpha_D \operatorname{Proj}_{d_1} * \cdots * \operatorname{Proj}_{d_a}$ be a non-negative convolution of projections on $\mathcal{H}_n$. Then the $\mathbf{T}$-*Markov chain on* $\mathcal{B}_n$ has transition matrix

$$\left[\frac{1}{\beta_n}\mathbf{T}\right]^T_{\check{\mathcal{B}}_n}, \quad \text{where} \quad \beta_n := \sum_{d_1+\cdots+d_a=n} \alpha_D\binom{n}{d_1 \ldots d_a}.$$

The hypotheses of a state space basis ensure that $\eta(x)$ is never zero. In the shuffle algebra, $\eta(x) = 1$ always, so no rescaling is necessary.

Expressing the above transition probabilities in terms of a "natural" process requires careful analysis of the underlying Hopf algebra. Fortunately, one only needs to do this once for each Hopf algebra to interpret all its $\mathbf{T}$-Markov chains, as an analogue of (Pang 2014, Th. 4.4.1) shows that these have the form:

1. Choose a weak-composition $(d_1, \ldots, d_a)$ of n with probability $\frac{1}{\beta_n}\alpha_D\binom{n}{d_1\ldots d_a}$.

2. Choose a way to break into pieces of sizes $d_1, \ldots, d_a$.

3. Choose a way to combine these pieces.

Here, the probabilities of the choices in steps 2 and 3 depend only on the Hopf algebra, not on $\mathbf{T}$. (The exact expressions for these probabilities are unsightly and not instructive; the interested reader may consult (Pang 2014, Th. 4.4.1).) For example, a $\mathbf{T}$-*shuffle* (the $\mathbf{T}$-Markov chain on the shuffle algebra) of n cards is the following:

1,2. Cut the deck into piles of sizes $(d_1, \ldots, d_a)$ with probability $\frac{1}{\beta_n}\alpha_D\binom{n}{d_1\ldots d_a}$.

3.   Drop one-by-one the bottommost card from a pile chosen with probability proportional to the current pile size.

Aside from the GSR riffle-shuffle, its a-handed generalisation, and the top-to-random shuffle, here are some additional notable **T**-shuffles:

Example 3.2. (Assaf, Diaconis, and Soundararajan 2012) give the mixing time for shuffles with biased cuts, when the shuffler prefers to take more cards in one hand than the other. Here, the probability of cutting i cards off the top of a deck of n cards is the asymmetric binomial, $q^i(1 - q)^{n-i}\binom{n}{i}$, for some parameter $q \in (0, 1)$. The associated non-negative convolution of projections is

$$\mathbf{T} = \sum_{i=0}^{n} q^i(1 - q)^{n-i} \operatorname{Proj}_i * \operatorname{Proj}_{n-i}.$$

This has an obvious a-handed generalisation, with $a - 1$ parameters. Setting all parameters to $\frac{1}{a}$ then recovers the a-handed riffle-shuffle (even though the associated non-negative convolution of projections is then $a^{-n}\iota^{*a}$ instead of ι^{*a}, as these Markov chains depend on **T** only up to scaling).

Example 3.3. (Diaconis, Fill, and Pitman 1992, Sec. 2 and Sec. 6, Ex. 2) discuss two notions of top-m-to-random shuffles: $\mathbf{T} = \operatorname{Proj}_m *\iota$ corresponds to cutting off m cards and re-inserting them randomly, keeping their relative order, whilst $\mathbf{T} = \operatorname{Proj}_1^{*m} *\iota$ cuts m cards off and inserts them randomly in any order. For both schemes, they show that $\frac{n}{m} \log n$ shuffles suffice to randomise the deck.

Example 3.4. Taking $\mathbf{T} = \operatorname{Proj}_1 *\iota + \iota * \operatorname{Proj}_1$ produces a shuffle where the pile sizes are $(1, n-1)$ or $(n - 1, 1)$, each with probability $\frac{1}{2}$. In other words, flip a fair coin, and perform a top-to-random shuffle if the coin shows heads, and a bottom-to-random shuffle if it shows tails. This is the (symmetric) *top-or-bottom-to-random shuffle* of (Diaconis, Fill, and Pitman 1992, Sec. 6, Ex 4). It is easy to introduce an asymmetry here: for $q \in [0, 1]$, take $\mathbf{T} = q \operatorname{Proj}_1 *\iota + (1 - q)\iota * \operatorname{Proj}_1$. Setting $q = 1$ then recovers the top-to-random shuffle. Theorem 5.1 below exhibits an eigenbasis for this map on cocommutative Hopf algebras.

The Markov chains coming from the above choices of **T**, on other Hopf-algebras, are also interesting.

Example 3.5. The irreducible representations of the symmetric groups form a basis of a Hopf algebra, with product being external induction, and coproduct coming from restriction to Young subgroups. It is straightforward to adapt (Pang 2014, Ex. 4.4.3) to give the following description for each step of a **T**-Markov chain, starting from a representation x of $\mathfrak{S}_n$:

1.   Choose a Young subgroup $\mathfrak{S}_{d_1} \times \cdots \times \mathfrak{S}_{d_a}$ of $\mathfrak{S}_n$ with probability $\frac{1}{\beta_n}\alpha_D\binom{n}{d_1 \ldots d_a}$.

2.   Restrict the starting state x to the chosen subgroup.

3.   Induce it back up to $\mathfrak{S}_n$, then pick an irreducible constituent with probability proportional to the dimension of its isotypic component.

In particular, the $\operatorname{Proj}_1 *\iota$-chain is restricting to $\mathfrak{S}_{n-1}$, inducing back to $\mathfrak{S}_n$, then choosing an irreducible constituent as in step 3. This chain previously appeared in the work of (Fulman 2009), where it generates central limit theorems for character ratios.

For a more involved example, see Example 5.3 regarding the $(q \operatorname{Proj}_1 *\iota + (1-q)\iota * \operatorname{Proj}_1)$-Markov chain on trees.

4   Projection Theorem and Eigenvalue Multiplicities

As outlined in the introduction, viewing Markov chains in this Hopf-theoretic framework gives two useful consequences. The first is the construction of Markov statistics from maps between Hopf algebras:

Theorem 4.1. *Let $\mathcal{H}$, $\bar{\mathcal{H}}$ be graded, connected Hopf algebras with bases $\mathcal{B}$, $\bar{\mathcal{B}}$ respectively. Suppose in addition that $\mathcal{B}$ is a state space basis. Let $\mathbf{T}$ be a non-negative convolution of projections. If $\theta : \mathcal{H} \to \bar{\mathcal{H}}$ is a Hopf-morphism such that $\theta(\mathcal{B}_n) = \bar{\mathcal{B}}_n$ for all n, then the $\mathbf{T}$-chain on $\bar{\mathcal{B}}_n$ is the projection under θ of the $\mathbf{T}$-chain on $\mathcal{B}_n$.*

It follows that θ is a Markov statistic for the $\mathbf{T}$-chain on $\mathcal{B}_n$ - this fact would be interesting by itself even if the projected chain were not identified as the $\mathbf{T}$-chain on the target Hopf algebra.

The second profit of the Hopf formulation is the following expression for all the eigenvalues and multiplicites of these Markov chains, which hint at their long term behaviour. It follow from (Patras 1994, Th. II.7), (Krob, Leclerc, and Thibon 1997, Prop. 3.12), (Bonnafé and Pfeiffer 2008, Prop. 3.10) and the arguments of (Krob, Leclerc, and Thibon 1997, Th. 3.21) and of (Aguiar and Lauve 2013), or alternatively from the Poincaré-Birkhoff-Witt straightening algorithm and the Perron-Frobenius theorem.

Theorem 4.2. *Work in the setup of Definition 3.1. Given a partition $\lambda := (\lambda_1, \ldots, \lambda_l)$ and a weak-composition $D = (d_1, \ldots, d_a)$, let $\langle \lambda, D \rangle$ denote the number of set partitions $B_1 | \ldots | B_a$ of $\{1, 2, \ldots, l\}$ such that, for each $i \in \{1, \ldots a\}$, we have $\sum_{j \in B_i} \lambda_j = d_i$. (So $\langle \lambda, D \rangle$ is equal to the inner product $\langle p_\lambda, h_D \rangle$ of symmetric functions, hence the notation.) Then the eigenvalues of $\frac{1}{\beta_n} \mathbf{T} : \mathcal{H}_n \to \mathcal{H}_n$ are*

$$\left\{ \frac{\beta_\lambda}{\beta_n} := \frac{1}{\beta_n} \sum_{D \vDash n} \alpha_D \langle \lambda, D \rangle \,\middle|\, \lambda \vdash n \right\},$$

and the multiplicity of the eigenvalue $\frac{\beta_\lambda}{\beta_n}$ is the coefficient of $x_\lambda := x_{\lambda_1} \ldots x_{\lambda_l}$ in the generating function $\prod_i (1 - x_i)^{-b_i}$, where b_i satisfies

$$\sum_n \dim \mathcal{H}_n t^n = \prod_i (1 - t^i)^{-b_i}.$$

Futhermore, $\mathbf{T}$ is diagonalisable if $\mathcal{H}$ is commutative or cocommutative.

Note that $\beta_{(n)}$ agrees with the β_n of Definition 3.1, so this is not a point of confusion. Here's how this formula specialises to some key examples:

Example 4.3. Let $\mathbf{T} = \mathrm{Proj}_1 * \iota$, the top-to-random map. Recall that this corresponds to $\alpha_{(1,n-1)} = 1$, and all other $\alpha_D = 0$. So, for all partitions λ of n, it holds that $\beta_\lambda = \langle \lambda, (1, n-1) \rangle$, and this is the number of parts of size 1 in λ, which can be $0, 1, \ldots, n-2$, or n. So the eigenvalues of a top-to-random chain on any Hopf algebra are $\beta_\lambda / \beta_{(n)} = 0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-2}{n}, 1$.

In the case of the shuffle algebra, for a deck of distinct cards, (Diaconis, Fill, and Pitman 1992, Th. 4.1) show that the multiplicity of the eigenvalue $\frac{j}{n}$ is the number of permutations with $n - j$ fixed points, and find projection operators for each eigenspace (on the right). (Hivert et al. 2011, Sec. 4.6) produce an eigenbasis by associating each permutation with $n - j$ fixed points to an eigenvector of eigenvalue $\frac{j}{n}$. The present $\mathrm{Proj}_1 * \iota$-chain framework generalises this eigenbasis algorithm for decks with repeated cards; see the remark after Proposition 5.2. It follows from a multigraded refinement of Theorem 4.2 above that, for any $\mathbf{T}$-shuffle of a deck of distinct cards, the multiplicity of the eigenvalue $\beta_\lambda / \beta_{(n)}$ is the number of permutations of cycle type λ.

Example 4.4. Take $\mathbf{T} = (q\operatorname{Proj}_1 * \iota + (1-q)\iota * \operatorname{Proj}_1)$, the "asymmetric top-or-bottom-to-random" operator. Its eigenvalues are

$$\frac{\beta_\lambda}{\beta_n} = \frac{q\langle\lambda, (1, n-1)\rangle + (1-q)\langle\lambda, (n-1, 1)\rangle}{n}.$$

Note that the definition of $\langle\lambda, D\rangle$ depends only on the part sizes of the composition D and not on their order, so $\langle\lambda, (1, n-1)\rangle = \langle\lambda, (n-1, 1)\rangle$. Hence the eigenvalues of top-or-bottom-to-random are the same as for top-to-random in Example 4.3 above, with the same multiplicities.

Using the multiplicity in Theorem 4.2 for the eigenvalue $1 = \frac{\beta_{(n)}}{\beta_n}$ shows that the following expressions, which are easily shown to be linearly independent stationary distributions, do span the eigenspace of eigenvalue 1.

Theorem 4.5. *For a fixed state space basis $\mathcal{B}$, all $\mathbf{T}$-Markov chain on $\mathcal{B}_n$ have the same set of stationary distributions. These can be uniquely written as a linear combination of the functions*

$$\pi_{c_1,\dots,c_n}(x) := \frac{\eta(x)}{n!^2} \sum_{\sigma\in\mathfrak{S}_n} \text{coefficient of } x \text{ in the product } c_{\sigma(1)}\dots c_{\sigma(n)}$$

for each multiset $\{c_1,\dots,c_n\}$ in $\mathcal{B}_1$. (Here, $\eta(x)$ are the rescaling constants of Definition 3.1.)

As noted in (Pang 2014, Th. 4.5.1), $\pi_{c_1,\dots,c_n}(x)$ essentially enumerates the ways to build x out of $c_1,\dots,c_n$ using the multiplication of the combinatorial Hopf algebra, and to then break it into singletons. In the case of card-shuffling, the unique stationary distribution for all $\mathbf{T}$-shuffles is the uniform distribution.

5   Eigenvectors and Applications

Since the coefficients α_D of a non-negative convolution of projections can take any non-negative value, it's not surprising that there is no neat universal eigenbasis algorithm for these chains. However, one case which works out nicely, thanks to the theory of dual graded graphs of (Fomin 1994, Th. 1.6.6), is the top-or-bottom-to-random chain of Example 3.4:

Theorem 5.1. *Let $\mathcal{H}$ be a graded connected Hopf algebra, and $\mathcal{P}$ be a (graded) basis of its primitive subspace. Write $\mathcal{P}$ as the disjoint union $\mathcal{P}_1 \amalg \mathcal{P}_{>1}$, where $\mathcal{P}_1$ has degree 1. Set*

$$E_j := \left\{ \sum_{i=0}^{j} \sum_{\sigma\in\mathfrak{S}_j} \binom{j}{i} q^i(1-q)^{j-i} c_{\sigma(1)}\dots c_{\sigma(i)} \left(\sum_{\tau\in\mathfrak{S}_k} p_{\tau(1)}\dots p_{\tau(k)}\right) c_{\sigma(i+1)}\dots c_{\sigma(j)} \right\},$$

ranging over all multisets $\{c_1,\dots,c_j\}$ of $\mathcal{P}_1$, and all multisets $\{p_1,\dots,p_k\}$ of $\mathcal{P}_{>1}$ where $\deg p_1 + \dots + \deg p_k = n - j$. Then E_j is a linearly independent set of eigenvectors of eigenvalue $\frac{j}{n}$ for the operator $\frac{1}{n}(q\operatorname{Proj}_1 * \iota + (1-q)\iota * \operatorname{Proj}_1) : \mathcal{H}_n \to \mathcal{H}_n$. Furthermore, if $\mathcal{H}$ is cocommutative, then $\amalg_{j=0}^{n-2} E_j \amalg E_n$ is a basis.

Here are some simple applications of these eigenvectors to the top-or-bottom-to-random shuffle of a deck of distinct cards, analogous to the statements for riffle-shuffling in the fifth paragraph of the introduction. (The shuffle algebra is commutative, so its dual is cocommutative, and the eigenvectors of $\mathbf{T}$ on this dual give right eigenfunctions of the transition matrix, from which one deduces these results.)

9

Proposition 5.2. *Let $\{X_t\}$ denote the top-or-bottom-to-random shuffle, with parameter q, of a deck of n distinct cards. Suppose the starting deck X_0 is in ascending order. Let $\mathrm{Des}(X) \subseteq \{1, 2, \ldots, n-1\}$ and $\mathrm{Peak}(X) \subseteq \{1, 2, \ldots, n-2\}$ denote the positions of the descents and peaks of X respectively. Then*

$$\mathit{Expectation} \left\{ \sum_{i \in \mathrm{Des}(X_t)} \binom{n-2}{i-1} q^{i-1}(1-q)^{n-1-i} \right\} = \left(1 - \left(\frac{n-2}{n} \right)^t \right) \frac{1}{2};$$

$$\mathit{Expectation} \left\{ \sum_{i \in \mathrm{Peak}(X_t)} \binom{n-3}{i-1} q^{i-1}(1-q)^{n-2-i} \right\} = \left(1 - \left(\frac{n-3}{n} \right)^t \right) \frac{1}{3}.$$

Setting $q = 1$ in Theorem 5.1 gives the eigenvectors for the top-to-random operator $\mathrm{Proj}_1 * \iota$ as

$$\sum_{\sigma \in \mathfrak{S}_j} c_{\sigma(1)} \cdots c_{\sigma(j)} \sum_{\tau \in \mathfrak{S}_k} p_{\tau(1)} \cdots p_{\tau(k)}.$$

These are also the eigenvectors of $\mathrm{Proj}_1^{*m} * \iota$, the unordered version of top-m-to-random as in Example 3.3, with eigenvalue $\binom{j}{m}/\binom{n}{m}$. The reason is that, on a cocommutative Hopf algebra, the map $\mathrm{Proj}_1^{*m} * \iota$ is a polynomial in $\mathrm{Proj}_1 * \iota$. Similarly, the E_j in Theorem 5.1 are eigenvectors of any polynomial in $q\,\mathrm{Proj}_1 * \iota + (1-q)\iota * \mathrm{Proj}_1$. In particular, they have eigenvalue q_2^{n-j} for the following map, corresponding to the *trinomial-top-and-bottom-to-random shuffle* of (Diaconis, Fill, and Pitman 1992, Sec. 6, Ex 6):

$$\sum_{m_1+m_2+m_3=n} \frac{1}{m_1! m_3!} q_1^{m_1} q_2^{m_2} q_3^{m_3} \, \mathrm{Proj}_1^{*m_1} * \iota * \mathrm{Proj}_1^{*m_3},$$

(Here, $q_1 + q_2 + q_3 = 1$, and the previous q is $\frac{q_1}{q_1+q_3}$ in terms of the new parameters.)

To finish, here is an example away from the world of card-shuffling, to illustrate the diversity of Markov chains that this framework can analyse.

Example 5.3. We will study the trinomial-top-and-bottom-to-random Markov chain (of two paragraphs prior) on the Connes-Kreimer Hopf algebra of rooted forests. Take as state space basis the set of all rooted forests - that is, a disjoint union of trees, each of which has a distinguished root vertex. (The vertices are unlabelled, and the embedding of the tree in the plane is immaterial.) The degree of a forest is its number of vertices. The product of two trees is their disjoint union, and the coproduct of a tree T is the sum of $T \backslash S \otimes S$ over all connected subtrees S of T which are either empty or contain the root of T. Hence $\iota * \mathrm{Proj}_1$ corresponds to removing a root of a forest, and $\mathrm{Proj}_1 * \iota$ to removing a leaf. For a full definition of this algebra, see (Connes and Kreimer 1998, second half of Sec. 2).

The Hopf-power Markov chain on this algebra was the subject of (Pang 2014, Sec. 5.3). Adapting Theorem 5.3.8 there gives the following description of the trinomial-top-and-bottom-to-random chain:

Suppose a company has a forest structure, so all employees have at most one direct superior. All employees are either regular employees or VPs, and the superior of a VP is necessarily also a VP (so the VPs in each connected component of the company form a subtree containing the root).

Each month, each regular employee independently produces excellent work with probability q_3, average work with probability q_2, and subpar work with probability q_1 (where $q_1 + q_2 + q_3 = 1$.) For each employee who produced excellent work, one by one in a random order, the regular employee

furthest up the chain of superiority from him becomes a VP. Then the employees who produced subpar work are fired, one by one starting from the most superior. Each firing causes a cascade of promotions: first, someone further down the chain of superiority from the fired employee is uniformly selected to replace him. Then, if the promoted employee was superior to anyone, then one of those is uniformly selected and promoted to his position. This process continues until someone who is not superior to anyone is promoted.

The chain keeps track of the structure of the regular employees, but does not know which employee is taking which position in the forest structure, nor does it see the structure of the VPs.

The cases $j = 2, 3$ below are analogues of Proposition 5.2 for this chain. These are inequalities, rather than equalities, because the exact eigenvectors in Theorem 5.1 are very complicated (involving a second sum), so to obtain a slicker result, we use instead the estimates $\widetilde{\widetilde{\mathbf{f}}}_j$. Defining these requires some more notation: for a regular employee u, let $\mathrm{desc}(u)$ (resp. $\mathrm{anc}(u)$) denote the number of regular employees who are further down (resp. up) from u in the chain of superiority, including u himself in both counts. (In tree language, these are the descendants and the ancestors). Also, write $n'(u)$ for the size of the connected component of regular employees containing u.

Proposition 5.4. *Let $\{X_t\}$ denote the trinomial-top-and-bottom-to-random Markov chain on the Connes-Kreimer Hopf algebra of rooted forests, with interpretation and notation as above. For each integer $j \geq 2$, define the following functions on forests:*

$$\widetilde{\widetilde{\mathbf{f}}}_j(T) := \sum_{u \in T} q_1^{\mathrm{desc}(u)} q_3^{\mathrm{anc}(u)} \binom{\mathrm{desc}(u)}{j}.$$

(The binomial coefficient is 0 if $\mathrm{desc}(u) < j$.) Then

$$\mathrm{Expectation}\left\{ \widetilde{\widetilde{\mathbf{f}}}_j(X_t) \right\} \leq q_2^{jt} \widetilde{\widetilde{\mathbf{f}}}_j(X_0) \max_{u \in X_0 : \mathrm{desc}(u) \geq j} \left\{ \binom{n'(u)}{\mathrm{anc}(u) - 1} \right\}.$$

References

Aguiar, Marcelo and Aaron Lauve (2013). "Antipode and Convolution Powers of the Identity in Graded Connected Hopf Algebras". In: *25th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2013)*. Discrete Math. Theor. Comput. Sci. Proc., AS. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, pp. 1083–1094.

Assaf, Sami, Persi Diaconis, and Kannan Soundararajan (2012). "Riffle shuffles with biased cuts". In: *24th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2012)*. Discrete Math. Theor. Comput. Sci. Proc., AR. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, pp. 445–456.

Bayer, Dave and Persi Diaconis (1992). "Trailing the dovetail shuffle to its lair". In: *Ann. Appl. Probab.* 2.2, pp. 294–313. ISSN: 1050-5164.

Bidigare, Pat, Phil Hanlon, and Dan Rockmore (1999). "A combinatorial description of the spectrum for the Tsetlin library and its generalization to hyperplane arrangements". In: *Duke Math. J.* 99.1, pp. 135–174. ISSN: 0012-7094. DOI: 10.1215/S0012-7094-99-09906-4.

Bonnafé, C. and G. Pfeiffer (2008). "Around Solomon's descent algebras". In: *Algebr. Represent. Theory* 11.6, pp. 577–602. ISSN: 1386-923X. DOI: 10.1007/s10468-008-9090-9.

Connes, Alain and Dirk Kreimer (1998). "Hopf algebras, renormalization and noncommutative geometry". In: *Comm. Math. Phys.* 199.1, pp. 203–242. ISSN: 0010-3616. DOI: 10.1007/s002200050499.

Diaconis, Persi, James Allen Fill, and Jim Pitman (1992). "Analysis of top to random shuffles". In: *Combin. Probab. Comput.* 1.2, pp. 135–155. ISSN: 0963-5483. DOI: 10.1017/S0963548300000158.

Diaconis, Persi, C. Y. Amy Pang, and Arun Ram (2014). "Hopf algebras and Markov chains: two examples and a theory". In: *J. Algebraic Combin.* 39.3, pp. 527–585. ISSN: 0925-9899. DOI: `10.1007/s10801-013-0456-7`.

Foissy, L. (2012). "Free and cofree Hopf algebras". In: *J. Pure Appl. Algebra* 216.2, pp. 480–494. ISSN: 0022-4049. DOI: `10.1016/j.jpaa.2011.07.010`.

Fomin, Sergey (1994). "Duality of graded graphs". In: *J. Algebraic Combin.* 3.4, pp. 357–404. ISSN: 0925-9899. DOI: `10.1023/A:1022412010826`.

Fulman, Jason (2009). "Commutation relations and Markov chains". In: *Probab. Theory Related Fields* 144.1-2, pp. 99–136. ISSN: 0178-8051. DOI: `10.1007/s00440-008-0143-0`.

Gessel, Ira (2010). "Descents, Peaks, and Shuffles of Permutations and Noncommutative Symmetric Functions". Slides from a talk at Workshop on Quasisymmetric Functions, Banff International Research Station. URL: `http://www.birs.ca/workshops/2010/10w5031/files/gessel.pdf`.

Grinberg, Darij and Vic Reiner (2014). "Hopf algebras in combinatorics". In: *ArXiv e-prints.* arXiv: `1409.8356`.

Hivert, Florent, Jean-Gabriel Luque, Jean-Christophe Novelli, and Jean-Yves Thibon (2011). "The $(1 - \mathbb{E})$-transform in combinatorial Hopf algebras". In: *J. Algebraic Combin.* 33.2, pp. 277–312. ISSN: 0925-9899. DOI: `10.1007/s10801-010-0245-5`.

Klausner, Lukas Daniel (2011). *Coalgebras, Hopf Algebras and Combinatorics.* Thesis. URL: `dmg.tuwien.ac.at/drmota/DALukasKlausner.pdf`.

Krob, D., B. Leclerc, and J.-Y. Thibon (1997). "Noncommutative symmetric functions. II. Transformations of alphabets". In: *Internat. J. Algebra Comput.* 7.2, pp. 181–264. ISSN: 0218-1967. DOI: `10.1142/S0218196797000113`. URL: `http://dx.doi.org/10.1142/S0218196797000113`.

Pang, C. Y. Amy (2013). "A Hopf-power Markov chain on compositions". In: *25th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2013).* Discrete Math. Theor. Comput. Sci. Proc., AS. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, pp. 499–510.

— (2014). "Hopf algebras and Markov chains". In: *ArXiv e-prints.* A revised thesis. arXiv: `1412.8221`.

Patras, F. (1994). "L'algèbre des descentes d'une bigèbre graduée". In: *J. Algebra* 170.2, pp. 547–566. ISSN: 0021-8693. DOI: `10.1006/jabr.1994.1352`.