Journal of Algebra 655 (2024) 139–162

Contents lists available at ScienceDirect

Journal of Algebra

journal homepage: www.elsevier.com/locate/jalgebra

Enumerative theory for the Tsetlin library

Sourav Chatterjee, Persi Diaconis, Gene B. Kim *

Departments of Mathematics and Statistics, Stanford University, United States of America

ARTICLE INFO

Article history:
Received 28 June 2023
Available online 21 August 2023
Communicated by Alberto Elduque

In memory of Georgia Benkart

Keywords:
Tsetlin library
Luce model
Markov chain

ABSTRACT

The Tsetlin library is a well-studied Markov chain on the symmetric group S_n. It has stationary distribution $\pi(\sigma)$ the Luce model, a nonuniform distribution on S_n, which appears in psychology, horse race betting, and tournament poker. Simple enumerative questions, such as "what is the distribution of the top k cards?" or "what is the distribution of the bottom k cards?" are long open. We settle these questions and draw attention to a host of parallel questions on the extension to the chambers of a hyperplane arrangement.

1. Introduction

Let $\theta_1, \theta_2, \ldots, \theta_n$ be positive real numbers. The **Luce model** $\pi(\sigma)$ is a probability distribution on the symmetric group S_n driven by these weights. In words, "put n balls with weights $\theta_1, \theta_2, \ldots, \theta_n$ into an urn. Each time, withdraw a ball from the urn (sampling without replacement) with probability proportional to its weight (relative to the remaining balls)." Thus, if $w_n = \theta_1 + \cdots + \theta_n$,

* Corresponding author.
E-mail addresses: souravc@stanford.edu (S. Chatterjee), diaconis@math.stanford.edu (P. Diaconis), genebkim@stanford.edu (G.B. Kim).

https://doi.org/10.1016/j.jalgebra.2023.08.009
0021-8693/

$$\pi(\sigma) = \frac{\theta_{\sigma(1)}}{w_n} \frac{\theta_{\sigma(2)}}{w_n - \theta_{\sigma(1)}} \frac{\theta_{\sigma(3)}}{w_n - \theta_{\sigma(1)} - \theta_{\sigma(2)}} \cdots \frac{\theta_{\sigma(n)}}{\theta_{\sigma(n)}}. \tag{1}$$

In Section 2, a host of applied problems are shown to give rise to the Luce model. These include the celebrated Tsetlin library, a Markov chain on S_n, described as: "at each step, choose a card labeled i with probability proportional to θ_i and move it to the top" – $\pi(\sigma)$ is the stationary distribution of this Markov chain.

It is natural to ask basic enumerative questions: pick σ from $\pi(\sigma)$.

- What is the distribution of the number of fixed points, cycles, longest cycle, and order of σ?
- What is the distribution of the length of the longest increasing subsequence of σ?

Most of these questions are open at this time. The main results below settle

- What is the distribution of the top k cards $\sigma(1), \ldots, \sigma(k)$? (Section 3)
- What is the distribution of the bottom k cards $\sigma(n-k+1), \ldots, \sigma(n)$? (Section 4)

Weighted sampling without replacement from a finite population is a standard topic. This may be accessed from the Wikipedia entries of "Horvitz–Thompson estimator" [35] and "concomitant order statistics."

Coming closer to combinatorics, two papers by Rosen [44] treat the coupon collector's problem and other coverage problems.

The recent paper by Ben-Hamou, Peres and Salez [6] couples sampling with and without replacement so that tail and concentration bounds, derived for partial sums when sampling with replacement, are seen to apply "as is" to sampling without replacement.

In computer science applications, Tsetlin library is studied as the "move to the front rule." In these applications, interest centers on the search cost. Here, the search cost for item i at time t is the number of cards j above card labeled i at time t. Summing over i, say with weight θ_i/w_n, gives the average search cost and letting t tend to infinity gives the limiting search cost. These average and limiting distributions are studied in [29] and [28], which give extensive references to the literature. These statistics are close to the number of inversions, and it may be hoped that some of the novel techniques introduced in [29] and [28] can be used to get at the distribution of inversions.

A final item; throughout, we have assumed that the weights θ_i are fixed and known. It is also natural to consider random weights. For a full development, see [43].

Section 5 develops the connections of the Tsetlin library to the Bidigare–Hanlon–Rockmore (BHR) walk on the chambers of a real hyperplane arrangement. Understanding the stationary distributions of these Markov chains is almost completely open.

Section 2 begins with a review of enumerative group theory. These questions make sense for continuous groups. Georgia Benkart made fundamental contributions here through her work on decomposing tensor products.

2. Background

2.1. Enumerative group theory

Let G be a finite group. A classical question is "pick $g \in G$ at random. What does it look like?" For example, if $G = S_n$,

- What is the distribution of $F(g)$, the number of fixed points of g?
- How many cycles are typical?
- What is the expected length of the longest cycle?
- What about the length of the longest increasing subsequence of g?
- What about the descent structure of g?
- How many inversions are typical?

All of these questions have classical answers (references below).

For $G = GL_n(q)$, parallel questions involve the conjugacy class structure of a random $g \in G$. For a splendid development (for finite groups of Lie type), see Fulman [30], which has full references to the results above. The recent survey of Diaconis and Simper [23] brings this up to date. It focuses on enumeration by double cosets $H \setminus G/K$.

The questions above make sense for continuous groups, where they become "random matrix theory." For example, when $G = O_n$ (the real orthogonal group), one may study the eigenvalues of $g \in G$ under Haar measure by studying the powers of traces

$$\int_{O_n} (\mathrm{Tr}(g))^k \, \mathrm{d}g.$$

Patently this asks for the number of times the trivial representation appears in the kth tensor power of the usual n-dimensional representation of O_n. See [20] for details.

Georgia Benkart did extensive work on decomposing tensor powers of representations of classical (and more general) groups. She worked on this with many students and coauthors. Her monograph with Britten and Lemire [7] is a convenient reference. Most of this work can be translated into probabilistic limit theorems. We started to do this with Georgia during MSRI 2018, but got sidetracked into doing a parallel problem working over fields of prime characteristic in joint work with Benkart–Diaconis–Liebeck–Tiep [8].

Most all of the above is enumeration under the uniform distribution. A recent trend in enumerative (probabilistic) group theory is enumeration under natural *non*-uniform distributions. For example, on S_n,

- The Ewens measure $\pi_\theta(\sigma) = Z^{-1}(\theta)\theta^{C(\sigma)}$. Here, θ is a fixed positive real number, $C(\sigma)$ is the number of cycles of σ, and $Z^{-1}(\theta)$ is a simple normalizing constant. The Ewens measure originated in biology, but has blossomed into a large set of applications. See Crane [18].

- The Mallows measure $\pi_\theta(\sigma) = Z^{-1}(\theta)\theta^{I(\sigma)}$, where $I(\sigma)$ is the number of inversions of θ. This was originally studied for taste testing experiments but has again had a huge development.
- More generally, if G is a finite group and $S \subseteq G$ is a symmetric generating set, let $\ell(g)$ be the length function and define $P_\theta(g) = Z^{-1}(\theta)\theta^{\ell(g)}$. Ewens and Mallows models are special cases with $G = S_n$ and $S = \{\text{all transpositions}\}$ and $S = \{\text{all adjacent transpositions}\}$.

Most of the questions studied above under the uniform distribution have been fully worked out under Ewens and Mallows measures. See the survey by Diaconis and Simper [23] for pointers to a large literature.

The above can be amplified to "permutons" [34] and "theons" [17]. It shows that enumeration under non-uniform distributions is an emerging and lively subject. We turn next to the main subject of the present paper.

2.2. The Luce model

This section gives several applications where the Luce model appears.

2.2.1. Psychology

In psychophysics experiments, a panel of subjects are asked to rank things, such as:

- Here are seven shades of red; rank them in order of brightness.
- Here are five tones; rank them from high to low.
- The same type of task occurs in taste-testing experiments. Rank these five brands of chocolate chip cookies (or wines, etc.) in order of preference.

This generates a collection of rankings (permutations) and one tries to draw conclusions.

Patently, rankings vary stochastically; if the same person is asked the same question at a later time, we expect the answers to vary slightly.

Duncan Luce introduce the model (1) via the simple idea that each item has a true weight (say, θ_i) and the model (1) induces natural variability (which can then be compared with observed data).

Indeed, he did more, crafting a simple set of axioms for pairwise comparison and showing that any consistent ranking distribution has to follow (1) for some choice of θ_i. This story is well and clearly told in [36] and [37].

We would be remiss in not pointing to the widespread dissatisfaction over the "independence of irrelevant alternatives" axiom in Luce's derivation. The long Wikipedia article on "irrelevance of alternatives" chronicles experiments and theory disputing this, not only for Luce but in Arrow's paradox and several related developments. Amos Tversky's "elimination by aspects (EBA)" model is a well-liked alternative.

2.2.2. Exponential formulation

Luce's work followed fifty years of effort to model such rankings. Early work of Thurstone and Spearman postulated "true weights" $\theta_1, \ldots, \theta_n$ for the ordered values and supposed people perceived $\theta_i + \varepsilon_i, 1 \le i \le n$ with ε_i independent normal $\mathcal{N}(0, \sigma^2)$. They then reported the ordering of these perturbed values. See [26] for an up-to-date account of Thurstonian ranking models.

Yellott [48] noticed that if in fact the ε_i had an extreme value distribution, with distribution function $e^{-e^{-x/r}}, -\infty < x < \infty$, then the associated Thurstonian ranking model is exactly the Luce model!

It is elementary, that if the random variable Y has an exponential distribution ($P(Y > x) = e^{-x}$), then $\log Y$ has an extreme value distribution. This gives the following theorem (used in Section 4):

Theorem 2.1. *For $1 \le i \le n$, let X_i be independent exponential random variables on $[0, \infty)$ with density*

$$\theta_i e^{-x\theta_i}$$

(so $X_i = Y_i/\theta_i$ with Y_i the standard exponential). Then, the chance of the event

$$X_1 < X_2 < \cdots < X_n$$

is (with $w_n = \theta_1 + \cdots + \theta_n$)

$$\frac{\theta_1}{w_n} \cdot \frac{\theta_2}{w_n - \theta_1} \cdot \frac{\theta_3}{w_n - \theta_1 - \theta_2} \cdots \frac{\theta_n}{\theta_n}.$$

Proof. Consider the event $X_1 < X_2 < \cdots < X_n$. The chance of this is

$$\begin{aligned}
&\theta_1 \cdots \theta_n \int_{x_1=0}^{\infty} \int_{x_2=x_1}^{\infty} \cdots \int_{x_n=x_{n-1}}^{\infty} \exp\left\{ -\sum_{i=1}^{n} x_i \theta_i \right\} dx_1 \cdots dx_n \\
&= \frac{\theta_1 \cdots \theta_n}{\theta_n} \int_{x_1=0}^{\infty} \cdots \int_{x_{n-1}=x_{n-2}}^{\infty} \exp\left\{ -\sum_{i=1}^{n-2} x_i \theta_i - x_{n-1} (\theta_{n-1} + \theta_n) \right\} dx_1 \cdots dx_{n-1} \\
&= \frac{\theta_1 \cdots \theta_n}{\theta_n (\theta_n + \theta_{n-1})} \int_{x_1=0}^{\infty} \cdots \int_{x_{n-2}=x_{n-3}}^{\infty} \exp\left\{ -\sum_{i=1}^{n-3} x_i \theta_i - x_{n-2} (\theta_{n-2} + \theta_{n-1} + \theta_n) \right\} dx_1 \cdots dx_{n-2} \\
&= \frac{\theta_1 \cdots \theta_n}{\theta_n (\theta_n + \theta_{n-1})(\theta_n + \theta_{n-1} + \theta_{n-2}) \cdots (\theta_n + \cdots + \theta_1)},
\end{aligned}$$

which is indeed equal to

$$\frac{\theta_1}{w_n} \cdot \frac{\theta_2}{w_n - \theta_1} \cdot \frac{\theta_3}{w_n - \theta_1 - \theta_2} \cdots \frac{\theta_n}{\theta_n}.$$

Thus, the order statistics follow the Luce model (1). □

For an application of the exponential representation to survey sampling, see Gordon [31].

2.2.3. Tsetlin library

The great algebraist Tsetlin was forced to work in a library science institute. While there, he postulated (and solved) the following problem:

Consider n library books arranged in order $1, 2, \ldots, n$. Suppose book i has popularity θ_i. During the day, patrons come and pick up book labeled i with probability θ_i/w_n and after perusing, they replace it at the left end of the row.

This is a Markov chain on S_n and Tsetlin [47] showed that it has (1) as its stationary distribution.

The same model has been repeatedly rediscovered; in computer science, the books are discs in deep storage. When a disc is called for, it is replaced on the front of the queue to cut down on future search costs. See Dobrow and Fill [24].

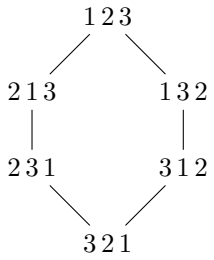
The model (and its stationary distribution) appear in genetics as the GEM (Griffiths–Engen–McCloskey) distribution [25].

Over the years, a host of properties of the Tsetlin chain have been derived. For example, Phatarfod [42] found a simple formula for the eigenvalues and Diaconis [21] found sharp rates of convergence to stationarity (including a cutoff) for a wide class of weights. See further Nestoridi [39]. All of this is now subsumed under "hyperplane walks"; see Section 5.

A monotonicity property. Suppose, without essential loss, that $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_n > 0$. Then,

- The largest $\pi(\sigma)$ is for $\sigma = \mathrm{id}$.
- The smallest $\pi(\sigma)$ is for $\sigma = (n, n-1, \ldots, 1)$.
- More generally, $\pi(\sigma)$ is monotone decreasing in the weak Bruhat order on permutations.

To explain, the weak Bruhat order is a partial order on S_n with cover relations $\sigma \preceq \sigma'$ if σ can be reached from σ' by a single adjacent transposition of the ith and $(i+1)$th symbols when $\sigma'(i) < \sigma'(i+1)$. Thus, when $n = 3$,



Proposition 2.2. *For $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_n$, $\pi(\sigma)$ is monotone decreasing in the weak Bruhat order.*

Proof. The formulas for $\pi(\sigma)$ and $\pi(\sigma')$ only differ in one term. If $\sigma_i > \sigma_{i+1}$ and these terms were transposed from σ', then

$$\frac{\pi(\sigma)}{\pi(\sigma')} = \frac{1 - \theta_{\sigma_1} - \theta_{\sigma_2} - \cdots - \theta_{\sigma_{i+1}}}{1 - \theta_{\sigma_1} - \theta_{\sigma_2} - \cdots - \theta_{\sigma_i}} < 1. \quad \square$$

Irrelevance of alternatives. The following property characterizes the Luce measure (1): let $\{i_1, i_2, \ldots, i_k\}$ be a subset of $\{1, 2, \ldots, n\}$. Fix $\theta_1, \ldots, \theta_n$ and pick σ from $\pi(\sigma)$. The distribution of cards labeled $\{i_1, \ldots, i_k\}$ follows the Luce model with parameters $\theta_{i_1}, \ldots, \theta_{i_k}$. For example, the chance that i is above j in σ is $\dfrac{\theta_i}{\theta_i + \theta_j}$. This is easy to see from the exponential representation.

2.2.4. Order statistics and a natural choice of weights

Many questions in probability and mathematical statistics can be reduced to the study of the order statistics of uniform random variables on $[0, 1]$ by using the simple fact that, if X is a real random variable with continuous distribution function $F(x)$ (so $P(X \leq x) = F(x)$), then $Y = F(X)$ is uniformly distributed on $[0, 1]$. This implies that standard goodness of fit tests (e.g., Kolmogorov–Smirnov) have distributions that are universal under the null hypothesis (they do not depend on F). If Y is uniform on $[0, 1]$, then $-\log Y$ is standard exponential as above, so order statistics of independent exponentials are a mainstream object of study. A marvelous introduction to this set of ideas is in Chapter 3 of [27] with Ronald Pyke's articles on spacings [41] providing deeper results.

With this background, let $Y_1, Y_2, \ldots, Y_n$ be independent standard exponentials on $(0, \infty)$. Denote the order statistics by $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$. The following property is easy to prove [27].

Theorem 2.3. *With above notation,*

$$Y_{(1)}, Y_{(2)} - Y_{(1)}, Y_{(3)} - Y_{(2)}, \ldots, Y_{(n)} - Y_{(n-1)}$$

are independent exponential random variables with distributions

$$Y_{(1)} \sim E_1/n, \qquad Y_{(2)} - Y_{(1)} \sim E_2/(n-1), \qquad \ldots \qquad Y_{(n)} - Y_{(n-1)} \sim E_n,$$

where $E_1, \ldots, E_n$ are independent standard exponentials (density e^{-x} on $(0, \infty)$).

It follows from our Luce calculations that the chance that the smallest spacing is $Y_{(1)}$ is $\dfrac{n}{\binom{n+1}{2}} = \dfrac{2}{n+1}$, and that the smallest spacing is $Y_{(2)} - Y_{(1)}$ is $\dfrac{n-1}{\binom{n+1}{2}}$, and so on.

Specifically, $Y_{(j+1)} - Y_{(j)}$ has probability $\frac{n-j}{\binom{n+1}{2}}$ of being smallest, and $Y_{(n)} - Y_{(n-1)}$ has chance $\frac{1}{\binom{n+1}{2}}$ of being smallest.

The whole permutation is given by the Luce model (1) with $\theta_i = n - i + 1$. This classical fact is due to Sukhatme [46]. We will call these **Sukhatme weights** in the following discussion.

2.2.5. Application to poker and the ICM (iterated card model)

In tournament poker (e.g., the World Series of Poker), suppose there are n players at the final table with player i having θ_i dollars. It is current practice among top players to assume that the order of the players, as they are eliminated, follows the Luce model (with the player having the largest θ_i least likely to be eliminated; thus most likely to win all the money), and so on. This is called the ICM (iterated card model) and is used as a basis for splitting the total capital and for calculating chances as the game progresses. For careful details and references, see Diaconis–Ethier [22], which disputes the model.

2.2.6. Applications to horse racing

In horse racing, players can bet on a horse to win, place (come in second), or show (come in third). The "crowd" does a good job of determining the chances of each of the n horses running to come in first. Call the amount bet on horse i just before closing, θ_i. However, the crowd does a poor job of judging the chance of a horse showing. Often, there is sufficient disparity between the crowd's bet and the true odds that money can be made (perhaps one race in four). This is despite the track's rake being 17% of the total. A group of successful bettors uses the θ_i's and the Luce model to evaluate the chance of placing. For details, see Hausch, Lo and Ziemba [33] or Harville [32].

With this list of applications, we trust we have sufficient motivation to ask "what does the distribution (1), $\pi(\sigma)$, look like?"

3. The top k cards

Throughout this section, without loss of generality, assume $\theta_1 + \cdots + \theta_n = 1$. For $\boldsymbol{\theta}$ and k fixed, let

$$P(\sigma_1\, \sigma_2\, \cdots\, \sigma_k) = \frac{\theta_{\sigma_1}\theta_{\sigma_2}\cdots\theta_{\sigma_k}}{(1-\theta_{\sigma_1})(1-\theta_{\sigma_1}-\theta_{\sigma_2})\cdots(1-\theta_{\sigma_1}-\cdots-\theta_{\sigma_{k-1}})} \tag{2}$$

denote the measure induced on the top k cards by the Luce measure. It is cumbersome to compute, e.g.,

$$P(\sigma_2) = \theta_{\sigma_2} \sum_{i \neq \sigma_2} \frac{\theta_i}{1-\theta_i}.$$

On the other hand, the Luce measure is just sampling from an urn without replacement. If $\{\theta_i\}$ are "not too wild" and k is small, then sampling with or without replacement should be "about the same." This is made precise in two metrics.

Let $Q(\sigma_1\,\sigma_2\,\cdots\,\sigma_k)$ be the product measure

$$Q(\sigma_1\,\sigma_2\,\cdots\,\sigma_k) = \theta_{\sigma_1}\theta_{\sigma_2}\cdots\theta_{\sigma_k}, \tag{3}$$

where $\sigma_1,\ldots,\sigma_k$ need not be distinct. Both P and Q depend on $\{\theta_i\}$ and k, but this is suppressed below. Define

$$d_\infty(P,Q) = \max_\sigma \left(1 - \frac{Q(\sigma_1\,\cdots\,\sigma_k)}{P(\sigma_1\,\cdots\,\sigma_k)}\right)$$

and

$$\|P-Q\|_{TV} = \frac{1}{2}\sum_\sigma |P(\sigma_1\,\cdots\,\sigma_k) - Q(\sigma_1\,\cdots\,\sigma_k)|\,,$$

where, in both formulas, $\sigma_1,\ldots,\sigma_k$ are not necessarily distinct, and $P(\sigma_1\,\cdots\,\sigma_k) = 0$ if they are not distinct. Clearly, $\|P-Q\|_{TV} \le d_\infty(P,Q)$.

Theorem 3.1. *For $\theta_1+\cdots+\theta_n = 1$, $\theta_i \le \frac{1}{2}$ for all $1 \le i \le n$,*

$$d_\infty(P,Q) \le 1 - \exp\left\{-2\left((k-1)\theta_{(1)} + (k-2)\theta_{(2)} + \cdots + \theta_{(k-1)}\right)\right\}.$$

Here, $\theta_{(1)} \ge \theta_{(2)} \ge \cdots \ge \theta_{(n)}$ are the ordered values.

Theorem 3.2. *As $n\to\infty$, suppose $\binom{k}{2}\sum_{i=1}^n \theta_i^2 \to \lambda$. Then,*

$$\|P-Q\|_{TV} \sim 1 - e^{-\lambda}.$$

In Theorem 3.2, $\{\theta_i\}$ form a triangular array, but again, this is suppressed in the notation. The remarks below point to non-asymptotic versions.

Proof of Theorem 3.1. From the definitions,

$$d_\infty(P,Q) = \max_\sigma \left(1 - (1-\theta_{\sigma_1})(1-\theta_{\sigma_1}-\theta_{\sigma_2})\cdots(1-\theta_{\sigma_1}-\cdots-\theta_{\sigma_{k-1}})\right),$$

where the maximum is over all $\sigma_1,\ldots,\sigma_k$ distinct (because, if they are not distinct, then we have that $1 - \frac{Q(\sigma_1\,\cdots\,\sigma_k)}{P(\sigma_1\,\cdots\,\sigma_k)} = -\infty$, which does not contribute to the maximum). Use $-2x \le \log(1-x) \le -x$ for $0 \le x \le \frac{1}{2}$. Since all $\theta_i \le \frac{1}{2}$,

$$d_\infty(P,Q) \le \max_\sigma 1 - \exp\left\{-2\left(\theta_{\sigma_1} + (\theta_{\sigma_1} + \theta_{\sigma_2}) + \cdots + (\theta_{\sigma_1} + \cdots + \theta_{\sigma_{k-1}})\right)\right\}.$$

The right-hand side is maximized for $\sigma_1, \ldots, \sigma_{k-1}$ with the largest weights. □

Proof of Theorem 3.2. A prepatory observation is useful:

$$\|P - Q\|_{TV} = \sum_{\sigma: P(\sigma) \ge Q(\sigma)} (P(\sigma) - Q(\sigma)) = 1 - P_Q(\sigma_1, \ldots, \sigma_k \text{ are distinct}).$$

This is just the chance that there are two or more balls in the same box if k balls are dropped independently into n boxes, the chance of box i being θ_i. This non-uniform version of the classical birthday problem has been well-studied. If X_{ij} is 1 or 0 as balls i, j are dropped into the same box and

$$W = \sum_{1 \le i < j \le k} X_{ij},$$

$E(W) = \binom{k}{2} \sum_{i=1}^n \theta_i^2$. Under the condition $E(W) \to \lambda$, W is known to have a limiting Poisson(λ) distribution and $P_Q(W = 0) \sim e^{-\lambda}$. See Chatterjee–Diaconis–Meckes [16] or Barbour–Holst–Janson [4] for further details and more quantitative bounds. □

Example 3.3. Consider the Sukhatme weights from Section 2.2.4:

$$\theta_i = \frac{n + 1 - i}{\binom{n+1}{2}}.$$

The exponent for the right-hand side of Theorem 3.1 is

$$\frac{2}{\binom{n+1}{2}} \left\{(k-1)n + (k-2)(n-1) + \cdots (n-k+2)\right\}.$$

Simple asymptotics show that for $k = c\sqrt{n}$, $c > 0$, this is

$$4c^2 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

So,

$$d_\infty(P,Q) \le 1 - e^{-4c^2 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)},$$

and $k \ll \sqrt{n}$ suffices for product measures to be a useful approximation to the first k-coordinates of the Luce measure. With $k = c\sqrt{n}$,

$$\binom{k}{2}\sum_{i=1}^{n}\theta_i^2 \sim \frac{c^2}{3} = \lambda$$

giving a similar approximation in total variation.

Example 3.4. The bound in Theorem 3.1 is useful when

$$(k-1)\theta_{(1)} + \cdots + \theta_{(k-1)}$$

is small. To see that this condition is needed, take $\theta_1 = \frac{1}{2}$, $\theta_i = \frac{1}{2(n-1)}$ for $2 \le i \le n$. For $k = 2$,

$$P(1\,2) = \frac{\theta_1\theta_2}{1-\theta_1}, \qquad Q(1\,2) = \theta_1\theta_2,$$

and so, $d_\infty(P,Q) = 1 - (1-\theta_1) = \frac{1}{2}$. This does not tend to zero when n is large. The two-sided bounds for $\log(1-x)$ show Theorem 3.1 is sharp in this sense for general k.

Example 3.5. As discussed above, if the infinity distance tends to zero, then total variation tends to zero. Here is a choice of weights θ_i so that the total variation convergence holds for the joint distribution of the first k coordinates of the Luce model to i.i.d., but not in infinity distance.

Fix $k, 1 \le k \le n$ and let $\theta_i = k^{-7/4}$ for $i \le k$ and

$$\theta_i = \frac{1-k^{-3/4}}{n-k}$$

for $i > k$ so that $\theta_1 \ge \theta_2 \ge \cdots \ge \theta_n > 0$ and $\sum_{i=1}^{n}\theta_i = 1$. Note that

$$\binom{k}{2}\sum_{i=1}^{n}\theta_i^2 \le k^2\sum_{i=1}^{k}\theta_i^2 + k^2\sum_{i=k+1}^{n}\theta_i^2 \le k^{-1/2} + \frac{k^2}{n-k}$$

while

$$\sum_{i=1}^{k-1}(k-i)\theta_i = k^{-7/4}\sum_{i=1}^{k-1}(k-i) = \frac{1}{2}\left(k^{-3/4}(k-1)\right).$$

Thus, if $1 \ll k \ll \sqrt{n}$, $\sum_{i=1}^{k-1}(k-i)\theta_i$ is large, but $\binom{k}{2}\sum_{i=1}^{n}\theta_i^2$ is small. Moreover, if $k = \lambda\sqrt{n}$, then $\binom{k}{2}\sum_{i=1}^{n}\theta_i^2 \to \frac{\lambda^2}{2}$, but $\sum_{i=1}^{k-1}(k-i)\theta_i \to \infty$.

Remark 3.6.

(a) Our proof of Theorem 3.2 used the Poisson approximation for the non-uniform version of the birthday problem. There are other possible limits which can be used to bound $\|P - Q\|_{TV}$. See [15].

(b) It is easy to see that

$$Q_\theta(\sigma_1, \ldots, \sigma_k \text{ distinct}) = k! e_k(\theta_1, \theta_2, \ldots, \theta_n),$$

where e_k is the kth elementary symmetric function. From here, Muirhead's theorem shows $\|P - Q\|_{TV}$ is a Schur-concave function of $\theta_1, \ldots, \theta_n$, smallest when $\theta_i = \frac{1}{n}$.

4. The bottom k cards

4.1. Introduction

For naturally occurring weights, the bottom k cards behave very differently from the top k cards. To illustrate by example, consider the Sukhatme weights of Section 2.2.4:

$$\theta_i = \frac{n+1-i}{\binom{n+1}{2}}, \qquad 1 \le i \le n.$$

The results of Section 3 show that, for large n,

$$P\left(\frac{\sigma_1}{n} \le x\right) \sim 2 \int_0^x (1-y)\, \mathrm{d}y.$$

That is, σ_1/n has a limiting $\beta(1,2)$ distribution.

Using Theorems 3.1 and 3.2, the same holds for σ_i/n for fixed $i \ll \sqrt{n}$. Of course, large numbers have higher probabilities, but all values in $\{1, 2, \ldots, n\}$ occur.

In contrast, consider the value of bottom card σ_n. Intuitively, this should be small since the high numbers have higher weights. We were surprised to find

$$P(\sigma_n = 1) \sim 0.516\ldots$$

In fact, we computed, using a result that follows, that

ℓ	$P(\ell$ is last$)$
1	0.516094
2	0.213212
3	0.107310
4	0.0597505
5	0.0354888
6	0.0220716
7	0.0142167
8	0.00941619
9	0.00638121
10	0.00440862

The section below sets up its own notation from first principles.

4.2. Main result

Let $\mathbb{N}$ denote the set of positive integers and let $\mathbb{N}^{\mathbb{N}}$ be the set of all maps from $\mathbb{N}$ into $\mathbb{N}$. Consider the topology of pointwise convergence on $\mathbb{N}^{\mathbb{N}}$. This topology is naturally metrizable with a complete separable metric, and so we can talk about convergence of probability measures on this space.

Now suppose that for each n, σ_n is a random element of the symmetric group S_n. We can extend σ_n to a random element of $\mathbb{N}^{\mathbb{N}}$, by defining $\sigma_n(i) = i$ for $i > n$.

Proposition 4.1. *Let σ_n be as above. Then σ_n converges in law as $n \to \infty$ if and only if for each k, the random vector $(\sigma_n(1), \ldots, \sigma_n(k))$ converges in law as $n \to \infty$.*

Proof. Since the coordinate maps on $\mathbb{N}^{\mathbb{N}}$ are continuous in the topology of pointwise convergence, one direction is clear.

For the other direction, suppose that for each k, $(\sigma_n(1), \ldots, \sigma_n(k))$ converges in law as $n \to \infty$. Notice that for any sequence of positive integers $a_1, a_2, \ldots$, the set

$$\left\{ f \in \mathbb{N}^{\mathbb{N}} : f(i) \le a_i \text{ for all } i \right\} \tag{4}$$

is a compact subset of $\mathbb{N}^{\mathbb{N}}$, since any infinite sequence in this set has a convergent subsequence by a diagonal argument. Take any $\varepsilon > 0$. By the given condition, $\sigma_n(i)$ converges in law as $n \to \infty$ for each i. In particular, $\{\sigma_n(i)\}_{n \ge 1}$ is a tight family (a family of measures is **tight** if it is almost compactly supported, see [10]), and so there is some number a_i such that for each n,

$$P(\sigma_n(i) > a_i) \le 2^{-i}\varepsilon.$$

Therefore if K denotes the set defined in (4) above, then for each n,

$$P(\sigma_n \in K) \ge 1 - \sum_{i=1}^{\infty} P(\sigma_n(i) > a_i) \ge 1 - \sum_{i=1}^{\infty} 2^{-i}\varepsilon = 1 - \varepsilon.$$

This proves that $\{\sigma_n\}_{n \ge 1}$ is a tight family of random variables on $\mathbb{N}^{\mathbb{N}}$. Therefore the proof will be complete if we can show that any probability measure on $\mathbb{N}^{\mathbb{N}}$ is determined by its finite dimensional distributions. But this is an easy consequence of Dynkin's π-λ theorem (see [10] for a theorem and proof of the Dynkin π-λ theorem). □

The above proposition implies, for instance, that if σ_n is a uniform random element of S_n, then σ_n does not converge in law on $\mathbb{N}^{\mathbb{N}}$, because $\sigma_n(1)$ does not converge in law.

Let $0 < \theta_1 \le \theta_2 \le \cdots$ be a non-decreasing infinite sequence of positive real numbers. For each n, consider the Luce model on S_n with parameters $\theta_1, \ldots, \theta_n$. Let σ_n be the *reverse* of a random permutation drawn from this model. That is, $\sigma_n(1)$ is the last ball

that was drawn and $\sigma_n(n)$ is the first. As we know from prior discussions, an equivalent definition is the following. Let $X_1, X_2, \ldots$ be an infinite sequence of independent random variables, where X_i has exponential distribution with mean $1/\theta_i$. Then $\sigma_n \in S_n$ is the permutation such that $X_{\sigma_n(1)} > X_{\sigma_n(2)} > \cdots > X_{\sigma_n(n)}$.

Theorem 4.2. *Let σ_n be as above. For each $x \geq 0$, let*

$$f(x) := \sum_{i=1}^{\infty} e^{-\theta_i x},$$

where we allow $f(x)$ to be ∞ if the sum diverges. Let

$$x_0 := \inf\{x : f(x) < \infty\},$$

with the convention that the infimum of the empty set is ∞. Then σ_n converges in law as $n \to \infty$ if and only if $x_0 < \infty$ and $f(x_0) = \infty$. Moreover, if this condition holds, then the limiting finite dimensional probability mass functions are given by the following formula: For any k and any distinct positive integers $a_1, \ldots, a_k$,

$$\lim_{n\to\infty} P(\sigma_n(1) = a_1, \ldots, \sigma_n(k) = a_k)$$
$$= \int_{x_1 > x_2 > \cdots > x_k > 0} \prod_{j=1}^{k} (\theta_{a_j} e^{-\theta_{a_j} x_j}) \prod_{i \notin \{a_1,\ldots,a_k\}} (1 - e^{-\theta_i x_k})\, \mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

Before proving the theorem, let us work out some simple examples. Suppose that $\theta_i = i$ for each i. This corresponds to the Luce model with the Sukhatme weights. Then clearly $f(x) < \infty$ for all $x > 0$, and hence $x_0 = 0$. Also, clearly, $f(0) = \infty$. Therefore in this case σ_n converges in law as $n \to \infty$. Moreover, by the formula displayed above,

$$\lim_{n\to\infty} P(\sigma_n(1) = 1) = \int_0^{\infty} e^{-x} \prod_{j=2}^{\infty} (1 - e^{-jx}) dx = \int_0^1 \prod_{j=2}^{\infty} (1 - y^j) dy = 0.516094\ldots$$

On the other hand, for the case of uniform random permutations, $\theta_i = 1$ for all i. In this case, $f(x) = \infty$ for all x, and hence $x_0 = \infty$. Thus, the theorem implies that σ_n does not converge in law (which we know already).

Next, suppose that $\theta_i = \beta \log(i+1)$ for some $\beta > 0$. Here $f(x) < \infty$ for $x > 1/\beta$ and $f(x) = \infty$ for $x \leq 1/\beta$. Thus, $x_0 = 1/\beta$ and $f(x_0) = \infty$, and so by the theorem, σ_n converges in law.

Strangely, σ_n *does not* converge in law if $\theta_i = \log(i+1) + 2\log\log(i+1)$. To see this, note that in this case,

$$f(x) = \sum_{i=1}^{\infty} \frac{1}{(i+1)^x (\log(i+1))^{2x}}.$$

Thus, $f(x) < \infty$ for $x > 1$ and $f(x) = \infty$ for $x < 1$, showing that $x_0 = 1$. But

$$f(x_0) = \sum_{i=1}^{\infty} \frac{1}{(i+1)(\log(i+1))^2} < \infty,$$

which violates the second criterion required for convergence. This shows that we cannot determine convergence purely by inspecting the rate of growth of θ_i. The criterion is more subtle than that.

What happens if the tightness criterion does not hold? In this case, the formula for the limit of $P(\sigma_n(1) = a_1, \ldots, \sigma_n(k) = a_k)$ remains valid, but it may not represent a probability mass function, i.e., the sum over all $a_1, \ldots, a_k$ may be strictly less than 1.

Proof of Theorem 4.2. Take any $k \geq 1$ and distinct positive integers $a_1, \ldots, a_k$. Take $n \geq \max_{1 \leq i \leq k} a_i$. Let E_n be the event $\{\sigma_n(1) = a_1, \ldots, \sigma_n(k) = a_k\}$. Then

$$\begin{aligned} P(E_n) &= P(X_{a_1} > X_{a_2} > \cdots > X_{a_k} > X_i \ \forall i \in [n] \setminus \{a_1, \ldots, a_k\}) \\ &= \int_{x_1 > x_2 > \cdots > x_k > 0} \prod_{j=1}^{k} (\theta_{a_j} e^{-\theta_{a_j} x_j}) \prod_{i \in [n] \setminus \{a_1, \ldots, a_k\}} (1 - e^{-\theta_i x_k}) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k. \end{aligned}$$

By the dominated convergence theorem, this gives

$$\lim_{n \to \infty} P(E_n) = \int_{x_1 > x_2 > \cdots > x_k > 0} \prod_{j=1}^{k} (\theta_{a_j} e^{-\theta_{a_j} x_j}) \prod_{i \notin \{a_1, \ldots, a_k\}} (1 - e^{-\theta_i x_k}) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

Thus, we have shown that for any k and distinct positive integers $a_1, \ldots, a_k$, $\lim_{n \to \infty} P(\sigma_n(1) = a_1, \ldots, \sigma_n(k) = a_k)$ exists, and also found the desired formula for the limit. However, we have not shown convergence in law because we have not established tightness. (This is not surprising, because we did not use any properties of the θ_i's yet.) From what we have done until now, it follows that $(\sigma_n(1), \ldots, \sigma_n(k))$ converges in law as $n \to \infty$ if and only if it is a tight family. But this holds if and only if $\{\sigma_n(i)\}_{n \geq 1}$ is a tight family for every i. We will now complete the proof of the theorem by showing that $\{\sigma_n(i)\}_{n \geq 1}$ is a tight family for every i if and only if $x_0 < \infty$ and $f(x_0) = \infty$.

First, suppose that $\{\sigma_n(1)\}_{n \geq 1}$ is a tight family. Then there is some a such that

$$\lim_{n \to \infty} P(\sigma_n(1) = a) > 0.$$

From the above calculation, we know that

$$\lim_{n \to \infty} P(\sigma_n(1) = a) = \int_0^{\infty} \theta_a e^{-\theta_a x} \prod_{i \neq a} (1 - e^{-\theta_i x}) \, \mathrm{d}x.$$

If this is nonzero, then there is at least one $x > 0$ for which

$$\prod_{i=1}^{\infty}(1 - e^{-\theta_i x}) > 0.$$

But this implies that

$$f(x) = \sum_{i=1}^{\infty} e^{-\theta_i x} < \infty.$$

Thus, $x_0 < \infty$. Next, we show that $f(x_0) = \infty$. Suppose not. Then $x_0 > 0$, since $f(0) = \infty$. Fix a positive integer a. For each $n \geq a$, and let A_n be the event $\{\sigma_n(1) \leq a\}$. Let F_n be the event $\{\max_{i \leq n} X_i \leq x_0\}$. Take any $x \in (0, x_0)$ and let G_n be the event $\{\max_{i \leq n} X_i \leq x\}$. Then

$$\begin{aligned} P(A_n) &\leq P(A_n \cap (F_n \setminus G_n)) + P((F_n \setminus G_n)^c) \\ &= P(A_n \cap (F_n \setminus G_n)) + P(F_n^c \cup G_n) \\ &\leq P(A_n \cap (F_n \setminus G_n)) + P(G_n) + P(F_n^c). \end{aligned}$$

If the event $A_n \cap (F_n \setminus G_n)$ happens, then $\max_{i \leq n} X_i$ belongs to the interval $(x, x_0]$, and one of $X_1, \ldots, X_a$ is the maximum among $X_1, \ldots, X_n$. Thus, in particular, one of $X_1, \ldots, X_a$ is in $(x, x_0]$. Plugging this into the above inequality, we get

$$P(A_n) \leq \sum_{i=1}^{a}(e^{-\theta_i x} - e^{-\theta_i x_0}) + \prod_{i=1}^{n}(1 - e^{-\theta_i x}) + 1 - \prod_{i=1}^{n}(1 - e^{-\theta_i x_0}).$$

Since $f(x) = \infty$, we have $\prod_{i=1}^{\infty}(1 - e^{-\theta_i x}) = 0$. Thus, taking $n \to \infty$ on both sides, we get

$$\lim_{n \to \infty} P(A_n) \leq \sum_{i=1}^{a}(e^{-\theta_i x} - e^{-\theta_i x_0}) + 1 - \prod_{i=1}^{\infty}(1 - e^{-\theta_i x_0}).$$

Now notice that the definition of A_n does not involve x. So we can take $x \nearrow x_0$ on the right, which makes the first term vanish and leaves the rest as it is. Thus,

$$\lim_{n \to \infty} P(A_n) \leq 1 - \prod_{i=1}^{\infty}(1 - e^{-\theta_i x_0}).$$

But the assumed finiteness of $f(x_0)$ implies that the product on the right is strictly positive. Thus, we get an upper bound on $\lim_{n \to \infty} P(A_n)$ which is less than 1. But observe that this upper bound does not depend on a. This contradicts the tightness of $\sigma_n(1)$, thereby completing the proof of one direction of the theorem.

Next, suppose that $x_0 < \infty$ and $f(x_0) = \infty$. We consider two cases. First, suppose that $x_0 = 0$. Then $f(x) < \infty$ for each $x > 0$. But

$$f(x) = \sum_{i=1}^{\infty} P(X_i > x). \tag{5}$$

Therefore by the Borel–Cantelli lemma, $X_i \to 0$ almost surely as $i \to \infty$. Now take any i and integers n and a bigger than i. Then the event $\sigma_n(i) \geq a$ implies that

$$\max_{j \geq a} X_j > \min\{X_1, \ldots, X_i\},$$

because otherwise the ith largest value among $(X_j)_{j=1}^n$ cannot be one of $(X_j)_{j \geq a}$. Thus,

$$P(\sigma_n(i) \geq a) \leq P\left(\max_{j \geq a} X_j > \min\{X_1, \ldots, X_i\}\right).$$

But the right side is a function of only a (and not n), and tends to zero as $a \to \infty$ because $X_j \to 0$ almost surely as $j \to \infty$. This proves tightness of $\{\sigma_n(i)\}_{n \geq 1}$ when $x_0 = 0$.

Next, consider the case $x_0 > 0$. For convenience, let us define the partial sums

$$f_n(x) := \sum_{i=1}^{n} e^{-\theta_i x}, \qquad g_n(x) := \prod_{i=1}^{n} \left(1 - e^{-\theta_i x}\right).$$

Take i, n and a as before. Let x be a real number bigger than x_0, to be chosen later. The event $\sigma_n(i) \geq a$ implies that at least one of the following two events must happen: (a) There are less than i elements of $(X_j)_{j=1}^n$ that are bigger than x, or (b) $X_j > x$ for some $j \geq a$. This gives

$$P(\sigma_n(i) \geq a) \leq \sum_{A \subseteq [n], |A| < i} \left(\prod_{j \in A} e^{-\theta_j x}\right) \left(\prod_{j \in [n] \setminus A} \left(1 - e^{-\theta_j x}\right)\right) + \sum_{j \geq a} e^{-\theta_j x}.$$

Now note that for any $A \subseteq [n]$ with $|A| < i$,

$$\prod_{j \in [n] \setminus A} \left(1 - e^{-\theta_j x}\right) \leq \frac{g_n(x)}{\prod_{j \in A} \left(1 - e^{-\theta_j x}\right)} \leq \frac{g_n(x)}{\left(1 - e^{-\theta_1 x_0}\right)^{i-1}}.$$

Therefore

$$\sum_{A \subseteq [n],\ |A| < i} \left(\prod_{j \in A} e^{-\theta_j x}\right) \left(\prod_{j \in [n] \setminus A} \left(1 - e^{-\theta_j x}\right)\right)$$

$$\leq \frac{g_n(x)}{(1-e^{-\theta_1 x_0})^{i-1}} \sum_{A\subseteq[n],\ |A|<i} \left(\prod_{j\in A} e^{-\theta_j x}\right)$$
$$\leq \frac{g_n(x)(1+f_n(x)+f_n(x)^2+\cdots+f_n(x)^{i-1})}{(1-e^{-\theta_1 x_0})^{i-1}}.$$

By the inequality $1-x\leq e^{-x}$, we have $g_n(x)\leq e^{-f_n(x)}$. Thus,

$$P(\sigma_n(i)\geq a)\leq \frac{e^{-f_n(x)}(1+f_n(x)+f_n(x)^2+\cdots+f_n(x)^{i-1})}{(1-e^{-\theta_1 x_0})^{i-1}} + \sum_{j\geq a} e^{-\theta_j x}.$$

Let m be the largest integer such that $\theta_m \leq 1/(x-x_0)$. Suppose that $n\geq m$. Then

$$f_n(x)\geq \sum_{j=1}^{m} e^{-\theta_j x} = \sum_{j=1}^{m} e^{-\theta_j x_0-\theta_j(x-x_0)} \geq e^{-1}\sum_{j=1}^{m} e^{-\theta_j x_0}.$$

But $m\to\infty$ as $x\searrow x_0$, and $f(x_0)=\infty$ by assumption. Thus, the above inequality shows that given any $L>0$, we can first choose x sufficiently close to x_0, and then choose n_0 sufficiently large, such that for all $n\geq n_0$, $f_n(x)\geq L$. Now take any $\varepsilon>0$ and find L so large that for all $y\geq L$,

$$\frac{e^{-y}(1+y+y^2+\cdots+y^{i-1})}{(1-e^{-\theta_1 x_0})^{i-1}} < \frac{\varepsilon}{2}.$$

Choose x and then n_0 as in the previous paragraph corresponding to this L. Then find a so large that

$$\sum_{j\geq a} e^{-\theta_j x} < \frac{\varepsilon}{2},$$

which exists since $f(x)<\infty$. For this choice of a, the above steps show that $P(\sigma_n(i)\geq a)\leq\varepsilon$ for all $n\geq n_0$. This proves tightness of $\{\sigma_n(i)\}_{n\geq 1}$ when $x_0>0$, completing the proof of the theorem. □

5. A vast generalization – hyperplane walks

5.1. Introduction

The Tsetlin library has seen vast generalizations in the past twenty years. In this section, we explain walks on the chambers of a hyperplane arrangement due to Bidigare–Hanlon–Rockmore [9] and Brown–Diaconis [13]. The Tsetlin library is a (very) special case of the braid arrangement. These Markov chains have a fairly complete theory (simple

forms for the eigenvalues and good rates of convergence to stationarity). But the description of the stationary distribution, the analog of the Luce model, is indirect, involving a weighted sampling without replacement scheme. Thus the problem

What does the stationary distribution of hyperplane walks look like?

Section 5.2 sets things up and states the main theorems (with examples). The few cases where something is known are reported in Section 5.3. The final section points to semigroup walks where parallel problems remain open. The main point of this section is to cast Sections 2–4 above as contributions to a general problem.

5.2. Hyperplane walks

We work in $\mathbb{R}^d$. Let $\mathcal{A} = \{H_1, H_2, \ldots, H_k\}$ be a finite collection of affine hyperplanes (translates of codimension one subspaces). These divide $\mathbb{R}^d$ into

- chambers (points not on any H_i). Let $\mathcal{C}$ be the chambers.
- faces (points on some H_i and on one side or another of others). Let $\mathcal{F}$ be the faces.

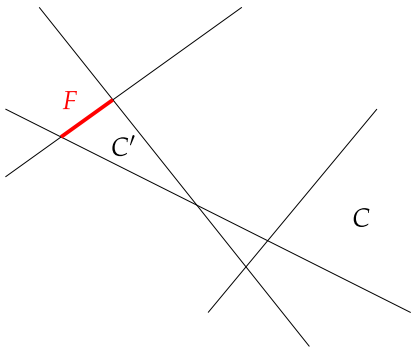


Fig. 1. Four lines in $\mathbb{R}^2$. There are 10 chambers and 30 faces (chambers, points of intersection and the empty face are faces).

A key notion is the projection of a chamber onto a face (Tits projection). For $C \in \mathcal{C}$ and $F \in \mathcal{F}$, PROJ $C \to F$ is the unique chamber adjacent to F and closest to C (in the sense of crossing the fewest number of H_i's). In the Fig. 1, PROJ $C \to F = C'$.

With these definitions, we are ready to walk. Choose face weights $\{w_F\}_{F \in \mathcal{F}}$ with $w_F \geq 0$ and $\sum_{F \in \mathcal{F}} w_F = 1$. Define a Markov chain $\kappa(C, C')$ on chambers via:

- from C, choose $F \in \mathcal{F}$ with probability w_F and move to PROJ $C \to F$.

Thus, $\kappa(C, C') = \sum_{\text{PROJ } C \to F = C'} w_F$.

Example 5.1 *(Boolean arrangements).* Let $H_i = \left\{x \in \mathbb{R}^d : x_i = 0\right\}, 1 \leq i \leq d$ be the usual coordinate hyperplanes. These divide $\mathbb{R}^d$ into 2^d chambers (the usual orthants) and 3^d faces. A face may be labeled by a vector of length d containing $0, \pm 1$ to delineate 0 (on) or on one side or the other of the ith hyperplane. Chambers are faces with no zeros. For PROJ $C \to F = C'$, set the ith coordinate of C' to the ith coordinate of F if this is ± 1 and leave it as the ith coordinate of C if the ith coordinate of F is 0.

Thus, a walk proceeds via: from C, pick a subset of coordinates and install ± 1 in them as determined by F. For example, if

$$w_F = \begin{cases} \dfrac{1}{2d} & F = (0, \ldots, 0, \pm 1, 0, \ldots, 0), \\ 0 & \text{otherwise,} \end{cases}$$

the walk becomes "pick a coordinate at random and replace it with ± 1 chosen uniformly." This is the celebrated Ehrenfest urn model of statistical physics. Dozens of natural specializations of these Boolean walks are spelled out in [13].

Example 5.2 *(Braid arrangement).* Take $H_{ij} = \left\{x \in \mathbb{R}^d : x_i = x_j\right\}, 1 \leq i < j \leq d$. Now, the chambers are points in $\mathbb{R}^d$ with no equal coordinates. It follows that the relative order is fixed within a chamber, so chambers can be labeled by permutations. The faces are indexed by "block ordered set partitions": coordinates within a block are equal and all coordinates in the first block are smaller than the coordinates in the second block, and so on.

For the projection, suppose the chamber labeled π is thought of as a deck of cards in arrangement π (with $\pi(i)$ the label of the card at position i). Suppose $d = 5$ and the face is $F = 1\,3/2/4\,5$. Remove cards labeled 1 and 3 from π (keeping them in their same relative order, then remove the card labeled 2 and place it under cards 1, 3. Finally, remove cards labeled 4, 5 and place them at the bottom of the five card deck. This is PROJ $\pi \to 1\,3/2/4\,5$.

The **Tsetlin library** arises from the choice

$$w_F = \begin{cases} \theta_i & \text{if } F = i/[n] \setminus i \\ 0 & \text{otherwise} \end{cases}.$$

That is the walk on S_n with "choose label i with probability θ_i and move this card to the top."

Riffle shuffling arises from

$$w_F = \begin{cases} \dfrac{1}{2^d} & \text{if } F = S,\ [d] - S \text{ where } S \text{ is not equal to } \emptyset \text{ or } [d] \\ \dfrac{1}{2^{d-1}} & \text{if } F = [d] \\ 0 & \text{otherwise} \end{cases}.$$

Another way to say this – label each of d cards in the current deck with a fair coin flip, remove all cards labeled "heads" keeping them in their same relative order, and place them on top. This is exactly "inverse riffle shuffling," the inverse of the Gilbert–Shannon–Reeds model studied by Bayer–Diaconis [5].

There are hundreds of other hyperplane arrangements where the chambers are labeled by natural combinatorial objects, and there are choices of face weights so that the walk is a natural object ot study. Indeed, any finite reflection group leads to a hyperplane arrangement with $H_{\mathbf{v}}$ being the hyperplane orthogonal to the vector $\mathbf{v}$ determining the reflection. Any finite graph leads to a "graphical arrangement." For a wonderful exposition, see Stanley [45].

As said, the Markov chains $\kappa(C, C')$ admit a complete theory with known eigenvalues and rates of convergence. We will not spell this out here; see [13], but turn to the main object of interest – the stationary distribution.

Let $\mathcal{A}$ be a general arrangement with chosen face weights $\{w_F\}_{F\in\mathcal{F}}$ and $\kappa(C, C')$ the associated Markov chain on C, the chambers of the arrangement. $\pi(C) \geq 0$ and $\sum_C \pi(C) = 1$ is stationary for κ if $\sum_C \pi(C)\kappa(C, C') = \pi(C')$ – thus π can be thought of as a left eigenvector with eigenvalue 1. When does a unique such π exist?

Theorem 5.3 *(Brown–Diaconis). Call $\{w_F\}$* ***separating*** *if they are not all supported in the same hyperplane (for $H \in \mathcal{A}$, there exists $H' \in \mathcal{A}$ and $w_F > 0$ for $F \subset H'$). Then κ has a unique stationary distribution $\pi(C)$ if and only if $\{w_F\}$ are separating.*

This π is the analog of the Luce model and becomes the Luce model for the braid arrangement as above. The following result gives a "weighted sampling without replacement characterization" of $\pi(C)$.

Theorem 5.4 *(Brown–Diaconis). Suppose $\{w_F\}$ are separating. The following algorithm generates a pick from $\pi(C)$:*

- *place all $\{w_F\}$ in an urn.*
- *draw them out, without replacement, with probability proportional to size (relative to what is left).*
- *say this results in the ordered list $F_1, F_2, \ldots, F_{|\mathcal{F}|}$.*
- *from any starting chamber C (the choice does not matter), project on $F_{|\mathcal{F}|}$, then on $F_{|\mathcal{F}|-1}$, and so on until F_1. The resulting chamber is exactly distributed as $\pi(C)$.*

Of course, for the Tsetlin library, this is just the Luce measure on permutations. The following subsection delineates the few examples where something can be said about π.

5.3. Understanding π

Suppose a group of orthogonal transformations acts transitively on $\mathcal{A}$ preserving $\kappa(C, C')$. Then, $\pi(C)$ is uniform over $\mathcal{C}$ (supposing separability). Examples include riffle shuffles, the Ehrenfest urn, and "random to top" (the Tsetlin library with $\theta_i = \frac{1}{n}, 1 \leq i \leq n$). For more on this, see [39].

Simple features of π can sometimes be calculated directly. See Pike [40] and its references.

Aside from the present paper, the only other examples that have been carefully studied are in the following graph coloring problems.

5.3.1. Graph coloring

Let G be a connected and undirected simple graph. Let $\mathcal{X}$ be the set of 2-colorings (say by $\pm$) of the vertex set of G. Define a Markov chain on $\mathcal{X}$ by

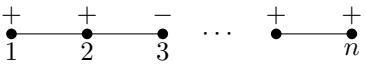
- from $x \in \mathcal{X}$
- pick an edge $e \in G$ uniformly at random
- change the two endpoints of e in x to be  or  with probability $\frac{1}{2}$.

Thus "neighbors are inspired to match, at random times." This is a close cousin of standard particle systems such as the voter model. All the theory works. The process is a hyperplane walk for the Boolean arrangement of dimension D, where D denotes the number of edges in the graph G. All eigenvalues and rates of convergence are easily available.

The only thing open is

"what can be said about the stationary distribution?"

To understand the question, suppose the graph is an n-point path



The distribution π is far from uniform. All $+$ or all $-$ have chance $\frac{1}{2}$ of staying, but $+-+-\cdots$ is impossible. Of course, $\pi(x)$ is invariant under switching $+$ and $-$. It is easy to show that, under π, the π process is a 1-dependent point process (see [11]). This means various central limit theorems are available.

How much more likely is "all $+$" than "many alternations"? This problem was carefully studied in a difficult paper by Chung and Graham [19] (see also [14]). They show, under

π, all + (or all −) have chance of order $C/2^n$, but many alternations has chance of order $C'/n!$. Very nice systems of recursive differential equations appear.

The point is, even in the simplest case, understanding the stationary distribution leads to interesting mathematics. We offer the present paper in this spirit.

5.4. Semigroups and beyond

The past ten years have shown yet broader generalization of the Tsetlin library. Kenneth Brown extended it to idempotent semigroups (allowing walks on the chambers of a building) [12]. Ben Steinberg, working with many coauthors, extended further in the semigroup direction. A convenient reference is the book length treatment [38].

In another direction, a sweeping generalization of much of modern algebra based on hyperplane and semigroup walks has been developed by Aguiar and Mahajan [1–3]. The three large volumes contain hundreds of fresh examples.

In *none* of these developments is the stationary measure understood.

Data availability

No data was used for the research described in the article.

Acknowledgment

Thanks to Jim Pitman and Jason Fulman for useful references.

References

[1] M. Aguiar, S. Mahajan, Monoidal Functors, Species and Hopf Algebras, AMS, 2010.
[2] M. Aguiar, S. Mahajan, Topics in Hyperplane Arrangements, AMS, 2017.
[3] M. Aguiar, S. Mahajan, Bimonoids for Hyperplane Arrangements, Cambridge University Press, 2020.
[4] A. Barbour, L. Holst, S. Janson, Poisson Approximation, Oxford Press, 1992.
[5] D. Bayer, P. Diaconis, Trailing the dovetail shuffle to its lair, Ann. Appl. Probab. 2 (1992) 294–313.
[6] A. Ben-Hamou, Y. Peres, J. Salez, Weighted sampling without replacement, Braz. J. Probab. Stat. 32 (2018) 657–669.
[7] G. Benkart, D. Britten, F. Lemire, Stability in Modules for Classical Lie Algebras: a Constructive Approach, AMS, 1990.
[8] G. Benkart, P. Diaconis, M. Liebeck, P. Tiep, Tensor product Markov chains, J. Algebra 561 (2020) 17–83.
[9] P. Bidigare, P. Hanlon, D. Rockmore, A combinatorial description of the spectrum for the Tsetlin library and its generalization to hyperplane arrangements, Duke Math. J. 99 (1999) 135–174.
[10] P. Billingsley, Probability and Measure, 3rd ed., Wiley, 2012.
[11] A. Borodin, P. Diaconis, J. Fulman, On adding a list of numbers (and other one-dependent determinantal processes), Bull. Am. Math. Soc. (N.S.) 47 (2010) 639–670.
[12] K. Brown, Semigroups, rings and Markov chains, J. Theor. Probab. 13 (2000) 871–938.
[13] K. Brown, P. Diaconis, Random walks and hyperplane arrangements, Ann. Probab. 26 (1998) 1813–1854.
[14] S. Butler, F. Chung, J. Cummings, R. Graham, Edge flipping in the complete graph, Adv. Appl. Math. 69 (2015) 46–64.

[15] M. Camarri, J. Pitman, Limit distributions and random trees derived from the birthday problem with unequal probabilities, Electron. J. Probab. 5 (2000) 1–18.
[16] S. Chatterjee, P. Diaconis, E. Meckes, Exchangeable pairs and Poisson approximation, Probab. Surv. 2 (2005) 64–106.
[17] L. Coregliano, A. Razborov, Semantic limits of dense combinatorial objects, Russ. Math. Surv. 75 (2020) 627–723.
[18] H. Crane, The ubiquitous Ewens sampling formula, Stat. Sci. 31 (2016) 1–19.
[19] F. Chung, R. Graham, Edge flipping in graphs, Adv. Appl. Math. 48 (2012) 37–63.
[20] P. Diaconis, The cutoff phenomenon in finite Markov chains, Proc. Natl. Acad. Sci. USA 93 (1996) 1659–1664.
[21] P. Diaconis, Patterns in eigenvalues: the 70th Josiah Willard Gibbs lecture, Bull. Am. Math. Soc. (N.S.) 40 (2003) 155–178.
[22] P. Diaconis, S. Ethier, Gambler's ruin and the ICM, Stat. Sci. 37 (2022) 289–305.
[23] P. Diaconis, M. Simper, Statistical enumeration of groups by double cosets, J. Algebra 607 (2022) 214–246.
[24] R. Dobrow, J. Fill, The Move-to-Front Rule for Self-Organizing Lists with Markov Dependent Requests, Discrete Probability and Algorithms, Springer, 1995.
[25] P. Donnelly, The heaps process, libraries, and size-biased permutations, J. Appl. Probab. 28 (1991) 321–335.
[26] S. Evans, R. Rivest, P. Stark, Leading the field: fortune favors the bold in Thurstonian choice models, Bernoulli 25 (2019) 26–46.
[27] W. Feller, An Introduction to Probability Theory and its Applications, vol. II, 3rd ed., Wiley, New York, 1971.
[28] J. Fill, Limits and rates of convergence for the distribution of serach cost under the move-to-front rule, Theor. Comput. Sci. 164 (1996) 185–206.
[29] J. Fill, L. Holst, On the distribution of search cost for the move-to-fron rule, Random Struct. Algorithms 8 (1996) 179–186.
[30] J. Fulman, Random matrix theory over finite fields, Bull. Am. Math. Soc. (N.S.) 39 (2001) 51–85.
[31] L. Gordon, Successive sampling in large finite populations, Ann. Stat. 11 (1983) 702–706.
[32] D. Harville, Assigning probabilities to the outcomes of multi-entry competitions, J. Am. Stat. Assoc. 68 (1973) 312–316.
[33] D. Hausch, V. Lo, W. Ziemba, Efficiency of Racetrack Betting Markets, World Scientific, 2008.
[34] C. Hoppen, Y. Kohayakawa, C. Moreira, B. Rath, R. Sampaio, Limits of permutation sequences, J. Comb. Theory, Ser. B 103 (2013) 93–113.
[35] D. Horvitz, D. Thompson, A generalization of sampling without replacement from a finite universe, J. Am. Stat. Assoc. 47 (1952) 663–685.
[36] D. Luce, Individual Choice Behavior: a Theoretical Analysis, Wiley, 1958.
[37] D. Luce, The choice axiom after twenty years, J. Math. Psychol. 15 (1977) 215–233.
[38] S. Margolis, F. Saliola, B. Steinberg, Cell Complexes, Poset Topology and the Representation Theory of Algebras Arising in Algebraic Combinatorics and Discrete Geometry, AMS, 2021.
[39] E. Nestoridi, Optimal strong stationary times for random walks on the chambers of a hyperplane arrangement, Probab. Theory Relat. Fields 174 (2019) 929–943.
[40] J. Pike, Eigenfunctions for random walks on hyperplane arrangements, Ph.D. thesis, USC, 2013.
[41] R. Pyke, Spacings revisited, Berkeley Symp. Math. Statist. Prob. 6 (1972) 417–427.
[42] R. Phatarfod, On the matrix occurring in a linear search problem, J. Appl. Probab. 28 (1991) 336–346.
[43] J. Pitman, N. Tran, Size-biased permutation of a finite sequence with independent and identically distributed terms, Bernoulli 21 (2015) 2484–2512.
[44] B. Rosen, Asymptotic theory for successive sampling with varying probabilities without replacement, II., Ann. Math. Stat. 43 (1972) 373–397.
[45] R. Stanley, An introduction to hyperplane arrangements, in: Geometric Combinatorics, AMS, 2007, pp. 389–496.
[46] P. Sukhatme, Tests of significance for samples of the chi square population with two degrees of freedom, Ann. Eugen. 8 (1937) 52–56.
[47] M. Tsetlin, Some problems in the behavior of finite automata, Dokl. Akad. Nauk SSSR 139 (1961) 830–833.
[48] J. Yellott, The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution, J. Math. Psychol. 15 (1977) 109–144.