J Theor Probab (2019) 32:1804–1844
https://doi.org/10.1007/s10959-018-0834-0

Lumpings of Algebraic Markov Chains Arise from Subquotients

C. Y. Amy Pang[1,2]

Received: 23 August 2017 / Revised: 23 May 2018 / Published online: 14 June 2018

Abstract A function on the state space of a Markov chain is a "lumping" if observing only the function values gives a Markov chain. We give very general conditions for lumpings of a large class of algebraically defined Markov chains, which include random walks on groups and other common constructions. We specialise these criteria to the case of descent operator chains from combinatorial Hopf algebras, and, as an example, construct a "top-to-random-with-standardisation" chain on permutations that lumps to a popular restriction-then-induction chain on partitions, using the fact that the algebra of symmetric functions is a subquotient of the Malvenuto–Reutenauer algebra.

Keywords Markov chain · Random walks on groups · Card shuffling · Combinatorial Hopf algebras

Mathematics Subject Classification (2010) 60J10 · 16T30 · 05E05

1 Introduction

Combinatorialists have built a variety of frameworks for studying Markov chains algebraically, most notably the theories of random walks on groups [18,62] and their extensions to monoids [10,14,16]. Further examples include [23,32,54]. These approaches usually associate some algebraic operator to the Markov chain, the advan-

✉ C. Y. Amy Pang
amypang@hkbu.edu.hk

1 Laboratoire de Combinatoire et d'Informatique Mathématique, Université du Québec à Montréal, CP 8888, Succ. Centre-ville, Montreal, QC H3C 3P8, Canada

2 Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong

tage being that the eigendata of the operator reflects the convergence rates of the chains. As an easy example, consider n playing cards laid in a row on a table, and imagine exchanging two randomly chosen cards at each time step. [24] represents this random transposition shuffle as multiplication by the sum of transpositions in the group algebra of the symmetric group, and deduces from the representation theory of the symmetric group that asymptotically $\frac{1}{2}n \log n$ steps are required to randomise the order of the cards. See Example 2.1 for more details of the setup.

Analogous to how these frameworks translated the convergence rate calculation into an algebraic question of representations and characters, the present paper gives very general algebraic conditions for a different probability problem: when is a function θ of an algebraic Markov chain $\{X_t\}$ a Markov function, meaning that the sequence of random variables $\{\theta(X_t)\}$ is itself a Markov chain? The motivation for this is that often, only certain functions of Markov chains are of interest. For example, the random transposition shuffle above may be in preparation for a card game that only uses the cards on the left (Example 2.11), or where only the position of one specific card is important (Example 2.9). One naturally suspects that randomising only half the cards or only the position of one card may take fewer than $\frac{1}{2}n \log n$ moves, as these functions can become randomised before the full chain does. The convergence rates of such functions of Markov chains are generally easier to analyse when the function is Markov, as the *lumped chain* $\{\theta(X_t)\}$ can be studied independently of the full chain $\{X_t\}$. The reverse problem is also interesting: a Markov chain X'_t that is hard to analyse directly may benefit from being viewed as $\{\theta(X_t)\}$ for a more tractable "lift" chain $\{X_t\}$. [17,25,28] are examples of this idea.

The aim of this paper is to expedite the search for lumpings and lifts of "algebraic" Markov chains by giving very general conditions for their existence. As formalised in Sect. 2.2, the chains under consideration are associated with a linear transformation $T : V \to V$, where the state space is a basis $\mathscr{B}$ of the vector space V. Our two main discoveries for such chains are:

- (Section 2.3, Theorem 2.7) if T descends to a well-defined map $\bar{T}$ on a quotient space $\bar{V}$ of V that "respects the basis" $\mathscr{B}$, then the quotient projection $\theta : V \to \bar{V}$ gives a lumping from any initial distribution on $\mathscr{B}$. The lumped chain is associated with $\bar{T} : \bar{V} \to \bar{V}$.
- (Section 2.4, Theorem 2.16) if V contains a T-invariant subspace V' that "respects the basis" $\mathscr{B}$, then $T : V' \to V'$ corresponds to a lumping that is only valid for certain initial distributions, i.e. a *weak lumping*.

Section 2 states and proves the above very general theorems, and illustrates them with numerous simple examples, both classical and new.

Section 3 specialises these general lumping criteria to descent operator chains on combinatorial Hopf algebras [54]—in essence, lumpings from any initial distribution correspond to quotient algebras, and weak lumpings to subalgebras. This is applied to two fairly elaborate examples. Sections 3.2, 3.3, 3.4, 3.5 demystifies a theorem of Jason Fulman [31, Th. 3.1]: the probability distribution of the RSK shape [64, Sec. 7.11] [33, Sec. 4] of a permutation, after t top-to-random shuffles (Example 2.1) from the identity, agrees with the probability distribution of a partition after t steps of a certain Markov chain that removes then re-adds a random box (see Sect. 3.2.1). Fulman remarked that the connection between these two chains was "surprising", perhaps because it is not a lumping (see the start of Sect. 3). Here, we use the new lumping criteria for descent operator chains to construct a similar chain to top-to-random shuffling that does lump to the chain on partitions, and prove that its probability distribution after t steps from the identity agrees with that of top-to-random.

The second application, in Sect. 3.6, is a Hopf-algebraic re-proof of a result of Christos Athanasiadis and Persi Diaconis [9, Ex. 5.8], that riffle-shuffles and related card shuffles lump via descent set.

2 General Theory

2.1 Matrix Notation

Given a matrix A , let $A(x, y)$ denote its entry in row x, column y, and write A^T for the transpose of A.

Let V be a vector space (over $\mathbb{R}$) with basis $\mathscr{B}$, and $\mathbf{T} : V \to V$ be a linear map. Write $[\mathbf{T}]_{\mathscr{B}}$ for the matrix of $\mathbf{T}$ with respect to $\mathscr{B}$. In other words, the entries of $[\mathbf{T}]_{\mathscr{B}}$ satisfy

$$\mathbf{T}(x) = \sum_{y \in \mathscr{B}} [\mathbf{T}]_{\mathscr{B}} (y, x) y$$

for each $x \in \mathscr{B}$.

V^* is the *dual vector space* to V, the set of linear functions from V to $\mathbb{R}$. Its natural basis is $\mathscr{B}^* := \{x^* | x \in \mathscr{B}\}$, where x^* satisfies $x^*(x) = 1$, $x^*(y) = 0$ for all $y \in \mathscr{B}$, $y \neq x$. The *dual map* to $\mathbf{T} : V \to V$ is the linear map $\mathbf{T}^* : V^* \to V^*$ satisfying $(\mathbf{T}^* f)(v) = f(\mathbf{T}v)$ for all $v \in V, f \in V^*$. Dualising a linear map is equivalent to transposing its matrix: $\left[\mathbf{T}^*\right]_{\mathscr{B}^*} = [\mathbf{T}]^T_{\mathscr{B}}$.

2.2 Markov Chains from Linear Maps via the Doob Transform

To start, here is a quick summary of the Markov chain facts required for this work. A (discrete time) Markov chain is a sequence of random variables $\{X_t\}$, where each X_t belongs to the *state space* Ω. All Markov chains here are time-independent and have a finite state space. Hence, they are each described by an $|\Omega|$-by-$|\Omega|$ *transition matrix* K: for any time t,

$$\begin{aligned} &P\{X_t = x_t | X_0 = x_0, X_1 = x_1, \ldots, X_{t-1} = x_{t-1}\} \\ &= P\{X_t = x_t | X_{t-1} = x_{t-1}\} := K(x_{t-1}, x_t). \end{aligned}$$

(Here, $P\{X|Y\}$ is the probability of event X given event Y.) If the probability distribution of X_t is expressed as a row vector g_t, then taking one step of the chain is equivalent to multiplication by K on the right: $g_t = g_{t-1} K$. (Some authors, notably [10], take the opposite convention, where $P\{X_t = y | X_{t-1} = x\} := K(y, x)$, and the distribution of X_t is represented by a column vector f_t with $f_t = Kf_{t-1}$.) Note that a matrix K specifies a Markov chain in this manner if and only if $K(x, y) \geq 0$ for all $x, y \in \Omega$, and $\sum_{y\in\Omega} K(x, y) = 1$ for each $x \in \Omega$. A probability distribution $\pi : \Omega \to \mathbb{R}$ is a *stationary distribution* if it satisfies $\sum_{x\in\Omega} \pi(x)K(x, y) = \pi(y)$ for each state $y \in \Omega$. We refer the reader to the textbooks [43,47] for more background in Markov chain theory.

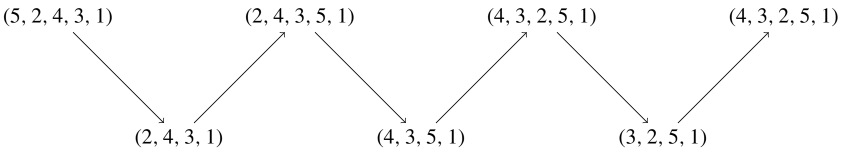
This paper concerns chains which arise from linear maps. A simple motivating example is a random walk on a group. Given a probability distribution Q on a group G (i.e. a function $Q : G \to \mathbb{R}$), consider the following Markov chain on the state space $\Omega = G$: at each time step, choose a group element g with probability $Q(g)$, and move from the current state x to the state xg. This chain is associated with the "right multiplication by $\sum_{g\in G} Q(g)g$" operator on the group algebra $\mathbb{R}G$, i.e. to the linear transformation $\mathbf{T} : \mathbb{R}G \to \mathbb{R}G$, $\mathbf{T}(x) := x\left(\sum_{g\in G} Q(g)g\right)$. To state the relationship more precisely: the transition matrix of the Markov chain is the transpose of the matrix of $\mathbf{T}$ relative to the basis G of $\mathbb{R}G$, i.e. $K = [\mathbf{T}]_G^T$. (One can define similar chains from left-multiplication operators.)

Before generalising this connection to other linear operators, here are some simple examples of random walks on groups that we will use to illustrate lumpings in later sections.

Example 2.1 (Card shuffling) Random walks on $G = \mathfrak{S}_n$, the symmetric group, describe many examples of card shuffling. The state space of these chains are the $n!$ possible orderings of a deck of n cards. For convenience, suppose the cards are labelled $1, 2, \ldots, n$, each label occurring once. There are various different conventions on how to represent such an ordering by a permutation, see [68, Sec. 2.2]. We follow the more modern notation in [8] (as opposed to [6,11]) and associate σ to the ordering where $\sigma(1)$ is the label of the top card, $\sigma(2)$ is the label of the second card from the top, ..., $\sigma(n)$ is the label of the bottom card. In other words, writing σ in one-line notation $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ (see Sect. 3.1) lists the card labels from top to bottom. Observe that in this convention, right-multiplication by a permutation τ moves the card in position $\tau(i)$ to position i.

Two simple shuffles that we will consider are:

– *top-to-random* [6,19]: remove the top card, then reinsert it into the deck at one of the n possible positions, chosen uniformly. A possible trajectory for a deck of five cards is



The associated distribution Q on $\mathfrak{S}_n$ is

$$Q(g) = \begin{cases} \frac{1}{n} & \text{if } g = (i\ i-1\ \ldots\ 1) \text{ for some } i \text{, in cycle notation;} \\ 0 & \text{otherwise.} \end{cases}$$

(Note that the identity permutation is the case $i = 1$.) Equivalently, the associated linear map is right-multiplication by

$$q = \frac{1}{n}\sum_{i=1}^{n}(i\ i-1\ \dots\ 1).$$

This Markov chain has been thoroughly analysed over the literature: [7, Sec. 1, Sec. 2] uses a strong uniform time to elegantly show that roughly $n \log n$ iterations are required to randomise the deck, and [19, Cor. 2.1] finds the explicit probabilities of achieving a particular permutation after any given number of shuffles. The time-reversal of the top-to-random shuffle is the equally well-studied Tsetlin library [67]: [41, Sec. 4.6] describes an explicit algorithm for an eigenbasis, [56] derives the spectrum for a weighted version, and [30] lists many more references.

– *random-transposition* [18,24, Chap. 3D]: choose two cards, possibly with repetition, uniformly and independently. If the same card was chosen twice, do nothing. Otherwise, exchange the two chosen cards. A possible trajectory for a deck of five cards is

$$(5, 2, 4, 3, 1) \longrightarrow (5, 4, 2, 3, 1) \longrightarrow (5, 4, 1, 3, 2) \longrightarrow (4, 5, 1, 3, 2)$$

The associated distribution Q on $\mathfrak{S}_n$ is

$$Q(g) = \begin{cases} \frac{1}{n} & \text{if } g \text{ is the identity;} \\ \frac{2}{n^2} & \text{if } g \text{ is a transposition;} \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, the associated linear map is right-multiplication by

$$q = \frac{1}{n}\,\mathrm{id} + \frac{2}{n}\sum \sigma;$$

summing over all transpositions σ.

The mixing time for the random-transposition shuffle is $\frac{1}{2}n \log n$, as shown in [24] using the representation theory of the symmetric group. [20] uses this chain to induce Markov chains on trees and on matchings. A recent extension to random involutions is [13].

Example 2.2 (Flip a random bit) [46]: Let $G = (\mathbb{Z}/2\mathbb{Z})^d$, written additively as binary strings of length d. At each time step, uniformly choose one of the d bits, and change it either from 0 to 1 or from 1 to 0. A possible trajectory for $d = 5$ is

$$(1, 0, 0, 1, 1) \longrightarrow (1, 0, 1, 1, 1) \longrightarrow (1, 0, 1, 1, 0) \longrightarrow (0, 0, 1, 1, 0)$$

The associated distribution Q is

$$Q(g) = \begin{cases} \frac{1}{d} & \text{if } g \text{ consists of } d-1 \text{ zeroes and 1 one;} \\ 0 & \text{otherwise.} \end{cases}$$

Equivalently, the associated linear map is right-multiplication by

$$q = \frac{1}{d}\left((1,0,\ldots,0) + (0,1,0,\ldots,0) + \cdots + (0,\ldots,0,1)\right).$$

(The addition in q is in the group algebra $\mathbb{R}G$, not within the group G.)

[46] investigated the return probabilities of this walk and similar walks on $(\mathbb{Z}/2\mathbb{Z})^d$ that allow changing more than one bit.

As detailed above, the transition matrix of a random walk on a group is $K = [\mathbf{T}]_G^T$, where $\mathbf{T}$ is the right-multiplication operator on the group algebra $\mathbb{R}G$ defined by $\mathbf{T}(x) := x\left(\sum_{g\in G} Q(g)g\right)$. This relationship between K and $\mathbf{T}$ allows the representation theory of G to illuminate the converge rates of the chain. A naive generalisation is to declare new transition matrices to be $K := [\mathbf{T}]_{\mathscr{B}}^T$, for other linear transformations $\mathbf{T}$ on a vector space with basis $\mathscr{B}$. The state space of such a chain is the basis $\mathscr{B}$, and the intuition is that the transition probabilities $K(x, y)$ would represent the chance of obtaining y when applying $\mathbf{T}$ to x.

In order for $K := [\mathbf{T}]_{\mathscr{B}}^T$ to be a transition matrix, we require $K(x, y) \geq 0$ for all x, y in $\mathscr{B}$, and $\sum_{y\in\mathscr{B}} K(x, y) = 1$. As noted by Persi Diaconis (personal communication), the non-negativity condition can be achieved by adding multiples of the identity transformation to $\mathbf{T}$, which essentially keeps the eigendata properties in Theorem 2.3. In any case, many linear operators arising from combinatorics already have non-negative coefficients with respect to natural bases, so we do not dwell on this problem.

The row-sum condition $\sum_{y\in\mathscr{B}} K(x, y) = 1$, though true for many important cases [10,14], is less guaranteed. There are many possible ways to adjust K so that its row sums become 1. One way which preserves the relationship between the eigendata of $\mathbf{T}$ and the convergence rates of the chain is to rescale $\mathbf{T}$ and the basis $\mathscr{B}$ using the Doob h-transform.

The Doob h-transform is a very general tool in probability, used to condition a process on some event in the future [26]. The simple case of relevance here is conditioning a (finite, discrete time) Markov chain on non-absorption. The Doob transform constructs the transition matrix of the conditioned chain out of the transition probabilities of the original chain between non-absorbing states, or, equivalently, out of the original transition matrix with the rows and columns for absorbing states removed. As observed in the multiple references below, the same recipe essentially works for any arbitrary non-negative matrix K.

The Doob transform relies on a positive *right eigenfunction* η of K, i.e. a positive function $\eta : \mathscr{B} \to \mathbb{R}$ satisfying $\sum_y K(x, y)\eta(y) = \beta\eta(x)$ for some positive number β, which is the eigenvalue. (Functions satisfying this condition with $\beta = 1$ are called *harmonic*, hence the name h-transform.) To say this in a basis-independent way, recall that $K = [\mathbf{T}]_{\mathscr{B}}^T = \left[\mathbf{T}^*\right]_{\mathscr{B}^*}$, so η (or more accurately, its linear extension in V^*) is an eigenvector of the dual map $\mathbf{T}^* : V^* \to V^*$ with eigenvalue β, i.e. $\eta \circ T = \beta\eta$ as functions on V.

Theorem 2.3 (Markov chains from linear maps via the Doob h-transform) [44, Def. 8.11, 8.12] [47, Sec. 17.6.1] [69, Lem. 4.4.1] [66, Lem. 1.4, Lem. 2.11] *Let*

V be a finite-dimensional vector space with basis $\mathscr{B}$, and $\mathbf{T} : V \to V$ be a linear map for which $K := [\mathbf{T}]_{\mathscr{B}}^{T}$ has all entries non-negative. Suppose K has a positive right eigenfunction η with eigenvalue $\beta > 0$. Then

i. *The matrix*

$$\check{K}(x, y) := \frac{1}{\beta} K(x, y) \frac{\eta(y)}{\eta(x)}$$

is a transition matrix. Equivalently, $\check{K} := \left[\frac{\mathbf{T}}{\beta}\right]_{\check{\mathscr{B}}}^{T}$, where $\check{\mathscr{B}} := \left\{\frac{x}{\eta(x)} : x \in \mathscr{B}\right\}$.

ii. *The left eigenfunctions $\mathbf{g}$ for $\check{K}$, with eigenvalue α (i.e. $\sum_x \mathbf{g}(x) K(x, y) = \alpha \mathbf{g}(y)$), are in bijection with the eigenvectors $g \in V$ of $\mathbf{T}$, with eigenvalue $\frac{\alpha}{\beta}$, via*

$$\mathbf{g}(x) := \eta(x) \times \textit{coefficient of } x \textit{ in } g.$$

iii. *The stationary distributions π for $\check{K}$ are precisely the functions of the form*

$$\pi(x) := \eta(x) \frac{\xi_x}{\sum_{x \in \mathscr{B}} \xi_x \eta(x)},$$

where $\sum_{x \in \mathscr{B}} \xi_x x \in V$ is an eigenvector of $\mathbf{T}$ with eigenvalue 1, whose coefficients ξ_x are all non-negative.

iv *The right eigenfunctions $\mathbf{f}$ for $\check{K}$, with eigenvalue α (i.e. $\sum_y K(x, y) \mathbf{f}(y) = \alpha \mathbf{f}(x)$), are in bijection with the eigenvectors $f \in V^*$ of the dual map $\mathbf{T}^*$, with eigenvalue $\frac{\alpha}{\beta}$, via*

$$\mathbf{f}(x) := \frac{1}{\eta(x)} f(x).$$

The function $\eta : V \to \mathbb{R}$ above is called the *rescaling function*. The output $\check{K}$ of the transform depends on the choice of rescaling function. Observe that, if $K := [\mathbf{T}]_{\mathscr{B}}^{T}$ already has each row summing to 1, then the constant function $\eta \equiv 1$ on $\mathscr{B}$ is a positive right eigenfunction of K with eigenvalue 1, and using this constant rescaling function results in no rescaling at all: $\check{K} = K$. This will be the case in all examples in Sects. 2.3, 2.4, so the reader may wish to skip the remainder of this section on first reading, and assume $\eta \equiv 1$ on $\mathscr{B}$ in all theorems. (Note that $\eta \equiv 1$ on $\mathscr{B}$ does not mean η is constant on V, since η is a linear function. Instead, η sends a vector v in V to the sum of the coefficients when v is expanded in the $\mathscr{B}$ basis.)

Proof To prove i, first note that $\check{K}(x, y) \geq 0$ because $\beta > 0$ and $\eta(x) > 0$ for all x. And the rows of $\check{K}$ sum to 1 because

$$\sum_y \check{K}(x, y) = \frac{\sum_y K(x, y)\eta(y)}{\beta\eta(x)} = \frac{\beta\eta(x)}{\beta\eta(x)} = 1.$$

Parts ii and iv are immediate from the definition of $\check{K}$. To see part iii, recall that a stationary distribution is precisely a positive left eigenfunction with eigenvalue 1 (normalised to be a distribution). □

Example 2.4 To illustrate the Doob transform, here is the down–up chain on partitions of size 3. (See Sect. 3.2.1 for a general description, and an interpretation in terms of restriction and induction of representations of symmetric groups.)

Let V_3 be the vector space with basis $\mathscr{B}_3$, consisting of the three partitions of size 3:

$$\mathscr{B}_3 := \left\{ \ydiagram{3}, \ydiagram{2,1}, \ydiagram{1,1,1} \right\}.$$

We define below a second vector space V_2, and linear transformations $\mathbf{D} : V_3 \to V_2$, $\mathbf{U} : V_2 \to V_3$ whose composition $\mathbf{T} = \mathbf{U} \circ \mathbf{D}$ will define our Markov chain on $\mathscr{B}_3$.

V_2 is the vector space with basis $\mathscr{B}_2$, the two partitions of size 2

$$\mathscr{B}_2 := \left\{ \ydiagram{2}, \ydiagram{1,1} \right\}.$$

For $x \in \mathscr{B}_3$, define $\mathbf{D}(x)$ to be the sum of all elements of $\mathscr{B}_2$ which can be obtained from x by deleting a box on the right end of any row. So

$$\mathbf{D}\left(\ydiagram{3}\right) = \ydiagram{2};$$
$$\mathbf{D}\left(\ydiagram{2,1}\right) = \ydiagram{2} + \ydiagram{1,1};$$
$$\mathbf{D}\left(\ydiagram{1,1,1}\right) = \ydiagram{1,1};$$

Then, for $x \in \mathscr{B}_2$, define $\mathbf{U}(x)$ to be the sum of all elements of $\mathscr{B}_3$ which can be obtained from x by adding a new box on the right end of any row, including on the row below the last row of x. So

$$\mathbf{U}\left(\yng(2)\right) = \yng(3) + \yng(2,1);$$

$$\mathbf{U}\left(\yng(1,1)\right) = \yng(2,1) + \yng(1,1,1);$$

An easy calculation shows that

$$K = [\mathbf{U} \circ \mathbf{D}]^T_{\mathscr{B}_3} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

which has all entries non-negative, but its rows do not sum to 1.

The function $\eta : \mathscr{B}_3 \to \mathbb{R}$ defined by

$$\eta\left(\yng(3)\right) = 1; \ \eta\left(\yng(2,1)\right) = 2; \ \eta\left(\yng(1,1,1)\right) = 1$$

is a right eigenfunction of K with eigenvalue $\beta = 3$. So applying the Doob transform with this choice of rescaling function amounts to dividing every entry of K by 3, then dividing the middle row by 2 and multiplying the middle column by 2, giving

$$\check{K} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ 0 & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

This is a transition matrix as its rows sum to 1. Observe that $\check{K} = \left[\frac{1}{3}\mathbf{U} \circ \mathbf{D}\right]^T_{\check{\mathscr{B}}_3}$, where

$$\check{\mathscr{B}}_3 := \left\{ \yng(3),\ \frac{1}{2}\yng(2,1),\ \yng(1,1,1) \right\}.$$

2.3 Quotient Spaces and Strong Lumping

As remarked in the introduction, sometimes only certain features of a Markov chain is of interest, that is, we wish to study a process $\{\theta(X_t)\}$ rather than $\{X_t\}$, for some function θ on the state space. The process $\{\theta(X_t)\}$ is called a lumping (or projection), because it groups together states with the same image under θ, treating them as a single state. The analysis of a lumping is easiest when $\{\theta(X_t)\}$ is itself a Markov chain. If this is true regardless of the initial state of the full chain $\{X_t\}$, then the lumping is strong; if it is dependent on the initial state, the lumping is weak. [43, Sec. 6.3, 6.4] is a very thorough exposition on these topics.

This section focuses on strong lumping; the next section will handle weak lumping.

Definition 2.5 (*Strong lumping*) Let $\{X_t\}, \{\bar{X}_t\}$ be Markov chains on state spaces $\Omega, \bar{\Omega}$, respectively, with transition matrices $K, \bar{K}$. Then $\{\bar{X}_t\}$ is a *strong lumping of* $\{X_t\}$ *via* θ if there is a surjection $\theta : \Omega \to \bar{\Omega}$ such that the process $\{\theta(X_t)\}$ is a Markov chain with transition matrix $\bar{K}$, irrespective of the starting distribution X_0. In this case, $\{X_t\}$ is a *strong lift of* $\{\bar{X}_t\}$ *via* θ.

The following necessary and sufficient condition for strong lumping is known as Dynkin's criterion:

Theorem 2.6 *(*Strong lumping for Markov chains*)* [43, Th. 6.3.2] *Let $\{X_t\}$ be a Markov chain on a state space Ω with transition matrix K, and let $\theta : \Omega \to \bar{\Omega}$ be a surjection. Then $\{X_t\}$ has a strong lumping via θ if and only if, for every $x_1, x_2 \in \Omega$ with $\theta(x_1) = \theta(x_2)$, and every $\bar{y} \in \bar{\Omega}$, the transition probability sums satisfy*

$$\sum_{y:\theta(y)=\bar{y}} K(x_1, y) = \sum_{y:\theta(y)=\bar{y}} K(x_2, y).$$

The lumped chain has transition matrix

$$\bar{K}(\bar{x}, \bar{y}) := \sum_{y:\theta(y)=\bar{y}} K(x, y)$$

for any x with $\theta(x) = \bar{x}$. □

When the chain $\{X_t\}$ arises from linear operators via the Doob h-transform, Dynkin's criterion translates into the statement below regarding quotient operators.

Theorem 2.7 (Strong lumping for Markov chains from linear maps) [52, Th. 3.4.1] *Let V be a vector space with basis $\mathscr{B}$, and $\mathbf{T} : V \to V, \eta : V \to \mathbb{R}$ be linear maps allowing the Doob transform Markov chain construction of Theorem 2.3. Let $\bar{V}$ be a quotient space of V, and denote the quotient map by $\theta : V \to \bar{V}$. Suppose that*

1. *the distinct elements of $\{\theta(x) : x \in \mathscr{B}\}$ are linearly independent, and*
2. *$\mathbf{T}, \eta$ descend to maps on $\bar{V}$ - that is, there exists $\bar{\mathbf{T}} : \bar{V} \to \bar{V}$, $\bar{\eta} : \bar{V} \to \mathbb{R}$, such that $\theta \circ \mathbf{T} = \bar{\mathbf{T}} \circ \theta$ and $\bar{\eta} \circ \theta = \eta$.*

Then, the Markov chain defined by $\bar{\mathbf{T}}$ (on the basis $\bar{\mathscr{B}} := \{\theta(x) : x \in \mathscr{B}\}$, with rescaling function $\bar{\eta}$) is a strong lumping via θ of the Markov chain defined by $\mathbf{T}$.

In the simplified case where $\eta \equiv 1$ on $\mathscr{B}$ and $\beta = 1$ (so no rescaling is required to define the chain on $\mathscr{B}$), such as for random walks on groups, taking $\bar{\eta} \equiv 1$ on $\bar{\mathscr{B}}$ satisfies $\bar{\eta} \circ \theta = \eta$, so condition 2 reduces to a condition on $\mathbf{T}$ only, and the lumped chain also does not require rescaling.

In the general case, the idea of the proof is that $\theta \circ \mathbf{T} = \bar{\mathbf{T}} \circ \theta$ is essentially equivalent to Dynkin's criterion for the unscaled matrices $K := [\mathbf{T}]_{\mathscr{B}}^T$ and $\bar{K} := [\bar{\mathbf{T}}]_{\bar{\mathscr{B}}}^T$, and this turns out to imply Dynkin's criterion for the Doob-transformed transition matrices.

Proof Let $K = [\mathbf{T}]_{\mathscr{B}}^T$, $\bar{K} = [\bar{\mathbf{T}}]_{\bar{\mathscr{B}}}^T$, and let β be the eigenvalue of η. The first step is to show that $\bar{\eta}$ is a possible rescaling function for $\bar{\mathbf{T}}$, i.e. $\bar{\eta}$ is an eigenvector of $\bar{\mathbf{T}}^*$

with eigenvalue β, taking positive values on $\bar{\mathscr{B}}$. In other words, the requirement is that $[\bar{\mathbf{T}}^*(\bar{\eta})]v = \beta\bar{\eta}v$ for every $v \in \bar{V}$, and $\bar{\eta}(v) > 0$ if $v \in \bar{\mathscr{B}}$. Since $\theta : V \to \bar{V}$ and its restriction $\theta : \mathscr{B} \to \bar{\mathscr{B}}$ are both surjective, it suffices to verify the above two conditions for $v = \theta(x)$ with $x \in V$ and with $x \in \mathscr{B}$, respectively.

Now

$$[\bar{\mathbf{T}}^*(\bar{\eta})](\theta x) = \bar{\eta} \circ \bar{\mathbf{T}}(\theta x) = \bar{\eta} \circ \theta \circ \mathbf{T}(x) = \eta \circ \mathbf{T}(x) = \mathbf{T}^*\eta(x) = \beta\eta(x) = [\beta\bar{\eta}]\theta(x).$$

And, for $x \in \mathscr{B}$, we have $\bar{\eta}(\theta(x)) = \eta(x) > 0$.

Now let $\check{K}, \check{\bar{K}}$ denote the transition matrices that the Doob transform constructs from K and $\bar{K}$. (Strictly speaking, we do not yet know that the entries of $\bar{K}$ are non-negative—this will be proved in Eq. (2.1) below—but the formula in the definition of the Doob transform remains well-defined nevertheless.) By Theorem 2.6 above, it suffices to show that, for any $x \in \mathscr{B}$ with $\theta(x) = \bar{x}$, and any $\bar{y} \in \bar{\mathscr{B}}$,

$$\check{\bar{K}}(\bar{x}, \bar{y}) = \sum_{y:\theta(y)=\bar{y}} \check{K}(x, y).$$

By definition of the Doob transform, this is equivalent to

$$\frac{1}{\beta}\bar{K}(\bar{x}, \bar{y})\frac{\bar{\eta}(\bar{y})}{\bar{\eta}(\bar{x})} = \frac{1}{\beta}\sum_{y:\theta(y)=\bar{y}} K(x, y)\frac{\eta(y)}{\eta(x)}.$$

Because $\bar{\eta}\theta = \eta$, the desired equality reduces to

$$\bar{K}(\bar{x}, \bar{y}) = \sum_{y:\theta(y)=\bar{y}} K(x, y). \tag{2.1}$$

Now expand both sides of $\bar{\mathbf{T}} \circ \theta(x) = \theta \circ \mathbf{T}(x)$ in the $\bar{\mathscr{B}}$ basis:

$$\sum_{\bar{y}\in\bar{\mathscr{B}}} \bar{K}(\bar{x}, \bar{y})\bar{y} = \theta\left(\sum_{y\in\mathscr{B}} K(x, y)y\right) = \sum_{\bar{y}\in\bar{\mathscr{B}}}\left(\sum_{y:\theta(y)=\bar{y}} K(x, y)\right)\bar{y}.$$

Equating coefficients of $\bar{y}$ on both sides completes the proof. □

Example 2.8 (Forget the last bit under "flip a random bit") Take $G = (\mathbb{Z}/2\mathbb{Z})^d$, the additive group of binary strings of length d, as in Example 2.2. Then $\bar{G} = (\mathbb{Z}/2\mathbb{Z})^{d-1}$ is a quotient group of G, by forgetting the last bit. The quotient map $\theta : G \to \bar{G}$ induces a surjective map $\theta : \mathbb{R}G \to \mathbb{R}\bar{G}$.

Recall that the "flip a random bit" chains comes from the linear transformation on $\mathbb{R}G$ of right-multiplication by $q = \frac{1}{d}\Big((1, 0, \ldots, 0) + (0, 1, 0, \ldots, 0) + \cdots + (0, \ldots, 0, 1)\Big)$. Since multiplication of group elements descends to quotient groups, forgetting the last bit is a lumping, and the lumped chain is associated with right-multiplication in $\mathbb{R}\bar{G}$ by the image of q in $\mathbb{R}\bar{G}$, which is $\frac{1}{d}\Big((1, 0, \dots, 0) + (0, 1, 0, \dots, 0) + \dots + (0, \dots, 0, 1) + (0, \dots, 0)\Big)$, where there are d summands each of length $d - 1$.

The analogous construction holds for any quotient $\bar{G}$ of any group G; see [8, App. IA].

The above principle extends to "quotient sets", i.e. a set of coset representatives, which need not be groups (and is extended further to double-coset representatives in [25]):

Example 2.9 ("Follow the ace of spaces" under shuffling) Consider $G = \mathfrak{S}_n$, and let $H = \mathfrak{S}_{n-1}$ be the subgroup of $\mathfrak{S}_n$ which permutes the last $n - 1$ objects. Then the transpositions $\tau_i := (1\ i)$, for $2 \leq i \leq n$, together with $\tau_1 :=$ id, give a set of right coset representatives of H. The coset $H\tau_i$ consists of all deck orderings where the card with label 1 is in the ith position from the top. Recall that right-multiplication is always well defined on the set of right cosets, so any card-shuffling model lumps by taking right cosets. This corresponds to tracking only the location of the card with label 1. [8, Sec. 2] analyses this chain in detail for the "riffle-shuffles" of [11].

The example of lumping to coset representatives can be further generalised to a framework concerning orbits under group actions; notice in the phrasing below that the theorem applies to more than random walks on groups.

Theorem 2.10 (Strong lumping to orbits under group actions) *Let V be a vector space with basis $\mathscr{B}$, and $\mathbf{T} : V \to V, \eta : V \to \mathbb{R}$ be linear maps allowing the Doob transform Markov chain construction of Theorem 2.3. Let $\{s_i : \mathscr{B} \to \mathscr{B}\}$ be a group of maps whose linear extensions to V commute with $\mathbf{T}$, and which satisfies $\eta \circ s_i = \eta$. Then the Markov chain defined by $\mathbf{T}$ lumps to a chain on the $\{s_i\}$-orbits of $\mathscr{B}$.*

This theorem recovers Example 2.9 above, of lumping a random walk on a group to right cosets, by letting s_i be left-multiplication by i, as i ranges over the subgroup H. Since s_i is left-multiplication and $\mathbf{T}$ is right-multiplication, they obviously commute. Hence any right-multiplication random walk on a group lumps via taking right cosets.

Proof Let $\bar{\mathscr{B}}$ be the sets of $\{s_i\}$-orbits of $\mathscr{B}$, and let $\theta : \mathscr{B} \to \bar{\mathscr{B}}$ send an element of $\mathscr{B}$ to its orbit. Let $\bar{V}$ be the vector space spanned by $\bar{\mathscr{B}}$. Then $\mathbf{T}$ descends to a well-defined map on $\bar{V}$ because of the following: if $\theta(x) = \theta(y)$, then $x = s_i(y)$ for some s_i, so $\mathbf{T}(x) = \mathbf{T} \circ s_i(y) = s_i \circ \mathbf{T}(y)$ (using s_i to denote the linear extension in this last expression), and so $\mathbf{T}(x)$ and $\mathbf{T}(y)$ are in the same orbit. And the condition $\eta \circ s_i = \eta$ ensures that $\bar{\eta}$ is well defined on the $\{s_i\}$-orbits. □

Below are two more specialisations of Theorem 2.10 that hold for random walks on any group as long as the element being multiplied is in the centre of the group algebra; we illustrate them with card-shuffling examples.

Example 2.11 (Values of the top k cards under random-transposition shuffling) This simple example appears to be new. Recall that the random-transposition shuffle corresponds to right-multiplication by $q = \frac{1}{n}\,\text{id} + \frac{2}{n}\sum_{i<j}(i\ \ j)$ on $\mathbb{R}\mathfrak{S}_n$. Because q is

a sum over all elements in two conjugacy classes, it is in the centre of $\mathbb{R}\mathfrak{S}_n$, hence right-multiplication by q commutes with any other right-multiplication operator. Let $s_i : \mathfrak{S}_n \to \mathfrak{S}_n$ be right-multiplication by $i \in \mathfrak{S}_{n-k}$, the subgroup of $\mathfrak{S}_n$ which only permutes the last $n-k$ objects. Then Theorem 2.10 implies that the random-transposition shuffle lumps to the orbits under this action, which are the left-cosets of $\mathfrak{S}_{n-k}$ (see also Example 2.9). The coset $\tau\mathfrak{S}_{n-k}$ consists of all decks whose top k cards are $\tau(1), \tau(2), \ldots, \tau(k)$ in that order (i.e. all decks that can be obtained from the identity by first applying τ and then permuting the bottom $n-k$ cards in any way). Hence the lumped chain tracks the values of the top k cards.

Example 2.12 (Coagulation–fragmentation) As noted in [20, Sec. 1.5], the following chain is one specialisation of the processes in [29], modelling the splitting and recombining of molecules.

Recall that the random-transposition shuffle corresponds to right-multiplication by the central element $q = \frac{1}{n}\,\mathrm{id} + \frac{2}{n}\sum_{i<j}(i\ j)$ on $\mathbb{R}\mathfrak{S}_n$. Let $s_i : \mathfrak{S}_n \to \mathfrak{S}_n$ be conjugation by the group element i, for all i in $\mathfrak{S}_n$. This conjugation action commutes with right-multiplication by q:

$$s_i \circ T(x) = i(qx)i^{-1} = q(ixi^{-1}) = T \circ s_i(x),$$

where the second equality uses that q is central. So Theorem 2.10 implies that the random-transposition shuffle lumps to the orbits under this conjugation action, which are the conjugacy classes of $\mathfrak{S}_n$. Each conjugacy class of $\mathfrak{S}_n$ consists precisely of the permutations of a specific cycle type, and so can be labelled by the multiset of cycle lengths, a partition of n (see Sect. 3.1). These cycle lengths represent the sizes of the molecules. As described in [21], right-multiplication by a transposition either joins two cycles or breaks a cycle into two, corresponding to the coagulation or fragmentation of molecules.

Our final example shows that the group $\{s_i\}$ inducing the lumping of a random walk on G need not be a subgroup of G:

Example 2.13 (The Ehrenfest Urn) [18, Chap. 3.1.3] Recall that the "flip a random bit" chain comes from the linear transformation on $\mathbb{R}\,(\mathbb{Z}/2\mathbb{Z})^d$ of right-multiplication by $q = \frac{1}{d}\,((1, 0, \ldots, 0) + (0, 1, 0, \ldots, 0) + \cdots + (0, \ldots, 0, 1))$. Let the symmetric group $\mathfrak{S}_d$ act on $(\mathbb{Z}/2\mathbb{Z})^d$ by permuting the coordinates. Because q is invariant under this action, right-multiplication by q commutes with this $\mathfrak{S}_d$ action. So the "flip a random bit" Markov chain lumps to the $\mathfrak{S}_d$-orbits, which track the number of ones in the binary string. The lumped walk is as follows: if the current state has k ones, remove a one with probability $\frac{k}{d}$; otherwise add a one. As noted by [18, Chap. 3.1.3], interpreting the state of k ones as k balls in an urn and $d-k$ balls in another urn gives the classical Ehrenfest urn model: given two urns containing d balls in total, at each step, remove a ball from either urn and place it in the other.

Remark In the previous three examples, the action $\{s_i : G \to G\}$ respects multiplication on G:

$$s_i(gh) = s_i(g)s_i(h) \tag{2.2}$$

Then the random walk from right-multiplication by q lumps to the $\{s_i\}$-orbits if and only if q is invariant under $\{s_i\}$. But Eq. 2.2 need not be true in all applications of Theorem 2.10 - see Example 2.9 where s_i is left-multiplication.

2.4 Subspaces and Weak Lumping

Now turn to the weaker notion of lumping, where the initial distribution matters.

Definition 2.14 (*Weak lumping*) Let $\{X_t\}, \{X'_t\}$ be Markov chains on state spaces Ω, Ω', respectively, with transition matrices K, K'. Then $\{X'_t\}$ is a *weak lumping of* $\{X_t\}$ *via* θ, *with initial distribution* X_0, if there is a surjection $\theta : \Omega \to \Omega'$ such that the process $\{\theta(X_t)\}$, started at the specified X_0, is a Markov chain with transition matrix K'. In this case, $\{X_t\}$ is a *weak lift of* $\{X'_t\}$ *via* θ.

[43, Th. 6.4.1] gives a complicated necessary and sufficient condition for weak lumping. (Note that they write π for the initial distribution and α for the stationary distribution.) Their simple sufficient condition [43, Th. 6.4.4] has the drawback of not identifying any valid initial distribution beyond the stationary distribution—such a result would not be useful for the many descent operator chains which are absorbing. So instead we appeal to a condition for continuous Markov processes [60, Th. 2], which when specialised to the case of discrete time and finite state spaces reads:

Theorem 2.15 (Sufficient condition for weak lumping for Markov chains) [60, Th. 2] *Let K be the transition matrix of a Markov chain $\{X_t\}$ with state space Ω. Suppose $\Omega = \coprod \Omega^i$, and there are distributions π^i on Ω such that*

1. *π^i is nonzero only on Ω^i,*
2. *The matrix*

$$K'(i, j) := \sum_{x \in \Omega^i, y \in \Omega^j} \pi^i(x) K(x, y)$$

satisfies the equality of row vectors $\pi^i K = \sum_j K'(i, j)\pi^j$ for all i, or equivalently

$$\sum_{x \in \Omega} \pi^i(x) K(x, y) = \sum_j K'(i, j)\pi^j(y)$$

for all i and all y.

Then, from any initial distribution of the form $\sum_i \alpha_i \pi^i$, for constants α_i, the chain $\{X_t\}$ lumps weakly to the chain on the state space $\{\Omega^i\}$ with transition matrix K'. □

(This condition was implicitly used in [43, Ex. 6.4.2].)

Remark As the proof of Theorem 2.16 will show, the case of chains from the Doob transform without rescaling corresponds to each π^i being the uniform distribution on Ω^i. In this case, the conditions above simplify: we require that $K'(i, j) := \frac{1}{|\Omega^i|} \sum_{x \in \Omega^i, y \in \Omega^j} K(x, y)$ satisfy $\frac{1}{|\Omega^i|} \sum_{x \in \Omega} K(x, y) = \frac{1}{|\Omega^j|} \sum_j K'(i, j)\pi^j(y)$ for all i and all $y \in \Omega^j$. In other words, the only requirement is that $\sum_{x\in\Omega} K(x, y)$ depends only on $\Omega^j \ni y$, not on y, a condition somewhat dual to Doob's.

For Markov chains arising from the Doob transform, the condition $\pi^i K = \sum_j K'(i, j)\pi^j$ translates to the existence of invariant subspaces. It may seem strange to consider the subspace spanned by $\left\{\sum_{x\in\mathscr{B}^i} x\right\}$, but Example 2.18 and Sect. 3.4.2 will show two examples that arise naturally, namely permutation statistics and congruence Hopf algebras.

Theorem 2.16 (Weak lumping for Markov chains from linear maps) *Let V be a vector space with basis $\mathscr{B}$, and $\mathbf{T} : V \to V, \eta : V \to \mathbb{R}$ be linear maps admitting the Doob transform Markov chain construction of Theorem 2.3. Suppose $\mathscr{B} = \amalg_i \mathscr{B}^i$, and write x^i for $\sum_{x\in\mathscr{B}^i} x$. Let V' be the subspace of V spanned by the x^i, and suppose $\mathbf{T}(V') \subseteq V'$. Define a map $\theta : \mathscr{B} \to \{x^i\}$ by setting $\theta(x) := x^i$ if $x \in \mathscr{B}^i$. Then, the Markov chain defined by $\mathbf{T} : V \to V$ lumps weakly to the Markov chain defined by $\mathbf{T} : V' \to V'$ (with basis $\mathscr{B}' := \{x^i\}$, and rescaling function the restriction $\eta : V' \to \mathbb{R}$) via θ, from any initial distribution of the form $P\{X_0 = x\} := \alpha_{\theta(x)} \frac{\eta(x)}{\eta(\theta(x))}$, where the αs are constants depending only on $\theta(x)$. In particular, if $\eta \equiv 1$ on $\mathscr{B}$ (so no rescaling is required to define the chain on $\mathscr{B}$), the Markov chain lumps from any distribution which is constant on each $\mathscr{B}^i$.*

Note that, in the simplified case $\eta \equiv 1$, it is generally not true that the restriction $\eta : V' \to \mathbb{R}$ is constant on $\mathscr{B}'$ - indeed, for $x^i \in \mathscr{B}'$, it holds that $\eta(x^i) = |\mathscr{B}^i|$. So a weak lumping chain from Theorem 2.16 will generally require rescaling.

Remark Suppose the conditions of Theorem 2.16 hold, and let $j : V' \hookrightarrow V$ be the inclusion map. Now the dual map $j^* : V^* \twoheadrightarrow V'^*$, and $\mathbf{T}^* : V^* \to V^*$, satisfy the hypotheses of Theorem 2.7, except that there may not be suitable rescaling functions $\eta : V^* \to \mathbb{R}$ and $\bar{\eta} : V'^* \to \mathbb{R}$. Because the Doob transform chain for $\mathbf{T}^*$ is the time-reversal of the chain for $\mathbf{T}$ [52, Th. 3.3.2], this is a reflection of [43, Th. 6.4.5].

The proof of Theorem 2.16 is at the end of this section.

Example 2.17 (Number of rising sequences under riffle-shuffling) [11, Cor. 2] The sequence $\{i, i+1, \ldots, i+j\}$ is a rising sequence of a permutation σ if those numbers appear in that order when reading the one-line notation of σ from left to right. Viewing σ as a deck of cards, this says that the card with label i is somewhere above the card with label $i+1$, which is somewhere above the card with label $i+2$, and so on, until the card with label $i+j$. Formally, $\sigma^{-1}(i) < \sigma^{-1}(i+1) < \cdots < \sigma^{-1}(i+j)$. Unless otherwise specified, a rising sequence is assumed to be maximal, i.e. $\sigma^{-1}(i-1) > \sigma^{-1}(i) < \sigma^{-1}(i+1) < \cdots < \sigma^{-1}(i+j) > \sigma^{-1}(i+j+1)$.

Following [11], write $R(\sigma)$ for the number of rising sequences in σ. This statistic is also written ides(σ), as it is the number of descents in σ^{-1}. For example, $R(2, 4, 5, 3, 1) = 3$, the three rising sequences being $\{1\}$, $\{2, 3\}$ and $\{4, 5\}$.

[11] studied the popular riffle–shuffle model, where the deck is cut into two according to a binomial distribution and interleaved. (We omit the details as this shuffle is

not the focus of the present paper). This arises from right-multiplication in $\mathbb{R}\mathfrak{S}_n$ by

$$q = \frac{n+1}{2^n}\,\mathrm{id} + \frac{1}{2^n}\sum_{R(\sigma)=2}\sigma.$$

[11, Cor. 2] shows that riffle-shuffling, if started from the identity, lumps weakly via the number of rising sequences. This result can be slightly strengthened by applying the present Theorem 2.16 in conjunction with [11, Cor. 3], which proves explicitly that q generates a subalgebra spanned by $x^i := \sum_{R(\sigma)=i}\sigma$, for $1 \le i \le n$. The authors recognised this subalgebra as equivalent to Loday's "number of descents" subalgebra [63]. (More precisely: the linear extension to $\mathbb{R}\mathfrak{S}_n$ of the inversion map $I(\sigma) := \sigma^{-1}$ is an algebra antimorphism—i.e. $I(\sigma\tau) = I(\tau)I(\sigma)$—and it sends x^i to the sum of permutations with $i-1$ descents, which span Loday's algebra.) Since the basis elements x^i have the form stipulated in Theorem 2.16, it follows that the lumping via the number of rising sequences is valid starting from any distribution that is constant on the summands of each x^i, i.e. on each subset of permutations with the same number of rising sequences.

The key idea in the previous example is that, if $q \in \mathbb{R}G$ generates a subalgebra V' of $\mathbb{R}G$ of the form described in Theorem 2.16, then this produces a weak lumping of the random walk on G given by right-multiplication by q. We apply this to the top-to-random shuffle:

Example 2.18 (Length of last rising sequence under top-to-random shuffling) This simple example appears to be new. In addition to the definitions in Example 2.17, more terminology is necessary. The *length* of the rising sequence $\{i, i+1, \ldots, i+j\}$ is $j+1$. Following [19], write $L(\sigma)$ for the length of the last rising sequence, meaning the one which contains n. For example, the rising sequences of $(2, 4, 3, 5, 1)$ have lengths $1, 2, 2$, respectively, and $L(2, 4, 3, 5, 1) = 2$.

Recall that the top-to-random shuffle is given by right-multiplication by

$$q = \frac{1}{n}\sum_{i=1}^{n}(i\ i-1\ \ldots\ 1).$$

The rising sequences of the cycles $(i\ i-1\ \ldots\ 1)$ are precisely $\{1\}$ and $\{2, 3, \ldots, n\}$, and these are the only permutations σ with $L(\sigma) = n-1$. [19, Th. 4.2] shows that the algebra generated by q is spanned by $x^i := \sum_{L(\sigma)=i}\sigma$, for $1 \le i \le n$. Thus, Theorem 2.16 shows that top-to-random shuffling weakly lumps via the length of the last rising sequence, starting from any distribution that is constant on permutations with the same last rising sequence length. In particular, since the identity is the only permutation with $L(\sigma) = n$, the lumping holds if the deck starts at the identity permutation.

Proof of Theorem 2.16 In the notation of Theorem 2.15, the distribution π^i is

$$\pi^i(x) = \begin{cases} \frac{\eta(x)}{\eta(x^i)} & \text{if } x \in \mathscr{B}^i; \\ 0 & \text{otherwise,} \end{cases}$$

which clearly satisfies condition 1.

To check condition 2, write $\mathbf{T}'$, η' for the restrictions of $\mathbf{T}$, η to V', and set $K = [\mathbf{T}]^T_{\mathscr{B}}$, $K' = [\mathbf{T}']^T_{\mathscr{B}'}$. As in the proof of Theorem 2.7, we start by showing that η' is a possible rescaling function for $\mathbf{T}'$:

$$(\mathbf{T}')^*(\eta') = \eta' \circ \mathbf{T}' = (\eta \circ \mathbf{T})|_{V'} = (\beta\eta)|_{V'} = \beta\eta',$$

so η' is an eigenvector of $\mathbf{T}'^*$ with eigenvalue β. And η' is positive on $\mathscr{B}'$ because $\eta'(x^i) = \sum_{x\in\mathscr{B}^i} \eta(x)$, a sum of positive numbers.

Write $\check{K}$, $\check{K}'$ for the associated transition matrices. (As in the proof of Theorem 2.7, we check that the entries of K' are non-negative later, in Eq. 2.3.) We need to show that, for all i and for all $y \in \mathscr{B}$,

$$\sum_{x\in\mathscr{B}^i} \pi^i(x)\check{K}(x, y) = \sum_j \check{K}'(x^i, x^j)\pi^j(y).$$

Note that $\pi^j(y)$ is zero unless $y \in \mathscr{B}^j$, so only one summand contributes to the right-hand side. By substituting for π^i, $\check{K}$ and $\check{K}'$, the desired equality is equivalent to

$$\sum_{x\in\mathscr{B}^i} \frac{\eta(x)}{\eta(x^i)}\frac{1}{\beta}K(x, y)\frac{\eta(y)}{\eta(x)} = K'(x^i, x^j)\frac{1}{\beta}\frac{\eta(x^j)}{\eta(x^i)}\frac{\eta(y)}{\eta(x^j)},$$

which reduces to

$$\sum_{x\in\mathscr{B}^i} K(x, y) = K'(x^i, x^j) \tag{2.3}$$

for $y \in \mathscr{B}^j$.

Now, by expanding in the $\mathscr{B}'$ basis,

$$\mathbf{T}'(x^i) = \sum_j K'(x^i, x^j)x^j = \sum_j K'(x^i, x^j) \sum_{y\in\mathscr{B}^j} y.$$

On the other hand, a $\mathscr{B}$ expansion yields

$$\mathbf{T}'(x^i) = \sum_{x\in\mathscr{B}^i} \mathbf{T}(x) = \sum_{x\in\mathscr{B}^i}\sum_{y\in\mathscr{B}} K(x, y)y.$$

So

$$\sum_{y\in\mathscr{B}}\sum_{x\in\mathscr{B}^i} K(x, y)y = \sum_j K'(x^i, x^j) \sum_{y\in\mathscr{B}^j} y = \sum_y \sum_{j:y\in\mathscr{B}^j} K'(x^i, x^j)y,$$

and since $\mathscr{B}$ is a basis, the coefficients of y on the two sides must be equal. □

3 Lumpings from Subquotients of Combinatorial Hopf Algebras

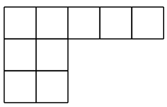
This section specialises the strong and weak lumping criteria of Part I to Markov chains from descent operators on combinatorial Hopf algebras [54]. Our main running example (which in fact motivated the entire paper) is a lift for the "down-up chain on partitions", where at each step a random box is removed and a possibly different random box added, according to a certain distribution (see Sect. 3.2.1). The stationary distribution of this chain is the Plancherel measure $\pi(\lambda) = \frac{(\dim \lambda)^2}{n!}$, where $\dim \lambda$ is the dimension of the symmetric group representation indexed by λ, or equivalently the number of standard tableaux of shape λ (see below for definitions). Because $(\dim \lambda)^2$ is the number of permutations whose RSK shape [64, Sec. 7.11] [33, Sec. 4] is λ, it is natural to ask if the down–up chain on partitions is the lumping of a chain on permutations with a uniform stationary distribution.

Fulman [31, Th. 3.1] proved that this is almost true for top-to-random shuffling: the probability distribution of the RSK shape after t top-to-random shuffles from the identity, agrees with the probability distribution after t steps of the down–up chain on partitions. However, it is not true that top-to-random shuffles lump via RSK shape. [15, Fig. 1] is an explicit 8-step trajectory of the partition chain that has no corresponding trajectory in top-to-random shuffling. In other words, it is possible for eight top-to-random shuffles to produce an RSK shape equal to the end of the exhibited partition chain trajectory, but no choice of intermediate steps will have RSK shapes equal to the given trajectory. (Technically, this figure is written for random-to-top, the time-reversal of top-to-random, so one should read it backwards from right to left, to apply it to top-to-random shuffles.)

The present Theorem 3.11 finds that top-to-random shuffling can be modified to give an honest weak lift of the down–up chain on partitions: every time a card is moved, relabel it with the current time that we moved the card, then track the (reversed) relative orders of the labels. (A different interpretation without cards is in Sect. 3.4.1) This lift is constructed in two stages—Sect. 3.3 builds a strong lift to tableaux using Hopf algebra quotients, and Sect. 3.4 builds a weak lift to permutations using Hopf subalgebras. Sect. 3.5 then shows that the multistep transition probabilities of the relabelled chain agree with the unmodified top-to-random shuffle, if both are started at the identity, thus recovering the Fulman result.

3.1 Notation

A *partition* λ is a weakly decreasing sequence of positive integers: $\lambda := (\lambda_1, \ldots, \lambda_l)$ with $\lambda_1 \geq \ldots \geq \lambda_l > 0$. This is a *partition of* n, denoted $\lambda \vdash n$, if $\lambda_1 + \cdots + \lambda_l = n$. We will think of a partition λ as a diagram of left-justified boxes with λ_1 boxes in the topmost row, λ_2 boxes in the second row, etc. For example, (5,2,2) is a partition of 9, and below is its diagram.



A *tableau of shape* λ is a filling of each of the boxes in λ with a positive integer. The *shift* of a tableaux T by an integer k, denoted $T[k]$, increases each filling of T by k. A tableau is *standard* if it is filled with $\{1, 2, \ldots, n\}$, each integer occurring once. If no two boxes of a tableau T has the same filling, then its *standardisation* $\mathrm{std}(T)$ is computed by replacing the smallest filling by 1, the second smallest filling by 2, and so on. Clearly $\mathrm{std}(T)$ is a standard tableau, of the same shape as T. A box b of T is *removable* if the difference $T \backslash b$ is a tableau. Below shows a tableau of shape (5, 2, 2), its shift by 3, and its standardisation. The removable boxes in the first tableau are 11 and 13.

1	2	5	10	13
4	8			
6	11			

T

4	5	8	13	16
7	11			
9	14			

$T[3]$

1	2	4	7	9
3	6			
5	8			

$\mathrm{std}(T)$

For a partition λ, write $\dim(\lambda)$ for the number of standard tableaux of shape λ, as this is the dimension of the symmetric group representation corresponding to λ [61, Chap. 2].

In the same vein, this paper will regard permutations as "standard words", using one-line notation: $\sigma := (\sigma(1), \ldots, \sigma(n))$. The *length* of a word is its number of letters. The *shift* of a word σ by an integer k, denoted $\sigma[k]$, increases each letter of σ by k. If a word σ has all letters distinct, then its *standardisation* $\mathrm{std}(\sigma)$ is computed by replacing the smallest letter by 1, the second smallest letter by 2, and so on. Clearly $\mathrm{std}(\sigma)$ is a permutation. For example, $\sigma = (6, 1, 4, 8, 2, 11, 10, 13, 5)$ is a word of length 9. Its shift by 3 is $\sigma[3] = (9, 4, 7, 11, 5, 14, 13, 16, 8)$, and its standardisation is $\mathrm{std}(\sigma) = (5, 1, 3, 6, 2, 8, 7, 9, 4)$.

We assume the reader is familiar with *RSK insertion*, a map from permutations to tableaux (only the insertion tableau is relevant here, not the recording tableau), see [64, Sec. 7.11] [33, Sec. 4].

A *weak-composition* D (also called a decomposition in [3]) is a list of non-negative integers $\left(d_1, d_2, \ldots, d_{l(D)}\right)$. This is a *weak-composition of* n, denoted $D \vdash n$, if $d_1 + \cdots + d_l = n$. For example, (1, 3, 0, 2, 2, 0, 1) is a weak-composition of 11.

A *composition* I is a list of positive integers $\left(i_1, i_2, \ldots, i_{l(I)}\right)$, where each i_k is a *part*. This is a *composition of* n, denoted $I \vdash n$, if $i_1 + \cdots + i_l = n$. For example, (1, 3, 2, 2, 1) is a composition of 11. Define a partial order on the compositions of n: say $J \leq I$ if J can be obtained by joining adjacent parts of I. For example, $(6, 2, 1) \leq (1, 3, 2, 2, 1)$, and also $(1, 3, 4, 1) \leq (1, 3, 2, 2, 1)$.

The *descent set* of a word $w = (w_1, \ldots, w_n)$ is defined to be $\Big\{ j \in \{1, 2, \ldots, n-1\}$ $| w_j > w_{j+1} \Big\}$. It is more convenient here to rewrite the descent set as a composition in the following way: a word w has *descent composition* $\mathrm{Des}(w) = I$ if i_j is the number of letters between the $j-1$th and jth descent, i.e. if $w_{i_1+\cdots+i_j} > w_{i_1+\cdots+i_j+1}$ for all j, and $w_r \leq w_{r+1}$ for all $r \neq i_1 + \cdots + i_j$. For example, the descent set of (6, 1, 4, 8, 2, 11, 10, 13, 5) is $\{1, 4, 6, 8\}$, and $\mathrm{Des}(6, 1, 4, 8, 2, 11, 10, 13, 5) =$

(1, 3, 2, 2, 1). Note that $\mathrm{Des}(\sigma^{-1})$ consists of the lengths of the rising sequences (as in Example 2.17) of σ.

3.2 Markov Chains from Descent Operators, and the Down–Up Chain on Partitions

The Markov chains in this and subsequent sections arise from descent operators on combinatorial Hopf algebras, through the framework of [54] as summarised below.

Loosely speaking, a combinatorial Hopf algebra is a graded vector space $\mathscr{H} = \bigoplus_{n=0}^{\infty} \mathscr{H}_n$ with a basis $\mathscr{B} = \amalg_n \mathscr{B}_n$ indexed by a family of "combinatorial objects", such as partitions, words, or permutations. The grading reflects the "size" of these objects. $\mathscr{H}$ admits a linear product map $m : \mathscr{H} \otimes \mathscr{H} \to \mathscr{H}$ and a linear coproduct map $\Delta : \mathscr{H} \to \mathscr{H} \otimes \mathscr{H}$ satisfying certain compatibility axioms; see the survey [39] for details. These two operations encode, respectively, how the combinatorial objects combine and break. The concept was originally due to Joni and Rota [42], and the theory has since been expanded in [1,2,12,40] and countless other works.

To define the descent operators, it is necessary to introduce a refinement of the coproduct relative to the grading. Given a weak-composition $D = \left(d_1, d_2, \ldots, d_{l(D)}\right)$ of n, follow [2] and define $\Delta_D : \mathscr{H}_n \to \mathscr{H}_{d_1} \otimes \cdots \otimes \mathscr{H}_{d_{l(D)}}$ to be a projection to the graded subspace $\mathscr{H}_{d_1} \otimes \cdots \otimes \mathscr{H}_{d_{l(D)}}$ of the iterated coproduct $(\Delta \otimes \mathrm{id}^{\otimes l(D)-1}) \circ \cdots \circ (\Delta \otimes \mathrm{id} \otimes \mathrm{id}) \circ (\Delta \otimes \mathrm{id}) \circ \Delta$. So Δ_D models breaking an object into $l(D)$ pieces, of sizes $d_1, \ldots, d_{l(D)}$, respectively. See the examples below.

Example 3.1 An instructive example of a combinatorial Hopf algebra is the *shuffle algebra* $\mathscr{S}$. Its basis is the set of all words in the letters $\{1, 2, \ldots, N\}$ (for some N, whose exact value is often unimportant). View the word $(w_1, \ldots, w_n)$ as the deck of cards with card w_1 on top, card w_2 second from the top, and so on, so card w_n is at the bottom. The degree of a word is its number of letters, i.e. the number of cards in the deck. The product of two words, also denoted by $\sqcup\!\sqcup$, is the sum of all their interleavings (with multiplicity), and the coproduct is deconcatenation, or cutting the deck. For example:

$$\begin{aligned} m((1,5) \otimes (5,2)) = (1,5) \sqcup\!\sqcup (5,2) &= 2(1,5,5,2) + (1,5,2,5) + (5,1,5,2) \\ &\quad + (5,1,2,5) + (5,2,1,5); \\ \Delta_{1,3}(1,5,5,2) &= (1) \otimes (5,5,2); \\ \Delta_{2,0,2}(1,5,5,2) &= (1,5) \otimes () \otimes (5,2). \end{aligned}$$

(Here, () denotes the empty word, the unit of $\mathscr{S}$.) Observe that

$$\begin{aligned} \frac{1}{4} m \circ \Delta_{1,3}(1,5,5,2) = \frac{1}{4} m((1) \otimes (5,5,2)) &= \frac{1}{4}(1,5,5,2) + \frac{1}{4}(5,1,5,2) \\ &\quad + \frac{1}{4}(5,5,1,2) + \frac{1}{4}(5,5,2,1). \end{aligned}$$

The four words that appear on the right-hand side are precisely all the possible results after a top-to-random shuffle of the deck (1, 5, 5, 2), and the coefficient of each word is the probability of obtaining it. The same is true for decks of n cards and the operator $\frac{1}{n}m \circ \Delta_{1,n-1}$.

Instead of removing only the top card—i.e. creating two piles of sizes 1 and $n-1$, respectively—consider cutting the deck into l piles of sizes $d_1, \dots, d_l$ for some weak-composition D of n. Then interleave these l piles together into one pile - i.e. uniformly choose an ordering of all n cards such that any two cards from the same one of the l piles stay in the same relative order. Such a shuffle is described by (a suitable multiple of) the composite operator $m \circ \Delta_D$. These composites (and their linear combinations) are the *descent operators* of [55], so named because, on a commutative or cocommutative Hopf algebra, their composition is equivalent to the multiplication in Solomon's descent algebra [63] of the symmetric group. This descent algebra view will be useful in the proof of Theorem 3.14, relating these shuffles to a different chain on permutations.

[19] studied more general "cut-and-interleave" shuffles where the cut composition D is random, according to some probability distribution P on the weak-compositions of n. These P-shuffles are described by

$$m \circ \Delta_P := \sum_D \frac{P(D)}{\binom{n}{d_1 \dots d_{l(D)}}} m \circ \Delta_D,$$

more precisely, their transition matrices are $[m \circ \Delta_P]^T_{\mathscr{B}_n}$, where $\mathscr{B}_n$ is the word basis of the shuffle algebra. (The notation $m \circ \Delta_P$, from [54], is non-standard and coined especially for this Markov chain application of descent operators.) Notice that, if P is concentrated at $(1, n-1)$, then $m \circ \Delta_P = \frac{1}{n} m \circ \Delta_{1,n-1}$, corresponding to the top-to-random shuffle as described in Example 3.1. The present paper will focus on the case where P is concentrated at $(n-1, 1)$, so $m \circ \Delta_P = \frac{1}{n} m \circ \Delta_{n-1,1}$ models the "bottom-to-random" shuffle.

[22,54] extend this idea to other combinatorial Hopf algebras, using $m \circ \Delta_P$ to construct Markov chains which model first breaking a combinatorial object into l pieces where the distribution of piece size is P, and then reassembling the pieces. In particular, $\frac{1}{n} m \circ \Delta_{n-1,1}$ describes removing a piece of size 1 and reattaching it. For general combinatorial Hopf algebras, this construction requires the Doob transform (Theorem 2.3).

To simplify the exposition, focus on the case where $|\mathscr{B}_1| = 1$, i.e. there is only one combinatorial object of size 1. (This is not true of the shuffle algebra, where $\mathscr{B}_1$ consists of all the different possible single card labels. Hence, we will ignore the shuffle algebra henceforth, until Sect. 3.5.) Writing $\bullet$ for this object, $\Delta_{1,\dots,1}(x)$ then is a multiple of $\bullet \otimes \cdots \otimes \bullet = \bullet^{\otimes \deg x}$. [54, Lem 3.3] showed that, under the conditions in Theorem 3.2, this multiple is a rescaling function. (Briefly, conditions i and ii guarantee that $K := [m \circ \Delta_P]$, the matrix before the Doob transform, has non-negative entries, and condition iii ensures η is positive on $\mathscr{B}_n$.)

Theorem 3.2 (Markov chains from descent operators) [54, Lem 3.3, Th. 3.4] *Suppose $\mathscr{H} = \bigoplus_n \mathscr{H}_n$ is a graded connected Hopf algebra with a basis $\mathscr{B} = \amalg_n \mathscr{B}_n$ satisfying:*

o. $\mathscr{B}_1 = \{\bullet\}$;
i. for all $w, z \in \mathscr{B}$, the expansion of $m(w \otimes z)$ in the $\mathscr{B}$ basis has all coefficients non-negative;
ii. for all $x \in \mathscr{B}$, the expansion of $\Delta(x)$ in the $\mathscr{B} \otimes \mathscr{B}$ basis has all coefficients non-negative;
iii. for all $x \in \mathscr{B}_n$ with $n > 1$, it holds that $\Delta(x) \neq 1 \otimes x + x \otimes 1$ (i.e. $\mathscr{B}_n$ contains no primitive elements when $n > 1$).

Then, for any fixed n and any probability distribution $P(D)$ on weak-compositions D of n, the corresponding descent operator $m \circ \Delta_P : \mathscr{B}_n \to \mathscr{B}_n$ given by

$$m \circ \Delta_P := \sum_D \frac{P(D)}{\binom{n}{d_1 \dots d_{l(D)}}} m \circ \Delta_D$$

and rescaling function $\eta : \mathscr{B}_n \to \mathbb{R}$ given by

$$\eta(x) := \textit{coefficient of } \bullet^{\otimes n} \textit{ in } \Delta_{1,\dots,1}(x)$$

admit the Doob transform construction of Theorem 2.3.

The stationary distributions of these chains are easy to describe:

Theorem 3.3 [54, Th. 3.12] *The unique stationary distribution of the Markov chains constructed in Theorem 3.2 is*

$$\pi(x) = \frac{1}{n!}\eta(x) \times \textit{coefficient of } x \textit{ in } \bullet^{\otimes n},$$

independent of the distribution P.

[54, Th. 3.5] derives the eigenvalues of all descent operator chains. For our main example of $\frac{1}{n}m \circ \Delta_{n-1,1}$, these eigenvalues are:

Theorem 3.4 [54, Th. 4.4.i] *For the descent operator $\frac{1}{n}m \circ \Delta_{n-1,1}$, the eigenvalues of the chains constructed in Theorem 3.2 are $\frac{j}{n}$ for $0 \leq j \leq n$, $j \neq n-1$, and their multiplicities are $\dim \mathscr{H}_{n-j} - \dim \mathscr{H}_{n-j-1}$.*

[54, Th. 4.4] also describes some eigenvectors. Since their formulae are complicated and they are not the focus of the present paper, we do not go into detail here.

3.2.1 Example: the Down–Up Chain on Partitions

We explain in detail below the Markov chain that arises from applying the Doob transform to $\frac{1}{n}m \circ \Delta_{n-1,1}$ on the algebra of symmetric functions. This chain is one focus of [31].

Work with the algebra of symmetric functions Λ [64, Chap. 7], with basis the Schur functions $\{s_\lambda\}$, which are indexed by partitions. For clarity, we will often write λ in place of s_λ. The degree of λ is the number of boxes in its diagram.

As described in [39, Sec. 2.5], Λ carries the following Hopf structure:

$$m(s_\nu \otimes s_\mu) = s_\nu s_\mu = \sum_\lambda c^\lambda_{\nu\mu} s_\lambda;$$

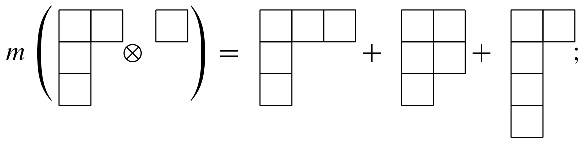
$$\Delta(s_\lambda) = \sum_{\nu,\mu} c^\lambda_{\nu\mu} s_\nu \otimes s_\mu,$$

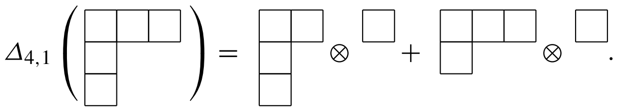
where $c^\lambda_{\nu\mu}$ are the Littlewood-Richardson coefficients. These simplify greatly when μ is the partition (1) - namely $c^\lambda_{\nu\mu} = 1$ if the diagrams of λ and ν differ by one box, and $c^\lambda_{\nu\mu} = 0$ otherwise. (This is one case of the *Pieri rule*.) Writing $\lambda \sim \nu \cup \square$ and $\nu \sim \lambda \backslash \square$ when the diagram of λ can be obtained by adding one box to the diagram of ν, the above can be summarised as

$$m(\nu \otimes (1)) = \sum_{\lambda:\lambda\sim\nu\cup\square} \lambda;$$

$$\Delta_{\deg \nu-1,1}(\lambda) = \left(\sum_{\nu:\nu\sim\lambda\backslash\square} \nu\right) \otimes (1).$$

For example,





To investigate how the Doob transform turns this data into probabilities, it is necessary to first understand the rescaling function η. By Theorem 3.2, $\eta(\lambda)$ is the coefficient of $(1)^{\otimes \deg(\lambda)}$ in $\Delta_{1,\dots,1}(\lambda)$, i.e. the number of ways to remove boxes one by one from λ. Since such ways are in bijection with the standard tableaux of shape λ, it holds that $\eta(\lambda) = \dim \lambda$. Hence, the Doob transform creates the following transition matrix from $\frac{1}{n} m \circ \Delta_{n-1,1}$:

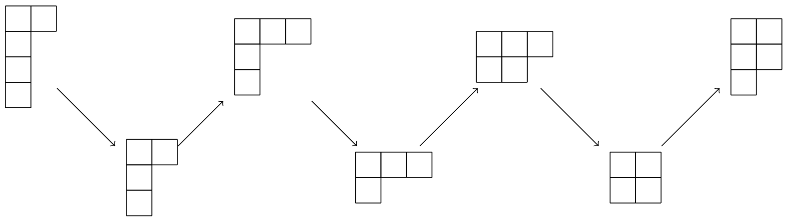
$$\check{K}(\lambda, \mu) = \sum_{\nu:\nu\sim\lambda\backslash\square, \mu\sim\nu\cup\square} \frac{1}{n} \frac{\dim \mu}{\dim \lambda}$$

$$= \sum_{\nu:\nu\sim\lambda\backslash\Box,\mu\sim\nu\cup\Box} \frac{1}{n}\frac{\dim\mu}{\dim\nu}\frac{\dim\nu}{\dim\lambda}.$$

The second expression suggests a decomposition of each time step into two parts:

1. Remove a box from λ to obtain ν, with probability $\frac{\dim\nu}{\dim\lambda}$.
2. Add a box to ν to obtain μ, with probability $\frac{1}{n}\frac{\dim\mu}{\dim\nu}$.

The two parts are illustrated by down-right and up-right arrows, respectively, in the following example trajectory in degree 5:



One easy way to implement step 1, the box removal, is via the *hook walk* of [37]: uniformly choose a box b, then uniformly choose a box in the *hook* of b—that is, to the right or below b—and continue uniformly picking from successive hooks until you reach a removable box. Similarly, step 2 can be implemented using the *complimentary hook walk* of [38]: start at the box (outside the partition diagram) in row n, column n, uniformly choose a box in its *complimentary hook*—that is, to its left or above it, and outside of the partition—and continue uniformly picking from complimentary hooks until you reach an addable box.

The transition matrix of this chain in degree 3 is

	(3)	(2, 1)	(1, 1, 1)
(3)	$\frac{1}{3}$	$\frac{2}{3}$	0
(2, 1)	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
(1, 1, 1)	0	$\frac{2}{3}$	$\frac{1}{3}$

.

To interpret the Markov chain on partitions from other descent operators $m \circ \Delta_P$, it is necessary to view a partition of n as an irreducible representation of $\mathfrak{S}_n$, as explained in [61, Chap. 2]. Then, the multiplication and comultiplication of partitions come, respectively, from the induction of the external product and the restriction to Young subgroups—for irreducible representations corresponding to the partitions $\mu \vdash i$, $\nu \vdash j$ and $\lambda \vdash n$,

$$\mu\nu = \mathrm{Ind}_{\mathfrak{S}_i \times \mathfrak{S}_j}^{\mathfrak{S}_{i+j}} \mu \times \nu; \quad \Delta_{n-i,i}(\lambda) = \mathrm{Res}_{\mathfrak{S}_{n-i} \otimes \mathfrak{S}_i}^{\mathfrak{S}_n} \lambda.$$

So the chains from $m \circ \Delta_P$ model restriction-then-induction, as detailed below.

Definition 3.5 Each step of the *P-restriction-then-induction* chain on irreducible representations of the symmetric group $\mathfrak{S}_n$ goes as follows:

1. Choose a weak-composition $D = (d_1, \ldots, d_{l(D)})$ of n with probability $P(D)$.
2. Restrict the current irreducible representation to the chosen Young subgroup $\mathfrak{S}_{d_1} \times \cdots \times \mathfrak{S}_{d_{l(D)}}$.
3. Induce this representation to $\mathfrak{S}_n$, then pick an irreducible constituent with probability proportional to the dimension of its isotypic component.

[31] considered similar chains for subgroups of any group $H \subseteq G$ instead of $\mathfrak{S}_{d_1} \times \cdots \times \mathfrak{S}_{d_{l(D)}} \subseteq \mathfrak{S}_n$.

To calculate the common unique stationary distribution of all these descent operator chains, using Theorem 3.3, first note that, for $\lambda \vdash n$, the coefficient of λ in $(1)^n$ is given by

$$|\{(\nu_2, \ldots, \nu_{n-1}) : \nu_2 \sim (1) \cup \square, \nu_3 \sim \nu_2 \cup \square, \ldots, \lambda \sim \nu_{n-1} \cup \square\}| = \dim \lambda.$$

To obtain $\pi(\lambda)$, multiply the above by $\frac{\eta(\lambda)}{n!}$. As $\eta(\lambda)$ is also $\dim \lambda$, this means $\pi(\lambda) = \frac{(\dim \lambda)^2}{n!}$, the Plancherel measure.

3.3 Quotient Algebras and a Lift to Tableaux

The following theorem is a specialisation of Theorem 2.7, about the strong lumping of Markov chains from linear maps, to the case of descent operator chains.

Theorem 3.6 (Strong lumping for descent operator chains) [53, Th. 4.1] *Let $\mathcal{H}$, $\bar{\mathcal{H}}$ be graded, connected Hopf algebras with bases $\mathcal{B}$, $\bar{\mathcal{B}}$, respectively, that both satisfy the conditions in Theorem 3.2. 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 Markov chain on $\mathcal{B}_n$ which the Doob transform fashions from the descent operator $m \circ \Delta_P$ lumps strongly via θ to the Doob transform chain from the same operator on $\bar{\mathcal{B}}_n$.*

Proof Condition 1 of Theorem 3.2 requires the distinct images of $\mathcal{B}_n$ under θ to be linearly independent—this is true here by hypothesis.

Condition 2 requires $\theta \circ (m \circ \Delta_P) = (m \circ \Delta_P) \circ \theta$, and $\bar{\eta} \circ \theta = \eta$. The former is true because θ is a Hopf-morphism. To check the latter, apply both sides to an arbitrary $x \in \mathcal{B}_n$ and multiply by $\bar{\bullet}^{\otimes n}$, where $\bar{\bullet}$ is the unique element of $\bar{\mathcal{B}}_1$; then the condition required is equivalent to

$$\bar{\eta}(\theta(x))\bar{\bullet}^{\otimes n} = \eta(x)\bar{\bullet}^{\otimes n}. \tag{3.1}$$

The left-hand side is $\Delta_{1,\ldots,1}(\theta(x))$, by definition of $\bar{\eta}$. Because θ is a Hopf-morphism, this is equal to $(\theta \otimes \cdots \otimes \theta) \circ \Delta_{1,\ldots,1}(x) = (\theta \otimes \cdots \otimes \theta)\left(\eta(x)\bullet^{\otimes n}\right)$, by definition of η (writing $\bullet$ for the unique element of $\mathcal{B}_1$). Hence this is $(\theta(\bullet) \otimes \cdots \otimes \theta(\bullet))\eta(x)$. Since $\theta(\mathcal{B}_n) = \bar{\mathcal{B}}_n$ for all n, it is true for $n = 1$, which means $\theta(\bullet) \in \bar{\mathcal{B}}_1$. Since $\bar{\mathcal{B}}_1 = \{\bar{\bullet}\}$, it must be that $\theta(\bullet) = \bar{\bullet}$. Hence $(\theta(\bullet) \otimes \cdots \otimes \theta(\bullet))\eta(x) = \eta(x)\bar{\bullet}^{\otimes n}$, proving Eq. 3.1. □

Remark Observe that the proof does not fully use the assumption $\theta(\mathscr{B}_n) = \bar{\mathscr{B}}_n$ for all n—all that is required is that $\theta(\mathscr{B}_n) = \bar{\mathscr{B}}_n$ for the single value of n of interest, and that $\theta(\mathscr{B}_1) = \bar{\mathscr{B}}_1$. Indeed, if $\mathscr{B}_n$ can be partitioned into communication classes $\mathscr{B}_n = \amalg_i \mathscr{B}_n^{(i)}$ for the $m \circ \Delta_P$ Markov chain (i.e. it is impossible to move between distinct $\mathscr{B}_n^{(i)}$ using the $m \circ \Delta_P$ Markov chain), so there is effectively a separate chain on each $\mathscr{B}_n^{(i)}$, then, to prove a lumping for the chain on one $\mathscr{B}_n^{(i)}$, it suffices to require $\theta(x) \in \bar{\mathscr{B}}_n$ only for $x \in \mathscr{B}_n^{(i)}$ (and $\theta(\mathscr{B}_1) = \bar{\mathscr{B}}_1$). This will be useful in Sect. 3.6 for showing that cut-and-interleave shuffles of n distinct cards lump via descent set, as this lumping is false for non-distinct decks.

3.3.1 Example: the Down–Up Chain on Standard Tableaux

To use Theorem 3.6 to lift the descent operator chains on partitions of the previous section, we need a Hopf algebra whose quotient is Λ, and the quotient map must come from a map from the basis elements of the new, larger Hopf algebra to partitions (or more accurately, to Schur functions). Below describes one such algebra, the Poirer–Reutenauer Hopf algebra of standard tableaux. It was christened $(\mathbb{Z}T, *, \delta)$ in [57], but we follow [27, Sec. 3.5] and denote it by **FSym**, for "free symmetric functions". Its distinguished basis is $\{\mathbf{S}_T\}$, where T runs over the set of standard tableaux. As with partitions, it will be convenient to write T in place of $\mathbf{S}_T$. This algebra is graded by the number of boxes in T. The quotient map $\mathbf{FSym} \to \Lambda$ is essentially taking the shape of the standard tableaux—the image of $\mathbf{S}_T$ in Λ is $s_{\text{shape}(T)}$.

Because the product and coproduct of **FSym** are fairly complicated, involving Jeu de Taquin and other tableaux manipulations, we describe here only $m : \mathscr{H}_{n-1} \otimes \mathscr{H}_1 \to \mathscr{H}_n$ and $\Delta_{n-1,1}$, and direct the interested reader to [57, Sec. 5c, 5d] for details.

If T is a standard tableaux with $n-1$ boxes, then the product $m(T \otimes \boxed{1})$ is the sum of all ways to add a new box, filled with n, to T. For example,

$$m\left(\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1} 4\\\cline{1-1}\end{array} \otimes \boxed{1}\right) = \begin{array}{|c|c|c|}\hline 1&2&5\\\hline 3\\\cline{1-1} 4\\\cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 1&2\\\hline 3&5\\\hline 4\\\cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1} 4\\\cline{1-1} 5\\\cline{1-1}\end{array} .$$

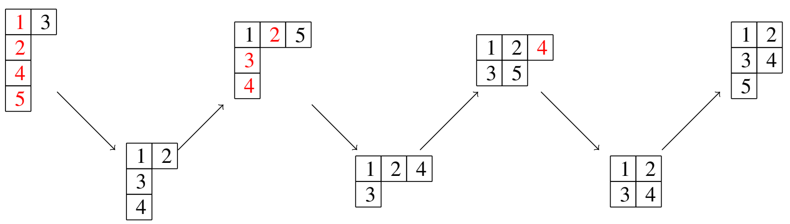
The coproduct $\Delta_{n-1,1}$ is "unbump and standardise". [64, fourth paragraph of proof of Th. 7.11.5] explains unbumping as follows: for a removable box b in row i, remove b, then find the box in row $i-1$ containing the largest integer smaller than b. Call this filling b_1. Replace b_1 with b, then put b_1 in the box in row $i-2$ previously filled with the largest integer smaller than b_1, and continue this process up the rows. In the second term in the example below, these displaced fillings are 4, 3, 2. What unbumping achieves is this: if b was the last number to be inserted in an RSK insertion that resulted in T, then unbumping b from T recovers the tableaux before b was inserted. The coproduct $\Delta_{n-1,1}(T)$ is the sum of unbumpings over all removable boxes b of T, then standardising the unbumped tableaux, for example

$$\begin{aligned}\Delta_{4,1}\left(\begin{array}{|c|c|c|}\hline 1&2&5\\\hline 3\\\cline{1-1} 4\\\cline{1-1}\end{array}\right) &= \mathrm{std}\left(\begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1} 4\\\cline{1-1}\end{array}\right)\otimes\begin{array}{|c|}\hline 1\\\hline\end{array} + \mathrm{std}\left(\begin{array}{|c|c|c|}\hline 1&3&5\\\hline 4\\\cline{1-1}\end{array}\right)\otimes\begin{array}{|c|}\hline 1\\\hline\end{array}\\ &= \begin{array}{|c|c|}\hline 1&2\\\hline 3\\\cline{1-1} 4\\\cline{1-1}\end{array}\otimes\begin{array}{|c|}\hline 1\\\hline\end{array} + \begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3\\\cline{1-1}\end{array}\otimes\begin{array}{|c|}\hline 1\\\hline\end{array}.\end{aligned}$$

To describe the down–up chain on standard tableaux (i.e. the chain which the Doob transform fashions from the map $\frac{1}{n}m\circ\Delta_{n-1,1}$), it remains to calculate the rescaling function $\eta(T)$. This is the coefficient of $\begin{array}{|c|}\hline 1\\\hline\end{array}^{\otimes n}$ in $\Delta_{1,\dots,1}(T)$, which the description of $\Delta_{n-1,1}$ above rephrases as the number of ways to successively choose boxes to unbump from T. Such ways are in bijection with the standard tableaux of the same shape as T, so $\eta(T) = \dim(\text{shape } T)$. Hence one step of the down–up chain on standard tableaux, starting from a tableau T of n boxes, has the following interpretation:

1. Pick a removable box b of T with probability $\frac{\dim(\text{shape}(T\backslash b))}{\dim(\text{shape } T)}$, and unbump b. (As for partitions, one can pick b using the hook walk of [37].)
2. Standardise the remaining tableaux and call this T'.
3. Add a box labelled n to T', with probability $\frac{1}{n}\frac{\dim(\text{shape}(T'\cup n))}{\dim(\text{shape } T')}$. (As for partitions, one can pick where to add this box using the complimentary hook walk of [38].)

Here are a few steps of a possible trajectory in degree 5 (the red marks the unbumping paths):



The transition matrix of this chain in degree 3 is

	1 2 3	1 2 / 3	1 3 / 2	1 / 2 / 3
1 2 3	$\frac{1}{3}$	$\frac{2}{3}$	0	0
1 2 / 3	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$
1 3 / 2	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$
1 / 2 / 3	0	0	$\frac{2}{3}$	$\frac{1}{3}$

According to Proposition 3.3, the unique stationary distribution of the down–up chain on tableau is $\pi(T) = \frac{1}{n!}\eta(T)\times$coefficient of T in $\boxed{1}^n$. Note that there is a unique way of adding outer boxes filled with $1, 2, \ldots$ in succession to build a given tableau T, so each tableau of n boxes appears precisely once in the product $\boxed{1}^n$. Hence $\pi(T) = \frac{1}{n!}\eta(T) = \frac{1}{n!}\dim(\text{shape } T)$.

As explained at the beginning of this subsection, the symmetric functions Λ is a quotient of **FSym** [57, Th. 4.3.i], and the quotient map sends $\mathbf{S}_T$ to $s_{\text{shape}(T)}$. Applying Theorem 3.6 then gives:

Theorem 3.7 *The down–up Markov chain on standard tableaux lumps to the down–up Markov chain on partitions via taking the shape.* □

By the same argument, the P-restriction-then-induction chains on partitions, for any probability distribution P, lift to $m \circ \Delta_P$ chains on tableaux, but these are hard to describe.

3.4 Subalgebras and a Lift to Permutations

The following theorem is a specialisation of Theorem 2.16, about the weak lumping of Markov chains from linear maps, to the case of descent operator chains.

Theorem 3.8 (Weak lumping for descent operator chains) *Let $\mathcal{H}$, $\mathcal{H}'$ be graded, connected Hopf algebras with state space bases $\mathcal{B}$, $\mathcal{B}'$, respectively, that both satisfy the conditions in Theorem 3.2. Suppose for all n that $\theta : \mathcal{B}_n \to \mathcal{B}'_n$ is such that the "preimage sum" map $\theta^* : \mathcal{B}'_n \to \mathcal{H}_n$, defined by $\theta^*(x') := \sum_{x\in\mathcal{B},\theta(x)=x'} x$, extends to a Hopf-morphism. Then the Markov chain on $\mathcal{B}_n$ which the Doob transform fashions from the descent operator $m \circ \Delta_P$ lumps weakly via θ to the Doob transform chain from the same map on $\mathcal{B}'_n$, from any starting distribution X_0 where $\frac{X_0(x)}{\eta(x)} = \frac{X_0(y)}{\eta(y)}$ whenever $\theta(x) = \theta(y)$.*

Proof First observe that θ^* sends $\mathscr{B}'_n$ to a linearly independent set in $\mathscr{H}_n$, so $\theta^* : \mathscr{H}' \to \mathscr{H}$ is injective, hence it is legal to identify $x' \in \mathscr{H}'$ with $\sum_{x\in\mathscr{B},\theta(x)=x'} x$ and view $\mathscr{H}'$ as a Hopf subalgebra of $\mathscr{H}$. So $\mathscr{H}'$ is an invariant subspace of $\mathscr{H}$ under $m \circ \Delta_P$. The other requirements of Theorem 2.16 are that each element in the basis $\mathscr{B}'_n$ should be a sum over disjoint subsets of $\mathscr{B}_n$, which is true by hypothesis; and that the restriction of the rescaling function $\eta : \mathscr{H}_n \to \mathbb{R}$ to $\mathscr{B}'_n$ is the natural rescaling function for $\mathscr{H}'_n$. The definition of η' is that, for all $x \in \mathscr{B}'_n$, it holds that $\Delta_{1,\dots,1}(x') = \eta'(x')\bullet'^{\otimes n}$, where $\bullet'$ is the unique element of $\mathscr{B}'_1$. Since θ sends $\mathscr{B}_1$ to $\mathscr{B}'_1$, it must be true that $\bullet' = \theta(\bullet)$, i.e. $\bullet = \theta^*(\bullet')$, or $\bullet = \bullet'$ when viewing $\mathscr{H}'$ as a subalgebra of $\mathscr{H}$. Hence $\Delta_{1,\dots,1}(x') = \eta'(x')\bullet^{\otimes n}$. The left-hand side is

$$\Delta_{1,\dots,1}\left(\sum_{x\in\mathscr{B},\theta(x)=x'} x\right) = \sum_{x\in\mathscr{B},\theta(x)=x'} \Delta_{1,\dots,1}(x) = \sum_{x\in\mathscr{B},\theta(x)=x'} \eta(x)\bullet^{\otimes n} = \eta(x')\bullet^{\otimes n},$$

where the second equality uses the definition of η, and the third equality uses the linearity of η. Hence $\eta'(x') = \eta(x')$. □

3.4.1 Example: the Down–Up Chain on Permutations

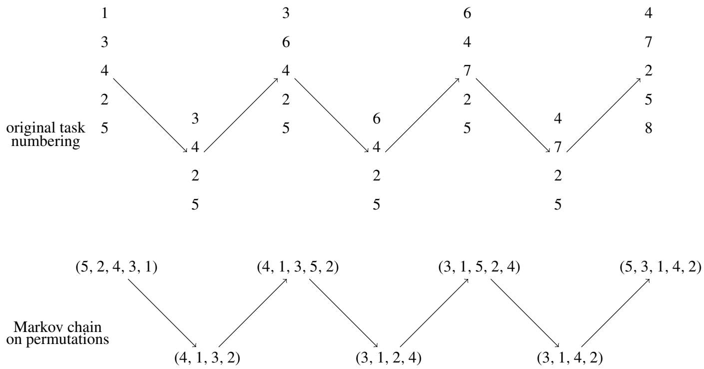
Recall that the previous section lifted the down–up chain on partitions to the down–up chain on tableaux, using that the symmetric functions Λ is a quotient of the Hopf algebra **FSym** of tableaux. To further lift this chain to permutations using Theorem 3.8, we express **FSym** as a subalgebra of a Hopf algebra of permutations, namely the Malvenuto–Reutenauer algebra.

Before a full description of this Hopf structure on permutations, here is an intuitive interpretation of its down–up chain, from $\frac{1}{n}m \circ \Delta_{n-1,1}$. (This is a mild variant of [54, Ex. 1.2], which is for $\frac{1}{n}m \circ \Delta_{1,n-1}$.) You keep an electronic to-do list of n tasks. Each day, there are two changes to the list: first, you complete the task at the top of the list and delete it; next, you receive a new task, which you add to the list in a position depending on its urgency (more urgent tasks are placed closer to the top). Assume that each incoming task is equally distributed in urgency relative to the $n-1$ tasks presently on the list, so each new task is inserted into the list in a uniform position.

To construct a Markov chain from this process, we assign a permutation (in one-line notation) to each to-do list. For the end-of-day lists, this will be a permutation of n; for mid-day lists (after the task-completion and before the new task arrives), this will be a permutation of $n-1$. This permutation assignment is rather complicated and requires three steps as follows:

1. write i for the task received on day i, so the to-do list becomes a vertical list of numbers;
2. standardise this list; now the numbers indicate the relative time that each task has spent on the list: 1 denotes the task that's been on the list for the longest time, 2 for the next oldest task, and so on, so n denotes the task you received today;
3. read these numbers from the bottom of the list to the top, so the bottommost number in the list appears leftmost in the permutation (in one-line notation).

The diagram below shows one possibility over days 5, 6, 7 and 8 (thus three transitions) of both the original task numbering before standardisation and the Markov chain on permutations from the standardised task numbering; each down arrow indicates the completion of the top task, and each up arrow the introduction of the new task, and these two parts together give one timestep of the Markov chain.



The same chain arises from performing the top-to-random shuffle and keeping track of the relative last times that the cards were last touched, instead of their values.

As mentioned above, this is the chain associated with $\frac{1}{n}m \circ \Delta_{n-1,1}$ on the Malvenuto-Reutenauer Hopf algebra of permutations, denoted $(\mathbb{Z}S, *, \Delta)$ in [50, Sec. 3], $(\mathbb{Z}S, *, \delta)$ in [57], and $\mathfrak{S}Sym$ in [4]. We follow the recent Parisian literature, such as [27], and call this algebra **FQSym**, for "free quasisymmetric functions".

The basis of concern here is the fundamental basis $\{\mathbf{F}_\sigma\}$, as σ ranges over all permutations (of any length). As in the previous sections, we often write $\mathbf{F}_\sigma$ simply as σ. The degree of σ is its length when considered as a word.

We explain the Hopf structure on **FQSym** by example. The product $\sigma_1\sigma_2$ is $\sigma_1 \sqcup\!\sqcup\, \sigma_2[\deg \sigma_1]$, the sum of all "interleavings" or "shuffles" of σ_1 with the shift of σ_2 by $\deg(\sigma_1)$:

$$\begin{aligned}(3,1,2)(2,1) &= (3,1,2) \sqcup\!\sqcup\, (5,4)\\ &= (3,1,2,5,4) + (3,1,5,2,4) + (3,1,5,4,2)\\ &\quad + (3,5,1,2,4) + (3,5,1,4,2)\\ &\quad + (3,5,4,1,2) + (5,3,1,2,4) + (5,3,1,4,2)\\ &\quad + (5,3,4,1,2) + (5,4,3,1,2).\end{aligned}$$

The coproduct is "deconcatenate and standardise":

$$\Delta(\sigma) = \sum_{\sigma_1 \cdot \sigma_2 = \sigma} \mathrm{std}(\sigma_1) \otimes \mathrm{std}(\sigma_2),$$

where $\cdot$ denotes concatenation. Thus

$$\Delta(4,1,3,2)$$

$$= () \otimes (4, 1, 3, 2) + \mathrm{std}(4) \otimes \mathrm{std}(1, 3, 2) + \mathrm{std}(4, 1) \otimes \mathrm{std}(3, 2)$$
$$+ \ \mathrm{std}(4, 1, 3) \otimes \mathrm{std}(2) + (4, 1, 3, 2) \otimes ()$$
$$= () \otimes (4, 1, 3, 2) + (1) \otimes (1, 3, 2) + (2, 1) \otimes (2, 1)$$
$$+ \ (3, 1, 2) \otimes (1) + (4, 1, 3, 2) \otimes ().$$

Recall that we are primarily interested in $\frac{1}{n} m \circ \Delta_{n-1,1}$. Note that $\Delta_{n-1,1}$ removes the last letter of the word and standardises the result, whilst right-multiplication by (1) yields the sum of all ways to insert the letter n. Since $\Delta_{n-1,1}(\sigma)$ contains only one term, we see inductively that the rescaling function is $\eta(\sigma) \equiv 1$. Hence one step of the down–up chain on **FQSym**, starting at $\sigma \in \mathfrak{S}_n$, has the following description:

1. Remove the last letter of σ.
2. Standardise the remaining word.
3. Insert the letter n into this standardised word, in a uniformly chosen position.

The transition matrix of this chain in degree 3 is (all empty entries are 0)

	(1, 2, 3)	(1, 3, 2)	(3, 1, 2)	(2, 3, 1)	(2, 1, 3)	(3, 2, 1)
(1, 2, 3)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$			
(1, 3, 2)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$			
(3, 1, 2)				$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
(2, 3, 1)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$			
(2, 1, 3)				$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
(3, 2, 1)				$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

According to Proposition 3.3, the unique stationary distribution of this chain is $\pi(\sigma) = \frac{1}{n!}\eta(\sigma) \times$coefficient of σ in $(1)^n$. Since there is a unique way of inserting the letters $1, 2, \ldots$ in that order to obtain a given permutation, each permutation of length n appears precisely once in the product $(1)^n$. Hence $\pi(\sigma) \equiv \frac{1}{n!}$.

Recall that the point of discussing this chain is that it is a weak lift of the down–up chain on tableaux of the previous section, if in the initial distribution any two permutations having the same RSK insertion tableau are equally probable (such permutations are said to belong to the same *plactic class*). Let RSK denote the map sending a permutation to its insertion tableau under the Robinson–Schensted–Knuth algorithm (this tableau is often called P). [57, Th. 4.3.iii] shows that **FSym** is a subalgebra of **FQSym** under the injection $\theta^*(\mathbf{S}_T) := \sum_{\mathrm{RSK}(\sigma)=T} \mathbf{F}_\sigma$, so Theorem 2.16 applies. (Note that the rescaling function $\eta(T) = \dim(\text{shape } T)$ is indeed the restriction of $\eta(\sigma) \equiv 1$, since the number of terms $\mathbf{F}_\sigma$ in the image of $\mathbf{S}_T$ is $\dim(\text{shape } T)$.) Thus

Theorem 3.9 *The down–up Markov chain on permutations lumps weakly to the down–up Markov chain on tableaux via taking RSK insertion tableau, whenever the initial distribution is constant on plactic classes.* □

This lumping can be "concatentated" with the lumping of Theorem 3.7 from tableaux to partitions. Thus the down–up chain on permutations lumps weakly to the down–up chain on partitions via taking the shape of the RSK insertion tableau, whenever the initial distribution is constant on plactic classes. By the same reasoning, this is true for chains from any descent operator $m \circ \Delta_P$. We call such chains on permutations the P-shuffles-with-standardisation.

Definition 3.10 Fix an integer n, and let $P(D)$ be a probability distribution on the weak-compositions of n. Each step of the *P-shuffle-with-standardisation* Markov chain on the permutations $\mathfrak{S}_n$ (viewed in one-line notation) goes as follows:

1. Choose a weak-composition D of n with probability $P(D)$.
2. Deconcatenate the current permutation into a word w_1 of the first d_1 letters, w_2 of the next d_2 letters, and so on.
3. Replace the smallest letter in w_1 by 1, the next smallest by 2, and so on. Then replace the smallest letter in w_2 by $d_1 + 1$, the next smallest by $d_1 + 2$, and so on for all w_i.
4. Interleave these words uniformly (i.e. uniformly choose a permutation where the letters $1, 2, \ldots, d_1$ are in the same relative order as in the replaced w_1, where $d_1 + 1, d_1 + 2, \ldots, d_1 + d_2$ are in the same relative order as in the replaced w_2, etc.).

As discussed in the previous four paragraphs and at the end of Sect. 3.3, the Hopf-morphisms **FSym** $\hookrightarrow$ **FQSym** and **FSym** $\twoheadrightarrow \Lambda$ prove the following:

Theorem 3.11 *Fix an integer n, and let $P(D)$ be a probability distribution on the weak-compositions of n. The P-shuffle-with-standardisation chain on the permutations $\mathfrak{S}_n$ lumps weakly to the P-restriction-then-induction chain on partitions, via taking the shape of the RSK insertion tableau, whenever the initial distribution is constant on plactic classes.* □

The next section deduces from this theorem a result of Fulman [31, Th. 3.1], that the probability of obtaining a partition λ after t steps of P-restriction-then-induction, starting from the partition with a single part, is the probability that the RSK shape of a deck is λ after t iterations, starting from the identity, of a P analogue of the top-to-random shuffle.

3.4.2 Other Lumpings of P-Shuffles-with-Standardisation, Strong and Weak

Return to the weak lumping of permutations to tableaux via RSK insertion (i.e. ignore the second lumping to partitions). The P-shuffle-with-standardisation chain has many weak lumpings in this style, thanks to the general construction in [40, Th. 31] [58] of Hopf subalgebras of **FQSym** spanned by $\sum_{\theta(\sigma)=T} \mathbf{F}_\sigma$, for various functions θ. The criteria on $\theta : \amalg_n \mathfrak{S}_n \to \mathscr{C}$ (the codomain $\mathscr{C}$ can be any graded set) is that its extension to all words with distinct letters, defined via $\theta(\sigma) = \theta(\mathrm{std}(\sigma))$, should be compatible with concatenation and alphabet restriction in the following sense:

1. if $\theta(\sigma_1) = \theta(\tau_1)$ and $\theta(\sigma_2) = \theta(\tau_2)$ then $\theta(\sigma_1 \cdot \sigma_2) = \theta(\tau_1 \cdot \tau_2)$;
2. if $\theta(\sigma) = \theta(\tau)$, and $\sigma_{\leftarrow r}$ (resp. $\sigma_{r\rightarrow}$) contains the letters $1, 2, \ldots, r$ (resp. $r + 1, \ldots, n$) in the same order as in σ, and similarly for $\tau_{\leftarrow r}$ and $\tau_{r\rightarrow}$, then $\theta(\sigma_{\leftarrow r}) = \theta(\tau_{\leftarrow r})$ and $\theta(\sigma_{r\rightarrow}) = \theta(\tau_{r\rightarrow})$.

Many such θ can be expressed in terms of insertion algorithms similar to RSK. For example, taking θ to be binary tree insertion [40, Algo. 17] generates the Loday–Ronco Hopf algebra [48]. Thus the P-shuffle-with-standardisation chain lumps weakly to a chain on binary trees, and [36] gives a variant with twin binary trees. Another example is the rising sequence lengths, also known as the recoil or idescent set: $\theta(\sigma) = \mathrm{Des}(\sigma^{-1})$, associated with the hypoplactic insertion of [45, Sec. 4.8]. Thus the P-shuffle-with-standardisation chain lumps weakly via the set of rising sequence lengths.

The dual of this general construction creates quotient algebras of the dual algebra **FQSym***, which is isomorphic to **FQSym** via inversion of permutations. These quotients satisfy the strong lumping criterion of Theorem 3.6, so the P-shuffles-with-standardisation lump (strongly) via $\sigma \mapsto \theta(\sigma^{-1})$ for any θ satisfying the conditions above. Examples of such strong lumping maps include RSK recording tableau, decreasing binary tree (the map λ of [5], the "recording" part of the binary tree algorithm), and descent set.

To see another example and non-example of lumpings induced from Hopf-morphisms, consider the following commutative diagram from [57, Th. 4.3]:

$$
\begin{array}{ccc}
\mathbf{FQSym}\text{, permutations} & \xrightarrow{\quad \mathrm{Des} \quad} & QSym\text{, compositions} \\
\uparrow & & \uparrow \\
\mathbf{FSym}\text{, standard tableaux} & \xrightarrow{\quad \mathrm{shape} \quad} & \Lambda\text{, partitions}
\end{array}
$$

The main example in Sects. 3.2, 3.3, 3.4 lifts a chain on partitions to a chain on permutations via a chain on standard tableaux, on the bottom left. Let us see why it is not possible to construct a lift via compositions, on the top right, instead. The corresponding Hopf algebra here is the algebra of quasisymmetric functions [35].

There is no problem with the top Hopf-morphism, which sends a permutation $\mathbf{F}_\sigma$ to the fundamental quasisymmetric function $F_{\mathrm{Des}(\sigma)}$ associated with its descent composition. Since this map sends a basis of **FQSym** to a basis of $QSym$, Theorem 3.6 applies and P-shuffles-with-standardisation lump strongly via descent composition.

The problem is with the Hopf-morphism on the right - this inclusion is not induced from a set map from compositions to partitions. Theorem 3.8 only applies if $s_\lambda = \sum_{\theta(I)=\lambda} F_I$ for some function θ sending compositions to partitions. This condition does not hold, as the same F_I can occur in the expansion of multiple Schur functions: $s_{(3,1)} = F_{(1,3)} + F_{(2,2)} + F_{(3,1)}$, $s_{(2,2)} = F_{(1,2,1)} + F_{(2,2)}$.

3.5 Equidistribution of Shuffles from the Identity, with and without standardisation

In this section, we relate the P-shuffles-with-standardisation of Definition 3.10, to the "cut-and-interleave" shuffles of [19], which we call here *P-shuffles-without-standardisation* for clarity.

Definition 3.12 Fix an integer n, and let $P(D)$ be a probability distribution on the weak-compositions of n. Each step of the *P-shuffle-without-standardisation* Markov chain on the permutations $\mathfrak{S}_n$ (viewed in one-line notation) goes as follows:

1. Choose a weak-composition D of n with probability $P(D)$.
2. Deconcatenate the current permutation into a word w_1 of the first d_1 letters, w_2 of the next d_2 letters, and so on.
3. Interleave these words uniformly, i.e. uniformly choose a permutation where the letters of each w_i stay in the same relative order.

This chain comes from the descent operators $m \circ \Delta_P$ applied to the shuffle algebra of Ree [53].

In [31], Fulman showed the following:

Theorem 3.13 [31, Th. 3.1] *The probability of obtaining any partition as the RSK shape after P-shuffling-without-standardisation t times, starting from the identity permutation, is equal to its probability under t steps of P-restriction-then-induction, started at the trivial representation (i.e. the partition with a single row).*

Note that the result is only about the probabilities after all t steps of the chains; nothing can be deduced about the probabilities at intermediate times. In particular, it does not assert that P-shuffles-without-standardisation lumps, strongly or weakly, to P-restriction-then-induction; this is false, see the discussion in the second paragraph of the introduction to Sect. 3.

Fulman remarked that this connection is "surprising" and "quite mysterious", and perhaps a more enlightening proof is to combine Theorem 3.11 with the following.

Theorem 3.14 *The distribution on permutations after t iterates of P-shuffles-with-standardisation is the same as that after t iterates of P-shuffles-without-standardisation, if both are started from the identity permutation.*

The power of this theorem goes much beyond reproving Fulman's "almost lift": it allows results about either type of P-shuffle to apply to the other type. For example, [22, Ex. 5.8] showed that the expected number of descents after t riffle-shuffles of n cards, starting at the identity, is $\left(1-2^{-t}\right)\frac{n-1}{2}$, so this must also be the expected number of descents after t riffle-shuffles-with-standardisation starting at the identity. In the other direction, [54, Sec. 6] made the simple observation that, if one tracks only the relative orders of the bottom k cards under top-to-random-with-standardisation, then one sees a lazy version of top-to-random-with-standardisation on k cards, lazy meaning that at each time step no move is made with probability $\frac{n-k}{n}$. (This is a strong lumping, but not from Hopf algebras.) Thus the distribution of the relative orders of the bottom k cards after t iterates of top-to-random-without-standardisation from the identity is the distribution after t iterates of a lazy version of top-to-random-without-standardisation on k cards.

Another use of Theorem 3.14 is to obtain many analogues of Theorem 3.13, with various statistics in place of the RSK shape. This is because P-shuffles-with-standardisation is associated with **FQSym**, which has many subquotients as noted in Sect. 3.4.2, and hence has many strong and weak lumpings. Thus the probability distribution of the binary search tree, the decreasing tree, the rising sequence lengths, the descent set, and other statistics after t iterates of P-shuffles-without-standardisation can all be calculated from $m \circ \Delta_P$ Markov chains on the statistics themselves. (As Sect. 3.6 will explain, the descent set is actually a Markov statistic of P-shuffling-without-standardisation.)

Proof of Theorem 3.14 The case of $t = 1$ is clear, since performing a single P-shuffle-with-standardisation, starting from the identity permutation, does not actually require any standardisation. The key to showing this result for larger t is to express t iterates of a P-shuffle, with- or without-standardisation, as the same P''-shuffle, performed only once. This uses the (vector space) isomorphism identifying the descent operator $m \circ \Delta_D$ with the homogeneous non-commutative symmetric function S^D [54, Sec. 2.5]. (See [34] for background on non-commutative symmetric functions.) Write S^P for the non-commutative symmetric function associated with $m \circ \Delta_P$.

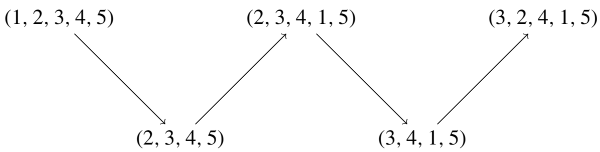
On a commutative Hopf algebra, such as the shuffle algebra, the composition $(m \circ \Delta_P) \circ (m \circ \Delta_{P'})$ corresponds to the internal product of non-commutative symmetric functions $S^{P'} S^P$ [55, Th. II.7]. So t iterates of a P-shuffle-without-standardisation correspond to $\left(S^P\right)^t$.

Now consider the P-shuffles-with-standardisation, on the algebra **FQSym**. Note that the coproduct of the identity permutation is

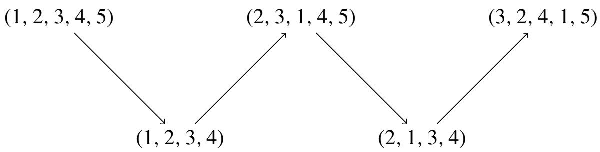
$$\Delta(1, \ldots, n) = \sum_{r=0}^{n} (1, \ldots, r) \otimes (1, \ldots, n-r), \tag{3.2}$$

and each tensor-factor is an identity permutation of shorter length. Thus the subalgebra of **FQSym** generated by identity permutations of varying length is closed under coproduct, and P-shuffling-with-standardisation from the identity stays within this sub-Hopf-algebra. Equation 3.2 shows that this sub-Hopf-algebra is cocommutative, so a composition of descent operators $(m \circ \Delta_P) \circ (m \circ \Delta_{P'})$ corresponds to the internal product of non-commutative symmetric functions $S^P S^{P'}$ [55, Th. II.7]. Despite this product being in the opposite order from the shuffle algebra case, t iterates of P-shuffles-with-standardisation are also described by $\left(S^P\right)^t$. □

It would be interesting to find a bijective proof of this equidistribution after t steps, i.e. to find a bijection between trajectories, starting at the identity, under P-shuffling-with- and P-shuffling-without-standardisation that have the same endpoint. Since the products of the associated non-commutative symmetric functions are in opposite orders for the two chains, such a bijection should probably be "order reversing" in some way. For example, the trajectory below of top-to-random-without-standardisation comes from inserting first in position 4, and then in position 2.



A possible trajectory of top-to-random-with-standardisation that has the same endpoint is to insert first in position $2 + 1 = 3$ and then in position 4:



Here, the "order reversal" makes intuitive sense, because when there is standardisation, 1 marks the most recently moved card, whereas in the shuffles-without-standardisation 1 is the first card moved.

To close this section, here are some similarities and differences between the P-shuffles with- and P-shuffles-without- standardisation. Since both are descent operator chains, by [54, Th. 3.5] they have the same eigenvalues, but different multiplicities. (Strictly speaking, the case without standardisation requires a multigraded version of this theorem.) For example, the eigenvalues for both top-to-random chains are $\frac{j}{n}$ for $j = 0, 1, 2, \ldots, n-3, n-2, n$. Their multiplicities for the without-standardisation chain are the number of permutations with j fixed points [19, Th. 4.1] [41, Sec. 4.6], whilst, for the chains with standardisation, they are the number of permutations fixing $1, 2, \ldots, j$ but not $j + 1$ (see Theorem 3.4 above). Hence in general the smaller eigenvalues have higher multiplicities in the chain with standardisation. The P-shuffles-without-standardisation are diagonalisable (because the shuffle algebra is commutative), and so are the top-to-random-with-standardisation and bottom-to-random-with-standardisation (because of a connection with dual graded graphs, see the Remark in [54, Sec. 4.2]), but Sage computer calculations show that the P-shuffles-with-standardisation are generally non-diagonalisable.

3.6 Lumping Card Shuffles by Descent Set

This section addresses an example separate from the long example of the previous sections. The goal is to apply a mild adaptation of Theorem 3.6 to reprove the following theorem of Athanasiadis and Diaconis:

Theorem 3.15 [9, Ex. 5.8] *The P-shuffles (without standardisation) lump via descent set.*

A weaker version of this result, for riffle-shuffles only, was announced in [51], along with eigenvectors of the lumped chain on compositions. The proof below is reproduced from the thesis [52, Sec. 6.3].

Recall from Example 3.1 that the P-shuffles (without standardisation) come from the descent operators $m \circ \Delta_P$ on the shuffle algebra $\mathscr{S}$. Thus, to prove Theorem 3.15, it suffices to construct a quotient Hopf algebra of $\mathscr{S}$ such that, for each word $w \in \mathscr{S}$ with distinct letters, the quotient map θ sends w to a basis element of the quotient algebra indexed by $\mathrm{Des}(w)$. This quotient Hopf algebra is $QSym$, the quasisymmetric functions of [35]. (Although $QSym$ is a well-known Hopf algebra, the quotient map θ is highly non-standard, in contrast to the Hopf-morphisms of previous sections.) The images under θ of the words with distinct letters will be the fundamental basis $\{F_I\}$, but the proof below will also require the monomial basis $\{M_I\}$. Since the product and coproduct of $QSym$ are fairly complicated, we omit the details here, and refer the interested reader to [35].

Theorem 3.16 [52, Th. 6.2.1] *There is a morphism of Hopf algebras $\theta : \mathscr{S} \to QSym$ such that, if w is a word with distinct letters, then $\theta(w) = F_{\mathrm{Des}(w)}$.*

Proof By [1, Th. 4.1], $QSym$ is the terminal object in the category of combinatorial Hopf algebras equipped with a multiplicative character. So, to define any Hopf-morphism to $QSym$, it suffices to define the corresponding character ζ on the domain. By [59, Th. 6.1.i], the shuffle algebra $\mathscr{S}$ is freely generated by *Lyndon words* [49, Sec. 5.1], which are strictly smaller than all their cyclic rearrangements. Hence any choice of the values of ζ on Lyndon words extends uniquely to a well-defined character on $\mathscr{S}$. For Lyndon u, set

$$\zeta(u) = \begin{cases} 1 & \text{if } u \text{ has all letters distinct and has no descents;} \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

We claim that, consequently, (3.3) holds for all words with distinct letters, even if they are not Lyndon. Assuming this for now, [1, Th. 4.1] defines

$$\begin{aligned} \theta(w) &= \sum_{I \vdash n} (\zeta \otimes \cdots \otimes \zeta)\,(\Delta_I(w))\, M_I \\ &= \sum_{I \vdash n} \zeta((w_1, \ldots, w_{i_1}))\zeta((w_{i_1+1}, \ldots, w_{i_1+i_2})) \ldots \zeta((w_{i_{l(I)-1}+1}, \ldots, w_n)) M_I. \end{aligned}$$

If w has distinct letters, then every consecutive subword $(w_{i_1+\cdots+i_j+1}, \ldots, w_{i_1 \ldots +i_{j+1}})$ of w also has distinct letters, so

$$\zeta((w_1, \ldots, w_{i_1})) \ldots \zeta((w_{i_{l(I)-1}+1}, \ldots, w_n)) = \begin{cases} 1 & \text{if } \mathrm{Des}(w) \le I; \\ 0 & \text{otherwise.} \end{cases}$$

Hence $\theta(w) = \sum_{\mathrm{Des}(w) \le I} M_I = F_{\mathrm{Des}(w)}$.

Now return to proving the claim that (3.3) holds whenever w has distinct letters. Proceed by induction on w, with respect to lexicographic order. [59, Th. 6.1.ii], applied to a word w with distinct letters, states that: if w has Lyndon factorisation $w = u_1 \cdots u_k$, then the product of these factors in the shuffle algebra satisfies

$$u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k = w + \sum_{v<w} \alpha_v v$$

where α_v is 0 or 1. The character ζ is multiplicative, so

$$\zeta(u_1) \ldots \zeta(u_k) = \zeta(w) + \sum_{v<w} \alpha_v \zeta(v). \tag{3.4}$$

If w is Lyndon, then the claim is true by definition; this includes the base case for the induction. Otherwise, $k > 1$ and there are two possibilities:

- None of the u_is have descents. Then the left-hand side of (3.4) is 1. Since the u_is together have all letters distinct, the only way to shuffle them together and obtain a word with no descents is to arrange the constituent letters in increasing order. This word is Lyndon, so it is not w, and, by inductive hypothesis, it is the only v in the sum with $\zeta(v) = 1$. So $\zeta(w)$ must be 0.
- Some Lyndon factor u_i has at least one descent. Then $\zeta(u_i) = 0$, so the left-hand side of (3.4) is 0. Also, no shuffle of $u_1, \ldots, u_k$ has its letters in increasing order. Therefore, by inductive hypothesis, all v in the sum on the right-hand side have $\zeta(v) = 0$. Hence $\zeta(w) = 0$ also.

□

Remark From the proof, one sees that the conclusion $\theta(w) = F_{\mathrm{Des}(w)}$ for w with distinct letters relies only on the value of ζ on Lyndon words with distinct letters. The proof took $\zeta(u) = 0$ for all Lyndon u with repeated letters, but any other value would also work. Alas, no definition of ζ will ensure that the images of all words are F_I for some I:

$$\theta((1,1)) = \frac{1}{2}\theta((1)\sqcup\!\sqcup(1)) = \frac{1}{2}\theta(1)\theta(1) = \frac{1}{2}M_{(1)}^2 = F_{(1,1)} + F_{(2)}.$$

Acknowledgements I would like to thank Nathan Williams for a question that motivated this research, and Persi Diaconis, Jason Fulman and Franco Saliola for numerous helpful conversations, and Federico Ardila, Grégory Châtel, Mathieu Guay-Paquet, Simon Rubenstein-Salzedo, Yannic Vargas and Graham White for useful comments. SAGE computer software [65] was very useful, especially the combinatorial Hopf algebras coded by Aaron Lauve and Franco Saliola.

References

1. Aguiar, M., Bergeron, N., Sottile, F.: Combinatorial Hopf algebras and generalized Dehn–Sommerville relations. Compos. Math. **142**(1), 1–30 (2006)
2. Aguiar, M., Mahajan, S.: Monoidal functors, species and Hopf algebras, *CRM Monograph Series*, vol. 29. American Mathematical Society, Providence, RI (2010). With forewords by Kenneth Brown and Stephen Chase and André Joyal
3. Aguiar, M., Mahajan, S.: Hopf monoids in the category of species. In: Hopf Algebras and Tensor categories. Contemporary Mathematics, vol. 585, pp. 17–124 (2013). https://doi.org/10.1090/conm/585/11665
4. Aguiar, M., Sottile, F.: Structure of the Malvenuto–Reutenauer Hopf algebra of permutations. Adv. Math. **191**(2), 225–275 (2005). https://doi.org/10.1016/j.aim.2004.03.007
5. Aguiar, M., Sottile, F.: Structure of the Loday–Ronco Hopf algebra of trees. J. Algebra **295**(2), 473–511 (2006). https://doi.org/10.1016/j.jalgebra.2005.06.021
6. Aldous, D., Diaconis, P.: Shuffling cards and stopping times. Am. Math. Mon. **93**(5), 333–348 (1986). https://doi.org/10.2307/2323590
7. Aldous, D., Diaconis, P.: Strong uniform times and finite random walks. Adv. Appl. Math. **8**(1), 69–97 (1987). https://doi.org/10.1016/0196-8858(87)90006-6
8. Assaf, S., Diaconis, P., Soundararajan, K.: A rule of thumb for riffle shuffling. Ann. Appl. Probab. **21**(3), 843–875 (2011). https://doi.org/10.1214/10-AAP701
9. Athanasiadis, C.A., Diaconis, P.: Functions of random walks on hyperplane arrangements. Adv. Appl. Math. **45**(3), 410–437 (2010). https://doi.org/10.1016/j.aam.2010.02.001

10. Ayyer, A., Schilling, A., Steinberg, B., Thiéry, N.M.: Markov chains, $\mathscr{R}$-trivial monoids and representation theory. Int. J. Algebra Comput. **25**(1–2), 169–231 (2015). https://doi.org/10.1142/S0218196715400081
11. Bayer, D., Diaconis, P.: Trailing the dovetail shuffle to its lair. Ann. Appl. Probab. **2**(2), 294–313 (1992)
12. Bergeron, N., Li, H.: Algebraic structures on Grothendieck groups of a tower of algebras. J. Algebra **321**(8), 2068–2084 (2009). https://doi.org/10.1016/j.jalgebra.2008.12.005
13. Bernstein, M.: A random walk on the symmetric group generated by random involutions. ArXiv e-prints (2016)
14. Bidigare, P., Hanlon, P., Rockmore, D.: A combinatorial description of the spectrum for the Tsetlin library and its generalization to hyperplane arrangements. Duke Math. J. **99**(1), 135–174 (1999). https://doi.org/10.1215/S0012-7094-99-09906-4
15. Britnell, J.R., Wildon, M.: Bell numbers, partition moves and the eigenvalues of the random-to-top shuffle in Dynkin types A, B and D. J. Combin. Theory Ser. A **148**, 116–144 (2017). https://doi.org/10.1016/j.jcta.2016.12.003
16. Brown, K.S.: Semigroups, rings, and Markov chains. J. Theor. Probab. **13**(3), 871–938 (2000). https://doi.org/10.1023/A:1007822931408
17. Corteel, S., Williams, L.K.: A Markov chain on permutations which projects to the PASEP. Int. Math. Res. Not. IMRN (17), Art. ID rnm055, 27 (2007). https://doi.org/10.1093/imrn/rnm055
18. Diaconis, P.: Group representations in probability and statistics. In: Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 11. Institute of Mathematical Statistics, Hayward, CA (1988)
19. Diaconis, P., Fill, J.A., Pitman, J.: Analysis of top to random shuffles. Combin. Probab. Comput. **1**(2), 135–155 (1992). https://doi.org/10.1017/S0963548300000158
20. Diaconis, P., Holmes, S.P.: Random walks on trees and matchings. Electron. J. Probab. **7**(6), 17 (2002). https://doi.org/10.1214/EJP.v7-105
21. Diaconis, P., Mayer-Wolf, E., Zeitouni, O., Zerner, M.P.W.: The Poisson–Dirichlet law is the unique invariant distribution for uniform split-merge transformations. Ann. Probab. **32**(1B), 915–938 (2004). https://doi.org/10.1214/aop/1079021468
22. Diaconis, P., Pang, C.Y.A., Ram, A.: Hopf algebras and Markov chains: two examples and a theory. J. Algebraic Combin. **39**(3), 527–585 (2014). https://doi.org/10.1007/s10801-013-0456-7
23. Diaconis, P., Ram, A.: Analysis of systematic scan Metropolis algorithms using Iwahori–Hecke algebra techniques. Michigan Math. J. **48**, 157–190 (2000). https://doi.org/10.1307/mmj/1030132713. Dedicated to William Fulton on the occasion of his 60th birthday
24. Diaconis, P., Shahshahani, M.: Generating a random permutation with random transpositions. Z. Wahrsch. Verw. Gebiete **57**(2), 159–179 (1981). https://doi.org/10.1007/BF00535487
25. Diaconis, P., Shahshahani, M.: Time to reach stationarity in the Bernoulli–Laplace diffusion model. SIAM J. Math. Anal. **18**(1), 208–218 (1987). https://doi.org/10.1137/0518016
26. Doob, J.L.: Conditional Brownian motion and the boundary limits of harmonic functions. Bull. Soc. Math. France **85**, 431–458 (1957)
27. Duchamp, G., Hivert, F., Thibon, J.Y.: Noncommutative symmetric functions. VI. Free quasi-symmetric functions and related algebras. Int. J. Algebra Comput. **12**(5), 671–717 (2002). https://doi.org/10.1142/S0218196702001139
28. Duchi, E., Schaeffer, G.: A combinatorial approach to jumping particles. J. Combin. Theory Ser. A **110**(1), 1–29 (2005). https://doi.org/10.1016/j.jcta.2004.09.006
29. Durrett, R., Granovsky, B.L., Gueron, S.: The equilibrium behavior of reversible coagulation–fragmentation processes. J. Theor. Probab. **12**(2), 447–474 (1999). https://doi.org/10.1023/A:1021682212351
30. Fill, J.A.: An exact formula for the move-to-front rule for self-organizing lists. J. Theor. Probab. **9**(1), 113–160 (1996). https://doi.org/10.1007/BF02213737
31. Fulman, J.: Card shuffling and the decomposition of tensor products. Pac. J. Math. **217**(2), 247–262 (2004). https://doi.org/10.2140/pjm.2004.217.247
32. Fulman, J.: Commutation relations and Markov chains. Probab. Theory Relat. Fields **144**(1–2), 99–136 (2009). https://doi.org/10.1007/s00440-008-0143-0
33. Fulton, W.: Young tableaux. In: London Mathematical Society Student Texts, vol. 35. Cambridge University Press, Cambridge (1997). With applications to representation theory and geometry
34. Gelfand, I.M., Krob, D., Lascoux, A., Leclerc, B., Retakh, V.S., Thibon, J.Y.: Noncommutative symmetric functions. Adv. Math. **112**(2), 218–348 (1995). https://doi.org/10.1006/aima.1995.1032

35. Gessel, I.M.: Multipartite P-partitions and inner products of skew Schur functions. In: Combinatorics and algebra (Boulder, Colo., 1983), *Contemp. Math.*, vol. 34, pp. 289–317. Amer. Math. Soc., Providence, RI (1984). https://doi.org/10.1090/conm/034/777705
36. Giraudo, S.: Algebraic and combinatorial structures on pairs of twin binary trees. J. Algebra **360**, 115–157 (2012). https://doi.org/10.1016/j.jalgebra.2012.03.020
37. Greene, C., Nijenhuis, A., Wilf, H.S.: A probabilistic proof of a formula for the number of Young tableaux of a given shape. Adv. Math. **31**(1), 104–109 (1979). https://doi.org/10.1016/0001-8708(79)90023-9
38. Greene, C., Nijenhuis, A., Wilf, H.S.: Another probabilistic method in the theory of Young tableaux. J. Combin. Theory Ser. A **37**(2), 127–135 (1984). https://doi.org/10.1016/0097-3165(84)90065-7
39. Grinberg, D., Reiner, V.: Hopf algebras in combinatorics. ArXiv e-prints (2014)
40. Hivert, F.: An introduction to combinatorial Hopf algebras—examples and realizations. In: Physics and theoretical computer science, *NATO Secur. Sci. Ser. D Inf. Commun. Secur.*, vol. 7, pp. 253–274. IOS, Amsterdam (2007)
41. Hivert, F., Luque, J.G., Novelli, J.C., Thibon, J.Y.: The $(1 - \mathbb{E})$-transform in combinatorial Hopf algebras. J. Algebraic Combin. **33**(2), 277–312 (2011). https://doi.org/10.1007/s10801-010-0245-5
42. Joni, S.A., Rota, G.C.: Coalgebras and bialgebras in combinatorics. Stud. Appl. Math. **61**(2), 93–139 (1979)
43. Kemeny, J.G., Snell, J.L.: Finite Markov Chains. The University Series in Undergraduate Mathematics. D. Van Nostrand Co., Inc., Princeton (1960)
44. Kemeny, J.G., Snell, J.L., Knapp, A.W.: Denumerable Markov Chains. D. Van Nostrand Co., Inc., Princeton (1966)
45. Krob, D., Thibon, J.Y.: Noncommutative symmetric functions. IV. Quantum linear groups and Hecke algebras at $q = 0$. J. Algebraic Combin. **6**(4), 339–376 (1997). https://doi.org/10.1023/A:1008673127310
46. Letac, G., Takács, L.: Random walks on an m-dimensional cube. J. Reine Angew. Math. **310**, 187–195 (1979)
47. Levin, D.A., Peres, Y., Wilmer, E.L.: Markov chains and mixing times. American Mathematical Society, Providence, RI With a Chapter by James G. Propp and David B, Wilson (2009)
48. Loday, J.L., Ronco, M.O.: Hopf algebra of the planar binary trees. Adv. Math. **139**(2), 293–309 (1998). https://doi.org/10.1006/aima.1998.1759
49. Lothaire, M.: Combinatorics on Words. Cambridge Mathematical Library. Cambridge University Press, Cambridge (1997). With a foreword by Roger Lyndon and a preface by Dominique Perrin; corrected reprint of the 1983 original, with a new preface by Perrin
50. Malvenuto, C., Reutenauer, C.: Duality between quasi-symmetric functions and the Solomon descent algebra. J. Algebra **177**(3), 967–982 (1995). https://doi.org/10.1006/jabr.1995.1336
51. Pang, C.Y.A.: 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, pp. 499–510. Assoc. Discrete Math. Theor. Comput. Sci., Nancy (2013)
52. Pang, C.Y.A.: Hopf algebras and Markov chains. ArXiv e-prints (2014). A revised thesis
53. Pang, C.Y.A.: Card-shuffling via convolutions of projections on combinatorial Hopf algebras. In: 27th International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2015), Discrete Math. Theor. Comput. Sci. Proc., AU, pp. 49–60. Assoc. Discrete Math. Theor. Comput. Sci., Nancy (2015)
54. Pang, C.Y.A.: Markov chains from descent operators on combinatorial Hopf algebras. ArXiv e-prints (2016). References are to a second version in preparation
55. Patras, F.: L'algèbre des descentes d'une bigèbre graduée. J. Algebra **170**(2), 547–566 (1994). https://doi.org/10.1006/jabr.1994.1352
56. Phatarfod, R.M.: On the matrix occurring in a linear search problem. J. Appl. Probab. **28**(2), 336–346 (1991)
57. Poirier, S., Reutenauer, C.: Algèbres de Hopf de tableaux. Ann. Sci. Math. Québec **19**(1), 79–90 (1995)
58. Priez, J.B.: A lattice of combinatorial Hopf algebras, application to binary trees with multiplicities. ArXiv e-prints (2013)
59. Reutenauer, C.: Free Lie Algebras, *London Mathematical Society Monographs. New Series*, vol. 7. The Clarendon Press Oxford University Press, New York (1993). Oxford Science Publications
60. Rogers, L.C.G., Pitman, J.W.: Markov functions. Ann. Probab. **9**(4), 573–582 (1981)

61. Sagan, B.E.: The symmetric group, *Graduate Texts in Mathematics*, vol. 203, second edn. Springer-Verlag, New York (2001). https://doi.org/10.1007/978-1-4757-6804-6. Representations, combinatorial algorithms, and symmetric functions
62. Saloff-Coste, L.: Random walks on finite groups. In: Kesten, H. (ed.) Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, pp. 263–346. Springer, Berlin (2004). https://doi.org/10.1007/978-3-662-09444-0_5
63. Solomon, L.: A Mackey formula in the group ring of a Coxeter group. J. Algebra **41**(2), 255–264 (1976)
64. Stanley, R.P.: Enumerative Combinatorics. Vol. 2, Cambridge Studies in Advanced Mathematics, vol. 62. Cambridge University Press, Cambridge (1999). With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin
65. Stein, W., et al.: Sage Mathematics Software (Version 6.6). The Sage Development Team (2015). http://www.sagemath.org
66. Swart, J.: Advanced topics in Markov chains (2012). http://staff.utia.cas.cz/swart/chain10.pdf. Lecture notes from a course at Charles University
67. Tsetlin, M.L.: Finite automata and models of simple forms of behaviour. Russ. Math. Surv. **18**(4), 1–28 (1963)
68. Zhao, Y.: Biased riffle shuffles, quasisymmetric functions, and the RSK algorithm (2009). http://yufeizhao.com/papers/shuffling.pdf
69. Zhou, H.: Examples of multivariate Markov chains with orthogonal polynomial eigenfunctions. ProQuest LLC, Ann Arbor, MI (2008). Thesis (Ph.D.)–Stanford University