Sampling Young Tableaux and Contingency Tables

Samuel King
Advised by Daniel Štefankovič and Sevak Mkrtchyan
Based on work previously completed with Mandar Juvekar

1 Contingency Tables and Markov Chains

The problem motivating this work is that of sampling contingency tables. A contingency table A is a matrix with nonnegative integer entries whose rows and columns sum to some specified values. In other words, given vectors $r = (r_i)_{i=1}^m$ and $c = (c_j)_{j=1}^n$ of positive integers, a contingency table with row sums r and column sums c is some $A \in \mathrm{Mat}_{m\times n}(\mathbb{N})$ such that

$$r_i = \sum_{t=1}^{n} A_{i,t} \quad \text{and} \quad c_j = \sum_{t=1}^{m} A_{t,j} \tag{1}$$

for each $i \in [m]$ and $j \in [n]$ (we use the notation $[k] = \{1, \ldots, k\}$). Notice that it must be that $\sum r_i = \sum c_j$ for such a contingency table to possibly exist. We are motivated by the problem of sampling uniformly from $\Omega(r, c)$, the space of all contingency tables with row sums r and column sums c.

Throughout the majority of this discussion, the primary method of sampling will be using Markov chains. The definitions and main results dealing with Markov chains are primarily acquired from Mark Jerrum's textbook *Counting, Sampling and Integrating: Algorithms and Complexity* [9]. A (finite-space) *Markov chain* is a sequence of random variables $X_0, \ldots, X_t, \ldots$ with values in a finite state space Ω such that

$$\Pr(X_t = a \mid X_0 = a_0, X_1 = a_1, \ldots, X_{t-1} = a_{t-1}) = \Pr(X_t = a \mid X_{t-1} = a_{t-1}) \tag{2}$$

for all $t \in \mathbb{N}$ and $a_0, \ldots, a_t \in \Omega$. Specifically, we consider *time-homogeneous* Markov chains, where the transition probability is independent of t. Thus, we may define

$$P(a, b) := \Pr(X_t = b \mid X_{t-1} = a) \tag{3}$$

and call P the Markov chain's *transition matrix*. It can be easily seen that for any $t \in \mathbb{N}$,

$$\Pr(X_t = b \mid X_0 = a) = P^t(a, b). \tag{4}$$

We define the *period* of state i to be $p_i := \gcd\{t \mid P^t(i, i) > 0\}$. We then say that a Markov chain is *aperiodic* if there exists a state i such that $p_i = 1$. We say a Markov chain is *irreducible* or *connected* if for every two states $i, j \in \Omega$, there exists some $t \in \mathbb{N}$ such that $P^t(i, j) > 0$. A Markov chain that is both aperiodic and irreducible is said to be *ergodic*.

For any probability distribution μ_t over the elements of Ω, we may think of $\mu_{t+1} = \mu_t P$ as the distribution acquired after performing a transition of the Markov chain on μ_t. We say that a distribution π on Ω is a *stationary distribution* if $\pi = \pi P$. With these definitions, we can state a fundamental result.

Theorem 1. *An ergodic Markov chain has a unique stationary distribution π; moreover, the Markov chain tends to π in the sense that $P^t(a, b) \to \pi_b$ as $t \to \infty$, for all $a, b \in \Omega$.*

This result implies that an ergodic Markov chain can be used to sample elements from a distribution close to π over Ω. We can start with any element $a \in \Omega$ and transition to other elements according to the rules defined by the Markov chain. How close the ending distribution is to π is dependent on t, the number of transitions and the transition matrix. In general, the more transitions we perform, the closer the distribution gets to π. Thus, a Markov chain is useful for sampling if its stationary distribution matches the desired distribution over Ω we wish to sample from and if it can quickly converge to the stationary distribution.

One useful result in computing the stationary distribution of a Markov chain is the following:

Theorem 2. *Suppose P is the transition matrix of a Markov chain. If the function $\pi' : \Omega \to [0, 1]$ satisfies*

$$\pi'(x)P(x, y) = \pi'(y)P(y, x) \tag{5}$$

for all $x, y \in \Omega$, and

$$\sum_{x \in \Omega} \pi'(x) = 1,$$

then π' is a stationary distribution of the Markov chain.

The condition described in eq. (5) is called the *detailed balance condition*, and a Markov chain whose stationary distribution satisfies this condition is said to be *reversible*.

To measure how far two given distributions are from each other, we define the *total variation distance* between the distributions μ and ν as

$$\mathrm{d_{TV}}(\mu, \nu) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = \frac{1}{2} ||\mu - \nu||_1. \tag{6}$$

From this, we can talk about the important notion of the *mixing time* $\tau(\epsilon)$ of an ergodic Markov chain, defined as

$$\tau(\epsilon) = \min\{t \mid \forall i \in \Omega,\ \mathrm{d_{TV}}(e_i P^t, \pi) \leq \epsilon\}, \tag{7}$$

where e_i is the distribution with $e_i(i) = 1$ and $e_i(j) = 0$ for $j \neq i$.

The problem of sampling using Markov chains is therefore the problem of designing an ergodic Markov chain on a given state space that has the desired stationary distribution (e.g. if you wish to sample uniformly from Ω, then π must be the uniform distribution) and which efficiently reaches the stationary distribution. Traditionally, we take "efficiently reaching the stationary distribution" or "the Markov chain mixes fast" to mean that the mixing time $\tau(\epsilon)$ is bounded by a polynomial in $1/\epsilon$ and any parameters of the given problem (e.g. m and n in the problem of sampling contingency tables).

1.1 Previous work in sampling contingency tables

In 2002, Morris [16] showed that there exists an efficient (polynomial time) algorithm to sample contingency tables provided that the row and column sums are sufficiently large, specifically $r_i = \Omega(n^{3/2} m \log m)$ and $c_j = \Omega(m^{3/2} n \log n)$. The algorithm they present is based on sampling continuously from the interior of a convex polytope acquired from the row and column sums.

Diaconis and Gangolli [3] present the following Markov chain for sampling contingency tables. Given a contingency table A, select a pair of distinct rows $i_1 \neq i_2$ uniformly from the set of all such pairs and select a pair of distinct columns $j_1 \neq j_2$ from the set of all such pairs. Then, create a new contingency table A' whose entries are the same as those in A except for the following adjustments:

$$A'_{i_1,j_1} = A_{i_1,j_1} + 1 \qquad A'_{i_1,j_2} = A_{i_1,j_2} - 1$$
$$A'_{i_2,j_1} = A_{i_2,j_1} - 1 \qquad A'_{i_2,j_2} = A_{i_2,j_2} + 1$$

The Markov chain then transitions from A to A' if the entries in A' are non-negative. Although it is easy to describe, researchers have had a difficult time analyzing the mixing time of this Markov chain.

In a paper from 2017, Kayibi et al. [12] attempt to show that this Markov chain mixes fast by using a canonical path argument. However, we managed to find a counterexample to one of the results used in the argument. Specifically, the paper states the following claim:

Proposition 3 (Corollary 8 from [12])**.** *Let N be the number of all $m \times n$ contingency tables of fixed row and column sums. The number of contingency tables having k fixed cells (in lexicographic ordering) is at most $N^{\frac{mn-k}{mn}}$.*

A counterexample to this proposition can be seen as follows: fix $n \in \mathbb{Z}^+$ and let $r = c = (1, 1, \ldots, 1)$. Then the set of $n \times n$ contingency tables with these row and column sums is exactly the set of tables acquired by permuting the rows of the $n \times n$ identity matrix. So in this case, $N = n!$. The set of these contingency tables with the first cell fixed to be 1 is the set of tables acquired by permuting the last $n-1$ rows of the identity matrix, so there are $(n-1)!$ such contingency tables. Consequently, there are $n! - (n-1)!$ tables with 0 fixed as the first entry. The greater of these two values is $n!-(n-1)!$. We used $k = 1$, so the bound given by the proposition is $N^{\frac{mn-k}{mn}} = (n!)^{\frac{n^2-1}{n^2}}$. However, if we take, for example, $n = 7$, our bound is $(7!)^{\frac{7^2-1}{7^2}} < 4235$ whereas the actual number of contingency tables with one fixed entry is $7! - 6! = 4320$. This shows that the proposition does not hold in general and that there must exist a breakdown in their argument. However, we can see that, using Stirling's approximation for $n!$,

$$\lim_{n\to\infty} \frac{n! - (n-1)!}{(n!)^{\frac{n^2-1}{n^2}}} = \lim_{n\to\infty} \frac{(n-1)!(n-1)}{n!}(n!)^{\frac{1}{n^2}} \tag{8}$$

$$= \lim_{n\to\infty} \left(1 - \frac{1}{n}\right)\left(\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\right)^{\frac{1}{n^2}} \tag{9}$$

$$= \left(\lim_{n\to\infty} 1 - \frac{1}{n}\right)\left(\lim_{n\to\infty} (2\pi n)^{\frac{1}{2n^2}}\right)\left(\frac{\lim_{n\to\infty} n^{\frac{1}{n}}}{\lim_{n\to\infty} e^{\frac{1}{n}}}\right) \tag{10}$$

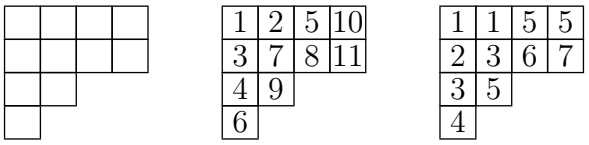
$$= 1. \tag{11}$$

Thus, it is unclear, given only this counterexample, whether this bound is only off by a constant or whether there is a more fundamental error in the argument provided by Kayibi et al. A more in-depth analysis of [12] is needed.

2 Young Tableaux

We now turn our attention to the study of Young tableaux. Young tableaux, as we illustrate below, are innately connected to contingency tables, so studying how to sample these objects may provide us with a method of sampling contingency tables.

A *Young diagram* (sometimes called a Ferrers diagram) of shape $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ is an array of k left-justified rows of cells such that row i has length λ_i and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 1$. A *standard Young tableau* of shape λ is a Young diagram filled with the integers from $[n]$ where $n = \sum_{i=1}^{r} \lambda_i$ such that the integers are strictly increasing both from left to right within each row and from top to bottom within each column (so each integer from $[n]$ appears precisely once). A *semistandard Young tableau* is a generalization in which the integers from $[n]$ are allowed to appear more than once, and the row condition is relaxed to require that integers are only weakly increasing from left to right. Such a tableau is said to have weight $\mu = (\mu_1, \ldots, \mu_n)$ if each integer i appears μ_i times. A standard Young tableau could be considered a semistandard Young tableau with weight $\mu = (1, 1, \ldots, 1)$. The following from left to right are examples of a Young diagram, a standard Young tableau, and a semistandard Young tableau, each with shape $\lambda = (4, 4, 2, 1)$:



(12)

2.1 Connection to contingency tables

For a given shape λ and a given weight μ, the number of semistandard Young tableaux of shape λ and weight μ is called the *Kostka number* $K_{\lambda,\mu}$. Young tableaux and Kostka numbers have connections to many other areas of mathematics such as representation theory, symmetric functions, and the study of longest increasing subsequences (see e.g. [14], [15], [19], [20]). We are most interested in the connection between Young tableaux and contingency tables through the Robinson-Schensted-Knuth (RSK) correspondence. This correspondence provides a bijection between the set of contingency tables with specified row sums r and column sums c and a set of pairs of semistandard Young tableaux of the same shape, one with weight r and one with weight c.

The RSK correspondence works by first giving a injection from contingency tables to two-lines arrays whose columns are in lexicographic order. Let A be a contingency table with row sums $r = (r_1, \ldots, r_m)$ and column sums

$c = (c_1, \ldots, c_n)$. Let $N = \sum r_i = \sum c_j$. We will create a $2 \times N$ array B by adding a total of $A_{i,j}$ columns $\binom{i}{j}$ to B. Placing these columns in lexicographic order (i.e. column $\binom{i}{j}$ comes before column $\binom{i'}{j'}$ if $i < i'$; if $i = i'$, then $\binom{i}{j}$ comes before column $\binom{i'}{j'}$ if $j < j'$) yields a unique two-line array. The resulting array will have r_i i's in its first row and c_j j's in its second row. For example, the following contingency table maps to the given two-line array:

$$\begin{bmatrix} 1 & 0 & 1 \\ 3 & 2 & 2 \\ 0 & 0 & 4 \end{bmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & 3 \\ 1 & 3 & 1 & 1 & 1 & 2 & 2 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \end{pmatrix} \tag{13}$$

This two-line array can then be read column-by-column from left to right to simultaneously build two semistandard Young tableaux of the same shape using the Schensted "row-bumping" algorithm. The shape of the tableaux may be different for different contingency tables. The details of this process is outside the scope of this work, but a thorough explanation of this algorithm and the RSK correspondence can be found in [6] and [13].

This correspondence provides a possible route of sampling contingency tables through the sampling of semistandard Young tableaux. That is, if we can determine how to efficiently sample pairs of Young tableaux of the same shape with weights r and c, then we can use the RSK correspondence to efficiently acquire a sampled contingency table with row sums r and column sums c.

2.2 Connection between sampling and counting

The problem of sampling is innately connected to the problem of counting when it comes to Young tableaux. If we have an efficient method of counting the number of Young tableaux with any given shape and weight, then we can efficiently sample Young tableaux of a given shape and weight. The other direction holds as well with a slight caveat; if we have an efficient method of sampling Young tableaux of any given shape and weight, then we can efficiently *approximate* the number of Young tableaux with a given shape and weight Because our work focuses on sampling Young tableaux, we will specify how counting would give us sampling, but the other direction uses the same general idea but in reverse.

To see that efficient counting implies efficient sampling, assume that we have an efficient method of calculating $K_{\lambda,\mu}$ for any λ and μ. Now fix some shape $\lambda = (\lambda_j)_{j=1}^k$ and weight $\mu = (\mu_i)_{i=1}^n$, and let a be the largest entry in a tableau with weight μ, that is, $a = \max\{i \mid \mu_i > 0\}$. Let Ω be the set of all semistandard Young tableaux with shape λ and weight μ. We will sample

a random Young tableau from Ω by randomly selecting the location of the largest entry and recursively filling out the rest of the tableau.

A *corner* of a Young diagram is a cell at the end of both its row and its column. Let c be the number of corners of λ, and let r_t be the row on which the t^{th} corner lies. Let $\lambda^t = (\lambda_j^t)_{j=1}^k$ be the shape derived from λ by removing the t^{th} corner, so

$$\lambda_j^t = \begin{cases} \lambda_j & \text{if } j \neq r_t \\ \lambda_j - 1 & \text{if } j = r_t \end{cases}. \tag{14}$$

Additionally, let $\mu' = (\mu'_i)_{i=1}^n$ be the weight derived from μ by removing one count of the entry a, so

$$\mu'_i = \begin{cases} \mu_i & \text{if } i \neq a \\ \mu_i - 1 & \text{if } i = a \end{cases}. \tag{15}$$

Since a is the largest entry in μ, it must be located on a corner of any Young tableau with weight μ. Given the t^{th} corner of λ, we can describe the probability that a is located at that corner of a Young tableau uniformly selected from Ω by

$$p_t = \frac{K_{\lambda^t,\mu'}}{K_{\lambda,\mu}}. \tag{16}$$

But if we can efficiently calculate Kostka numbers, then we can efficiently calculate p_t. Thus, we can efficiently determine the correct probabilities with which to place a on a corner of our Young tableau. After placing a on a corner, the remainder of the tableau is just a smaller Young tableau, so we can recursively perform the same procedure to fill it out as well. When we have finished filling it out, the resulting tableau is one that has been uniformly selected from Ω.

3 Sampling Standard Young Tableaux

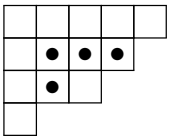
We first discuss the case of sampling standard Young tableaux. These tableaux have weight $\mu = (1, 1, \ldots, 1)$, and the RSK correspondence creates a bijection from contingency tables with row sums r and column sums c to pairs of Young tableaux with weights r and c. Thus, the ability to sample standard Young tableaux could provide us a way of sampling contingency tables with all row and column sums equal to 1. Specifically, we would need to be able to sample from the set of all Young tableaux with weight μ. This could be done, for example, by first sampling a Young diagram with shape λ with respect to the

Plancherel measure (for example, see [11]) and then sampling a Young tableau with shape λ. However, this special case of sampling contingency tables, as discussed in Section 1.1, is not particularly interesting, as the set of contingency tables with all row and column sums equal to 1 is just the set of square matrices whose rows are a permutation of the identity matrix of the same size, an easy set to sample from. However, we study the case of sampling standard Young tableaux both because it is mathematically interesting in its own right and also because understanding this special case may help us understand the more general problem of sampling semistandard Young tableaux.

3.1 Previous work in sampling standard Young tableaux

Mathematicians have long known formulas for the number of standard Young tableaux of a given shape. Frobenius in 1900 [5] and Alfred Young (for whom the tableaux are named) in 1902 [21] were the first to develop such a formula although it was not convenient to work with. Later, in 1954, Frame, Robinson, and Thrall [4] found a much simpler formula which is now known as the *hook length formula*.

Given a Young diagram of shape λ, the *hook* $H_\lambda(i,j)$ of the cell (i,j) is the set of cells to the right of (i,j) in the same row and the cells below (i,j) in the same column, along with (i,j) itself. For example, the hook of $(2,2)$ is marked in the following diagram:



(17)

The hook length, denoted $h_\lambda(i,j)$, is the number of cells in $H_\lambda(i,j)$. So, in eq. (17) above, $h_\lambda(2,2) = 4$. If we let f_λ be the number of Young tableaux of shape λ, then the following result states the hook length formula, established in [4].

Theorem 4. *Given a shape λ with size n,*

$$f_\lambda = \frac{n!}{\prod h_\lambda(i,j)}, \tag{18}$$

where the product is over all cells of the Young diagram with shape λ.

The hook length formula is an easy formula to compute. Therefore, using the connection between counting and sampling discussed in Section 2.2, we can use it to efficiently sample standard Young tableaux.

In 1979, Greene, Nijenhuis, and Wilf [7] provided an alternative probabilistic proof of the hook length formula. Their goal in reproving the result was to establish a better combinatorial explanation of why the hooks appear in the formula, as the proof provided by Frame, Robinson, and Thrall allows for no intuitive explanation. A convenient product of their new proof is that it gives an alternative and more efficient method of sampling standard Young tableaux, described as follows.

Given a Young diagram of shape λ with size n, randomly select a cell (i, j) with uniform probability $1/n$. Then, randomly select a new cell (i', j') from $H_\lambda(i, j) \setminus \{(i, j)\}$ with uniform probability $1/(h_\lambda(i, j) - 1)$. Select another cell from $H_\lambda(i', j') \setminus \{(i'j')\}$, and continue in the same manner until a corner is selected. Label this corner (α, β) and call this full process a single trial. To sample a standard Young tableau with shape λ, we will perform n trials. At the end of trial j, we will fill in the final cell (α, β) with the integer $n - j + 1$ and remove the cell from the diagram. At the end of the n trials, we will have filled out our original diagram completely and therefore have sampled a Young tableau. The probabilistic proof in [7] ensures that our tableau is selected from the uniform distribution over the set of standard Young tableaux of shape λ.

Each trial in the above algorithm will select a sequence of at most n cells, and n trials are performed. Thus, the algorithm has a time complexity of $O(n^2)$ and, hence, is an efficient way of sampling standard Young tableaux.

3.2 A proposed Markov chain

Although a process for sampling uniformly from the set of standard Young tableaux of a given shape λ already exists, we wish now to explore an alternative approach of algorithmic sampling by using a Markov chain. We do this because we hope that sampling via Markov chains will be fruitful in the semistandard case, and exploring this sampling method in the simpler case of sampling standard Young tableaux may aid our understanding of sampling in the semistandard case. Fix $\lambda = (\lambda_r)_{r=1}^k$ with $\lambda_1 \geq \cdots \geq \lambda_k$, and let $n = \sum \lambda_i$. Let Ω be the set of all standard Young tableaux with shape λ. We wish to sample uniformly from Ω.

We propose the following Markov chain $\mathcal{MC}_{\text{swap}}$ on Ω. Given $T \in \Omega$, let $T[i, j]$ with $i, j \in [n]$ be the table acquired from switching the locations of i and j in T. Now, for $T_t \in \Omega$, select i and j each uniformly at random from $[n]$. If $T_t[i, j]$ is a valid Young tableau (meaning the row and column constraints are satisfied), then let $T_{t+1} = T_t[i, j]$. Otherwise, let $T_{t+1} = T_t$. Let $\mathcal{P}$ be the corresponding transition matrix.

In the following discussion, it will be useful to prove the following lemma about swaps that can be performed on corners.

Lemma 5. *Let $T \in \Omega$. If $\alpha \in [n-1]$ is located at a corner of T, then $T[\alpha, \alpha+1]$ is a valid Young tableau.*

Proof. Consider the two spots of T local to both α and $\alpha+1$ which, in general, look like:

Because α is at a corner, the row and column conditions ensure that these two diagrams can only overlap at $\beta_1 = \gamma_2$ or $\beta_2 = \gamma_1$.

Now, if α and $\alpha+1$ were to be swapped, we need to check the resulting row and column conditions. Because $\alpha+1 < \gamma_3$, we immediately know $\alpha < \gamma_3$, and likewise for γ_4. Similarly, because $\alpha > \beta_1$, we immediately know $\alpha + 1 > \beta_1$, and likewise for β_2. Now, we have that $\gamma_1 < \alpha+1$. But because α only appears once in T, and $\gamma_1 \neq \alpha$, it must be that $\gamma_1 < \alpha$. The same can be said for γ_2. Thus, the row and column conditions local to α and $\alpha+1$ will still be satisfied after α and $\alpha+1$ are swapped. All the other row and column conditions for $T[\alpha, \alpha+1]$ are satisfied by virtue of their being satisfied in T. Thus, $T[\alpha, \alpha+1]$ is a valid Young tableau. □

Using this lemma, we can establish the following result.

Proposition 6. *$\mathcal{MC}_{swap}$ is irreducible.*

Proof. Fix $X_0, Y_0 \in \Omega$. We will show that there exists some t for which $\mathcal{P}^t(X_0, Y_0) > 0$; equivalently, we will show that we can make a series of swaps on X_0 to transition it into Y_0. We will do this by first making X_0 match Y_0 in the location of n, then in the location of $n-1$, etc., until they match for all integers from 1 to n.

Let r_n be the row that contains the number n in Y_0. Note that n must be located at the end of row r_n in Y_0 and that this location is a corner. We will make swaps on X_0 to move n to the end of row r_n. Let α be the number currently located at the end of row r_n on X_0. If $\alpha = n$, we are done, in which case call $X_0' = X_0$. Otherwise, by Lemma 5, we can make a series of swaps, say α with $\alpha+1$, then $\alpha+1$ with $\alpha+2$, etc., such that each intermediate tableau is a valid Young tableau, until we get the tableau

$X_0' = X_0[\alpha, \alpha+1][\alpha+1, \alpha+2]\cdots[n-1, n]$ which has n at the end of row r_n. Now X_0' and Y_0 match in the location of n.

Now, remove n from both X_0' and Y_0, giving us two smaller Young tablueaux of size $n-1$ with shape $\lambda' = (\lambda_r')_{r=1}^k$ with

$$\lambda_r' = \begin{cases} \lambda_r & \text{if } r \neq r_n \\ \lambda_r - 1 & \text{if } r = r_n \end{cases}. \tag{19}$$

Call these tableaux X_1 and Y_1. Now we can use the same process detailed above to transition X_1 to a Young tableau X_1' that matches Y_1 in the location of $n-1$. After removing $n-1$ from both X_1' and Y_1, we get two Young tableaux, X_2 and Y_2, of size $n-2$. We can repeat this process $n-2$ more times until the tableau derived from X_0 matches Y_0. Each swap that we perform has a positive probability of occurring in the Markov chain, so we have $\mathcal{P}^t(X_0, Y_0) > 0$, where t is the total number of swaps. □

It follows easily that our Markov chain is ergodic.

Proposition 7. *$\mathcal{MC}_{swap}$ is ergodic.*

Proof. First note that for any $X \in \Omega$, $\mathcal{P}(X, X) \geq \Pr(i = j) = 1/n$. Thus, the periodicity of X is 1, so the Markov chain is aperiodic. With Proposition 6, this implies that the Markov chain is ergodic. □

Now, by Theorem 1, we can conclude that this Markov chain has some stationary distribution π. Furthermore, just as we desire, the stationary distribution is the uniform distribution as shown here:

Proposition 8. *The stationary distribution π of $\mathcal{MC}_{swap}$ is uniform on Ω.*

Proof. Take any two distinct tableaux $X, Y \in \Omega$ such that Y differs from X by a single swap, i.e. there exist distinct $\alpha, \beta \in [n]$ such that $X[\alpha, \beta] = Y$. Then see that $\mathcal{P}(X, Y) = \Pr(\{i, j\} = \{\alpha, \beta\}) = 2/n^2$, and by symmetry $\mathcal{P}(Y, X) = 2/n^2$. For all other pairs X, Y that do not differ by a single swap (either $X = Y$, or $\mathcal{P}(X, Y) = 0$), we also have $\mathcal{P}(X, Y) = P(Y, X)$. Thus, $\mathcal{P}$ is symmetric.

Let π' be the uniform distribution on Ω. Then for any $x, y \in \Omega$, we get

$$\pi'(x)\mathcal{P}(x, y) = \pi'(y)\mathcal{P}(y, x). \tag{20}$$

By Theorem 2, we know that π' is a stationary distribution for our Markov chain. Since the chain is ergodic, we know by Theorem 1 that the stationary distribution is unique. Thus, $\pi = \pi'$ is the uniform distribution on Ω. □

So we now know that we can use $\mathcal{MC}_{\text{swap}}$ to sample uniformly from the set of all standard Young tableaux of some fixed shape. Now we would like to bound the mixing time of this Markov chain. There are two primary methods to bound the mixing time of a Markov chain, one which uses "couplings" and one which uses "canonical paths."

3.3 Coupling for $\mathcal{MC}_{\text{swap}}$

A *Markovian coupling* for a given Markov chain with space Ω and transition matrix P is a Markov chain (X_t, Y_t) on $\Omega \times \Omega$ with the following transition probabilities:

$$\begin{aligned} \Pr(X_{t+1} = a' \mid X_t = a, Y_t = b) &= P(a, a'), \\ \Pr(Y_{t+1} = b' \mid X_t = a, Y_t = b) &= P(b, b'). \end{aligned} \tag{21}$$

The coupling's transition matrix is often denoted $\hat{P}$. Essentially, it is a pair of two Markov chains run in parallel such that each individual chain looks like the Markov chain defined by P but which can be dependent on each other. Using such couplings can often be used to bound the mixing time of a Markov chain by using the following result often called the "Path Coupling theorem", first found in [2]:

Theorem 9. *For some Markov chain on Ω with transition matrix P, fix a coupling (X_t, Y_t). Let $G = (\Omega, E)$ be a graph and $d : E \to \mathbb{R}$ be a function that induces distances on $\Omega \times \Omega$. If there exists some $\alpha > 0$ such that for all $\{a, b\} \in E$,*

$$\mathbb{E}[d(X_{t+1}, Y_{t+1}) \mid X_t = a, Y_t = b] \leq (1 - \alpha) d(a, b), \tag{22}$$

then

$$\tau(\epsilon) \leq \frac{1}{\alpha} \log \left(\frac{d_{max}}{\epsilon} \right), \tag{23}$$

where $d_{max} = \max\{d(a, b) \mid (a, b) \in \Omega^2\}$.

[2] adds the remark that we can also get a bound on the mixing time if we assume the premise but with $\alpha = 0$, albeit a weaker one.

To investigate whether such a method could be used to bound the mixing time of $\mathcal{MC}_{\text{swap}}$, we construct the graph $G = (\Omega, E)$ where $E = \{\{a, b\} \mid \mathcal{P}(a, b) > 0\}$. The distance function we define on G is the natural one; the whole of d on Ω^2 is induced by letting $d(a, b) = 1$ for every $\{a, b\} \in E$ with $a \neq b$.

Unfortunately, however, we can show that with this definition of G and d, the Path Coupling theorem cannot be used to bound the mixing time of $\mathcal{MC}_{\text{swap}}$. We do this by defining a linear program that minimizes, over all possible couplings, the expected distance between two states after one transition of the Markov chain. It is defined more completely as follows.

Fix some $(a, b) \in E$. For every $(a', b') \in \Omega^2$, let $x_{a',b'} = \hat{\mathcal{P}}((a, b), (a', b'))$ be a variable to be determined by our linear program. Our transition matrix $\hat{\mathcal{P}}$ must define a coupling, so we must satisfy the constraints in eq. (21). These constraints are equivalent to

$$\begin{aligned} \sum_{b' \in \Omega} \hat{\mathcal{P}}((a, b), (a', b')) &= \mathcal{P}(a, a') \quad \text{for all } a' \in \Omega, \\ \sum_{a' \in \Omega} \hat{\mathcal{P}}((a, b), (a', b')) &= \mathcal{P}(b, b') \quad \text{for all } b' \in \Omega. \end{aligned} \tag{24}$$

Translating this into the variables in our linear program, we get the following constraints:

$$\begin{aligned} \sum_{b' \in \Omega} x_{a',b'} &= \mathcal{P}(a, a') \quad \text{for all } a' \in \Omega, \\ \sum_{a' \in \Omega} x_{a',b'} &= \mathcal{P}(b, b') \quad \text{for all } b' \in \Omega. \end{aligned} \tag{25}$$

Because all possible outcomes are represented by the probabilities $x_{a',b'}$, we may consider including the constraint

$$\sum_{a' \in \Omega} \sum_{b' \in \Omega} x_{a',b'} = 1, \tag{26}$$

but this is taken care of by the constraints in eq. (25) and the fact that $\mathcal{P}$ is a transition matrix, as

$$\sum_{a' \in \Omega} \sum_{b' \in \Omega} x_{a',b'} = \sum_{a' \in \Omega} \mathcal{P}(a, a') = 1. \tag{27}$$

The only other constraints we need are those that force our variables to represent probabilities:

$$0 \leq x_{a',b'} \leq 1 \quad \text{for all } (a', b') \in \Omega^2. \tag{28}$$

Note, however, that the constraints listed in eq. (25) automatically provide an upper bound on the variables, so we do not need to include the upper bound here; that is, we only need

$$x_{a',b'} \geq 0 \quad \text{for all } (a', b') \in \Omega^2. \tag{29}$$

Now, we want to know if there exists a coupling such that the expected value found in eq. (22) is strictly less than $d(a, b) = 1$. Thus, we wish to minimize the following objective function:

$$\mathbb{E}[d(X_t, Y_t) \mid X_t = a, Y_t = b] = \sum_{(a',b')\in\Omega} d(a', b')x_{a',b'}. \tag{30}$$

With this linear program now defined, we can use it to show that the Path Coupling theorem will not help us bound the mixing time of our Markov chain in general.

Proposition 10. *Using the naïve graph $G = (\Omega, E)$ and distance d defined above, the Path Coupling theorem cannot be used to bound the mixing time of $\mathcal{MC}_{swap}$.*

Proof. Fix the shape $\lambda = (3, 2, 1)$, and consider the pair of Young tableaux

$$a = \begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3&5\\\cline{1-2} 6\\\cline{1-1}\end{array} \quad \text{and} \quad b = \begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4&5\\\cline{1-2} 6\\\cline{1-1}\end{array}. \tag{31}$$

Since a and b differ by a swap, $\{a, b\} \in E$.

Notice because each $x_{a',b'} \geq 0$, the constraints defined in eq. (25) for which either $\mathcal{P}(a, a') = 0$ or which $\mathcal{P}(b, b') = 0$ force the corresponding variables $x_{a',b'} = 0$. Thus, the only nonzero variables are those $x_{a',b'}$ such that both $\mathcal{P}(a, a') > 0$ and $\mathcal{P}(b, b') > 0$.

By analyzing b, we can see that the only tableau it can transition to, besides a and b itself, is

$$c = \begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4&6\\\cline{1-2} 5\\\cline{1-1}\end{array}. \tag{32}$$

The only tableaux that a can transition to, besides b and a itself, are

$$d = \begin{array}{|c|c|c|}\hline 1&3&4\\\hline 2&5\\\cline{1-2} 6\\\cline{1-1}\end{array}, \; e = \begin{array}{|c|c|c|}\hline 1&2&5\\\hline 3&4\\\cline{1-2} 6\\\cline{1-1}\end{array}, \; f = \begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3&6\\\cline{1-2} 5\\\cline{1-1}\end{array}, \text{ and } g = \begin{array}{|c|c|c|}\hline 1&2&6\\\hline 3&5\\\cline{1-2} 4\\\cline{1-1}\end{array}. \tag{33}$$

With the rules that govern $\mathcal{MC}_{\text{swap}}$, we have $\mathcal{P}(b, a) = \mathcal{P}(b, c) = \mathcal{P}(a, b) = \mathcal{P}(a, d) = \mathcal{P}(a, e) = \mathcal{P}(a, f) = \mathcal{P}(a, g) = 1/18$, $\mathcal{P}(b, b) = 16/18$, and $\mathcal{P}(a, a) =$

13/18. Thus, the constraints from eq. (25) translate to the following:

$$x_{a,a} + x_{b,a} + x_{d,a} + x_{e,a} + x_{f,a} + x_{g,a} = \frac{1}{18}, \tag{34}$$

$$x_{a,b} + x_{b,b} + x_{d,b} + x_{e,b} + x_{f,b} + x_{g,b} = \frac{16}{18}, \tag{35}$$

$$x_{a,c} + x_{b,c} + x_{d,c} + x_{e,c} + x_{f,c} + x_{g,c} = \frac{1}{18}, \tag{36}$$

$$x_{a,a} + x_{a,b} + x_{a,c} = \frac{13}{18}, \tag{37}$$

$$x_{b,a} + x_{b,b} + x_{b,c} = \frac{1}{18}, \tag{38}$$

$$x_{d,a} + x_{d,b} + x_{d,c} = \frac{1}{18}, \tag{39}$$

$$x_{e,a} + x_{e,b} + x_{e,c} = \frac{1}{18}, \tag{40}$$

$$x_{f,a} + x_{f,b} + x_{f,c} = \frac{1}{18}, \tag{41}$$

$$x_{g,a} + x_{g,b} + x_{g,c} = \frac{1}{18}. \tag{42}$$

Additionally, we still have the constraints $x_{i,j} \geq 0$ for each variable in consideration. By analyzing the tableaux, we get

$$d(a,c) = d(d,b) = d(e,b) = d(f,b) = d(g,b) = 2, \tag{43}$$

$$d(d,c) = d(e,c) = d(g,c) = 3, \tag{44}$$

$d(i,i) = 0$, and $d(i,j) = 1$ for all other pairs (i,j) in consideration. This gives us the following objective function from eq. (30):

$$\begin{aligned}\text{minimize} \quad Z = {} & x_{a,b} + 2x_{a,c} + x_{b,a} + x_{b,c} + x_{d,a} + 2x_{d,b} + 3x_{d,c} + x_{e,a} + 2x_{e,b} \\ & + 3x_{e,c} + x_{f,a} + 2x_{f,b} + x_{f,c} + x_{g,a} + 2x_{g,b} + 3x_{g,c}.\end{aligned} \tag{45}$$

Now that we have the linear program defined, we want to show that its optimum value is strictly larger than $d(a,b) = 1$, as this implies that the Path Coupling theorem does not apply. We will do this by considering the dual of our linear program:

$$\begin{aligned}\text{maximize} \quad Z' = {} & \frac{1}{18}y_1 + \frac{16}{18}y_2 + \frac{1}{18}y_3 + \frac{13}{18}y_4 + \frac{1}{18}y_5 + \frac{1}{18}y_6 + \frac{1}{18}y_7 \\ & + \frac{1}{18}y_8 + \frac{1}{18}y_9\end{aligned} \tag{46}$$

$$
\begin{aligned}
\text{subject to} \quad & y_1 + y_4 \leq 0 \qquad y_1 + y_7 \leq 1 && (47)\\
& y_2 + y_4 \leq 1 \qquad y_2 + y_7 \leq 2 && (48)\\
& y_3 + y_4 \leq 2 \qquad y_3 + y_7 \leq 3 && (49)\\
& y_1 + y_5 \leq 1 \qquad y_1 + y_8 \leq 1 && (50)\\
& y_2 + y_5 \leq 0 \qquad y_2 + y_8 \leq 2 && (51)\\
& y_3 + y_5 \leq 1 \qquad y_3 + y_8 \leq 1 && (52)\\
& y_1 + y_6 \leq 1 \qquad y_1 + y_9 \leq 1 && (53)\\
& y_2 + y_6 \leq 2 \qquad y_2 + y_9 \leq 2 && (54)\\
& y_3 + y_6 \leq 3 \qquad y_3 + y_9 \leq 3 && (55)
\end{aligned}
$$

Now, consider the assignment $(y_i)_{i=1}^{9} = (1, 2, 1, -1, -2, 0, 0, 0, 0)$. This assignment satisfies the constraints above, and with it, the objective function evaluates to $Z' = 19/18$. It is a well-know result (for example, see [8]) that all feasible solutions of the dual of a linear program provide a bound (lower or upper, depending on whether the objective is to minimize or maximize, respectively) on the optimum value of the primal linear program. Thus, we have $Z \geq 19/18$. This implies that our linear program has an optimum value strictly greater than $1 = d(a, b)$, and hence, the Path Coupling theorem cannot provide a bound for $\mathcal{MC}_{\text{swap}}$, given our choice of graph and distance function. □

The above counterexample was found using the code in the Appendix (Section 6) below. From this result, we can conclude that using this graph G with the distance function d, both of which were selected to be the most natural for this Markov chain, we are unable to use the Path Coupling theorem to bound the mixing time of our Markov chain. It is still open whether this method could work with a more complicated graph and/or distance function. However, even if our Markov chain has a polynomial mixing time, this coupling method may not work. For example, Kumar and Ramesh [1] show that there exists a fast-mixing Markov chain for the problem of sampling perfect and near perfect matchings of a given graph but for which any coupling-based argument will only ever give an exponential bound on the mixing time.

3.4 Canonical paths for $\mathcal{MC}_{\text{swap}}$

The canonical path method for bounding the mixing time of a reversible Markov chain was developed by Jerrum and Sinclair in [10]. To use the method, for every $(x, y) \in \Omega^2$ we define a path in the graph $G = (\Omega, E)$

(where $E = \{(u, v) \mid P(u, v) > 0\}$) from x to y labeled $\gamma_{x,y}$. Then we let the congestion of an edge be

$$Congestion(u, v) = \frac{1}{\pi(u)P(u, v)} \sum_{\substack{x,y \\ (u,v)\in\gamma_{x,y}}} \pi(x)\pi(y)|\gamma_{x,y}|. \tag{56}$$

We then have the following result:

Theorem 11. *Let* $\rho = \max\{Congestion(u, v) \mid (u, v) \in \Omega^2\}$. *Then*

$$\tau(\epsilon) \leq 2\rho \left(2 \ln \left(\frac{1}{\epsilon} \right) + \ln \left(\frac{1}{\min \pi(x)} \right) \right). \tag{57}$$

Thus, if we can describe canonical paths such that we can find a polynomial bound on the maximum congestion of an edge, we can get a polynomial bound on the mixing time of our Markov chain.

For $\mathcal{MC}_{\text{swap}}$, we define canonical paths using the process established in the proof of Proposition 6. Fix two tableaux $u, v \in \Omega$. Locate the position p of n in v. Starting with $w = v$, use swaps to increment the number at p in w by 1 repeatedly until n is also located at p for w. Repeat this process for $n - 1$, $n - 2$, etc., until w is identical to v. As justified in the proof of Proposition 6, each of the intermediate tableaux are valid Young tableaux, so this process defines a canonical path $\gamma_{u,v}$ from vertex u to vertex v in our graph.

With these paths established, we need to bound the congestion of the edges of our graph. For a given pair of Young tableaux $u, v \in \Omega$, we have the following:

$$Congestion(u, v) = \frac{1}{\pi(u)\mathcal{P}(u, v)} \sum_{\substack{x,y \\ (u,v)\in\gamma_{x,y}}} \pi(x)\pi(y)|\gamma_{x,y}| \tag{58}$$

$$\leq \frac{1}{|\Omega|} \sum_{\substack{x,y \\ (u,v)\in\gamma_{x,y}}} |\gamma_{x,y}| \tag{59}$$

$$\leq \frac{n(n+1)}{2|\Omega|} \sum_{\substack{x,y \\ (u,v)\in\gamma_{x,y}}} 1. \tag{60}$$

Unfortunately, however, we do not currently have a bound on the size of the set $\{(x, y) \in \Omega^2 \mid (u, v) \in \gamma_{x,y}\}$ that yields a polynomial bound on the congestion of the edge (u, v). We leave this for future work.

4 Sampling Semistandard Young Tableaux

We now switch to the problem of sampling semistandard Young tableaux. Recall that the RSK correspondence maps contingency tables to pairs of semistandard Young tableaux of the same shape. Thus, this more generalized case is more interesting to us than the case of sampling standard tableaux. Furthermore, once we fix the row and column sums for our contingency tables, the weights of the corresponding Young tableaux are determined while their shapes are not. Thus, our overall goal is to sample pairs of semistandard Young tableaux with fixed weights and the same, but unfixed, shape.

4.1 Complexity of counting Young tableaux

Because of the strong connection between Kostka numbers (and by extension, Young tableaux) and other areas of mathematics, a lot of work has gone into understanding and calculating these coefficients (for example, see [18]). However, definitive results have remained elusive, possibly due to the complexity of the problem. In fact, in 2006, Hariharan Narayanan [17] showed that the problem of computing arbitrary Kostka numbers is $\#P$-complete, meaning that unless $P = NP$, there does not exist an algorithm that can compute these numbers in polynomial time. However, even if $P \neq NP$, the question of whether we can efficiently sample Young tableaux is still open since efficient exact counting implies efficient sampling, but efficient sampling only implies efficient approximate counting, as discussed in Section 2.2.

4.2 Generalizing $\mathcal{MC}_{\text{swap}}$ will not work

One natural Markov chain to consider is the one that uses swaps as in the case of sampling standard Young tableaux, generalizing $\mathcal{MC}_{\text{swap}}$ to the semistandard case. Such a chain would sample Young tableaux with both fixed shapes and fixed weights. Although this would not immediately serve our end goal of sampling pairs of tableaux with fixed weights but variable shapes, this process is still worth studying. For example, such a chain could be useful if we also found a way of sampling a random shape with the proper distribution. Then, we could sample a random shape and subsequently use this Markov chain to sample two Young tableaux with the selected shape and the desired weights.

Unfortunately, however, we can show that generalizing $\mathcal{MC}_{\text{swap}}$ to the semistandard case does not yield an irreducible Markov chain in all cases. For

example, consider the following Young tableaux:

$$\begin{array}{|c|c|}\hline 1&2\\\hline 3&3\\\hline 4&5\\\hline\end{array} \quad \text{and} \quad \begin{array}{|c|c|}\hline 1&3\\\hline 2&4\\\hline 3&5\\\hline\end{array}. \tag{61}$$

These two tableaux both have the same shape and weight, so we would need to find a way of getting from one to the other using swaps. Recall that the constraints of semistandard Young tableaux are that the rows need to be weakly increasing while the columns need to be strictly increasing. Therefore, there are no swaps we can perform on either of these tableaux that yield a valid Young tableau. Consequently, there is no way of getting from one tableau to the other using only swaps, and our Markov chain is not connected. Thus, we must consider a different type of Markov chain in the semistandard case.

4.3 A Markov chain for variable shapes

We now propose a Markov chain to sample from all Young tableaux of a given weight regardless of shape. Fix a weight $\mu = (\mu_j)_{j=1}^n$, and let $N = \sum_{j=1}^n \mu_j$. Let Ω be the set of all semistandard Young tableaux with weight μ; we wish to sample uniformly from Ω.

We propose the following Markov chain $\mathcal{MC}_{\text{shuffle}}$ to sample from Ω. Given a Young tableau X_t with weight μ, select an entry uniformly at random from the tableau. Let this number be a and let the row it comes from be r_0. Now, randomly select r_1 uniformly from $[a]$. Remove the selected entry a from the row r_0, and shift all of the entries to the right of a in row r_0 to the left to fill in the gap. Move all entries in row r_1 greater than a to the right, and insert a into the gap that is created. If no such entries exist, simply insert a at the end of row r_1. If the resulting tableau, X_t', is a valid Young tableau, let $X_{t+1} = X_t'$. Otherwise, let $X_{t+1} = X_t$. Let $\mathcal{P}$ be the corresponding transition matrix. An example of a transition of this Markov chain is illustrated in fig. 1.

We now will show that $\mathcal{MC}_{\text{shuffle}}$ has many of the desired properties for sampling from Ω.

Proposition 12. *$\mathcal{MC}_{shuffle}$ is irreducible.*

Proof. Let $Y \in \Omega$ be the Young tableau with weight μ and only one row, and let $X \in \Omega$ be an arbitrary tableau. We will show that we can reach Y starting from X and vice versa, using only transitions from $\mathcal{MC}_{\text{shuffle}}$. This will imply that $\mathcal{MC}_{\text{shuffle}}$ is irreducible.

We first show that Y is reachable from X. Label $X_0 = X$. Let X_{i+1} be the tableau acquired by moving the largest entry a_i in X_i not in the first row

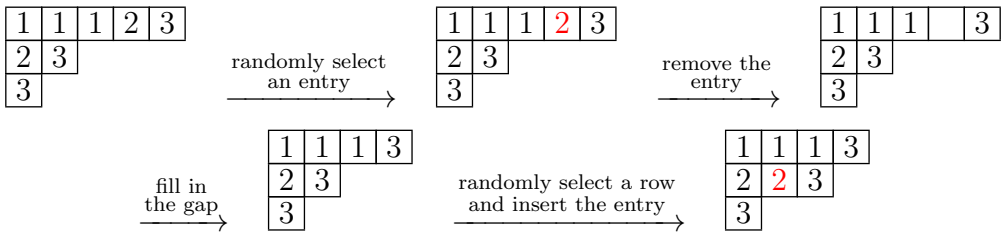


Figure 1: An example of a transition in $\mathcal{MC}_{\text{shuffle}}$. Because the resulting tableau is a valid Young tableau, the Markov chain will transition to it as its next state.

to the first row, in the manner defined by $\mathcal{MC}_{\text{shuffle}}$ (if multiple such entries exist, select the entry latest in lexicographic order). We will show that if X_i is a valid Young tableau, then so is X_{i+1}.

Note that a_i must be located on a corner of X_i, and therefore, it can be removed without violating any of the row or column constraints of a Young tableau. Now, after inserting a_i into the first row, we claim that no entry will be located beneath it. To see this, assume the opposite, and let c be the entry located directly beneath a_i and b be the entry previously located above b (so a_i has moved b and all of the following entries in the first row to the right). By our assumption that X_i is a valid Young tableau, we know that $b < c$. Furthermore, if we inserted a_i immediately before b, then $a_i < b$, and hence, $a_i < c$. But this contradicts the fact that a_i was selected as the largest entry of X_i not contained in the first row. Thus, after inserting a_i into the first row, no entries are located beneath it, and therefore no row or column constraints are violated. X_{i+1}, therefore, is a valid Young tableau.

Because there are finitely many entries in X, after finitely many steps, there will be some i' for which $X_{i'} = Y$. Therefore, we can reach Y from X using transitions allowed by $\mathcal{MC}_{\text{shuffle}}$.

Now we show that X is reachable from Y. We construct X from Y using transitions from $\mathcal{MC}_{\text{shuffle}}$ by building X up, entry-by-entry in lexicographic order, starting with the second row, and always removing from the first row. Because X is a valid Young tableau, each of the intermediate tableaux will necessarily satisfy all row constraints and all column constraints from the second row down. It is also easy to see that the column constraints between the first and second rows will also always be satisfied. Let b be an entry in the second row of X, and let a be the entry immediately above it in the first row (so $a < b$). Because Y starts with all entries in weakly increasing order in the first

row, once b is moved to the second row, the entries located above b throughout the process will never be any larger than a and, hence, will satisfy the column constraints of a Young tableau. Since all row and column constraints are satisfied, each intermediate tableau will be a valid Young tableau. In this way, we can reach X from Y using transitions allowed by $\mathcal{MC}_{\text{shuffle}}$. □

Proposition 13. *$\mathcal{MC}_{shuffle}$ is ergodic.*

Proof. Notice that any tableau can transition to itself by inserting an entry in the row that it started in. Thus, for any $X \in \Omega$, $\mathcal{P}(X, X) > 0$, which means the periodicity of every state in Ω is 1. So, $\mathcal{MC}_{\text{shuffle}}$ is aperiodic. Combining this with Proposition 12 gives us that $\mathcal{MC}_{\text{shuffle}}$ is ergodic. □

So, invoking Theorem 1, let π be the stationary distribution of $\mathcal{MC}_{\text{shuffle}}$. We now will show that π is uniform on Ω.

Proposition 14. *The stationary distribution π of $\mathcal{MC}_{shuffle}$ is uniform on Ω.*

Proof. Notice that every transition described by $\mathcal{MC}_{\text{shuffle}}$ is entirely reversible in the sense that the probability of selecting entry a in row r_0 of X and inserting it into row r_1 to yield a tableau Y is equal to the probability of selecting entry a in row r_1 of Y and inserting it into row r_0 to yield tableau X. Thus, for any two $X, Y \in \Omega$, $\mathcal{P}(X, Y) = \mathcal{P}(Y, X)$; i.e. $\mathcal{P}$ is symmetric. Thus, just as in the proof of Proposition 8, the uniform distribution satisfies the detailed balance condition, so we can conclude π is uniform on Ω. □

So, we know that $\mathcal{MC}_{\text{shuffle}}$ can be used to sample uniformly from Ω. At this time, we do not have any results related to the mixing time of this Markov chain, which we leave to future work.

5 Conclusion and Future Work

The study of sampling Young tableaux (and the related problem of computing Kostka numbers) is a rich problem with many ties to other areas of mathematics. Here, we propose potential methods of efficiently sampling Young tableaux from various distributions. In the future, we hope to continue and expand on the study of the proposed Markov chains in the hopes that it can build toward the eventual efficient sampling of contingency tables.

One major direction to continue in is to determine if other graphs and/or distance functions could be used for path coupling to bound the mixing time of $\mathcal{MC}_{\text{swap}}$, as discussed in Section 3.3. One distance function that we have

considered but not fully explored is that which calculates the total pairwise difference between two tableaux' entries. To define this more rigorously, let X and Y be two standard Young tableaux with the same shape. Let the entries of X in lexicographic order be $(a_1, a_2, \ldots, a_n)$ and those of Y be $(b_1, b_2, \ldots, b_n)$. Then let the distance function d be defined by

$$d(X, Y) = \sum_{i=1}^{n} |a_i - b_i|. \tag{62}$$

Such a distance function could be paired with, for example, the graph G with $E = \{\{a, b\} \mid \mathcal{P}(a, b) > 0\}$ and potentially allow for the use of the Path Coupling theorem.

Another direction would be to continue exploring methods of bounding the congestion on the edges of the canonical paths proposed in Section 3.4. This would mainly come down finding a sufficient bound on the size of the set $\{(x, y) \in \Omega^2 \mid (u, v) \in \gamma_{x,y}\}$ or showing that no such bound exists. Alternatively, different canonical paths between tableaux could be considered and explored.

A possibly more interesting (but also more difficult) direction would be to study couplings and/or canonical paths for $\mathcal{MC}_{\text{shuffle}}$ for sampling semistandard Young tableaux. Finding polynomial bounds on the mixing time of this chain would be a big step toward our ultimate goal of sampling contingency tables, although such bounds are not guaranteed to exist.

Finally, a more general future direction in the study of sampling contingency tables is to explore and identify different connections between contingency tables and other areas of mathematics. Clearly, we can continue to study Markov chains on sets of contingency tables, including the one proposed in [3], but it may be easier to approach the problem by way of other areas of mathematics as we attempt here.

References

[1] VS Anil Kumar and Hariharan Ramesh. Coupling vs. conductance for the jerrum–sinclair chain. *Random Structures & Algorithms*, 18(1):1–17, 2001.

[2] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 223–231. IEEE, 1997.

[3] Persi Diaconis and Anil Gangolli. Rectangular arrays with fixed margins. In David Aldous, Persi Diaconis, Joel Spencer, and J. Michael Steele, editors, *Discrete Probability and Algorithms*, pages 15–41, New York, NY, 1995. Springer New York.

[4] J Sutherland Frame, G de B Robinson, and Robert M Thrall. The hook graphs of the symmetric group. *Canadian Journal of Mathematics*, 6:316–324, 1954.

[5] Georg Ferdinand Frobenius, Ferdinand Georg Frobenius, Ferdinand Georg Frobenius, Ferdinand Georg Frobenius, and Germany Mathematician. *Über die Charaktere der symmetrischen Gruppe*. Königliche Akademie der Wissenschaften, 1900.

[6] William Fulton. *Young tableaux: with applications to representation theory and geometry*. Number 35. Cambridge University Press, 1997.

[7] Curtis Greene, Albert Nijenhuis, and Herbert S Wilf. A probabilistic proof of a formula for the number of young tableaux of a given shape. In *Young Tableaux in Combinatorics, Invariant Theory, and Algebra*, pages 17–22. Elsevier, 1982.

[8] Frederick S Hillier and Gerald J Lieberman. *Introduction to operations research*. McGraw-Hill Science, Engineering & Mathematics, 1995.

[9] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Springer Science & Business Media, 2003.

[10] Mark Jerrum and Alistair Sinclair. Conductance and the rapid mixing property for markov chains: the approximation of permanent resolved. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 235–244, 1988.

[11] Kurt Johansson. Discrete orthogonal polynomial ensembles and the plancherel measure. *Annals of Mathematics*, pages 259–296, 2001.

[12] Koko Kayibi, Shariefuddin Pirzada, and Tariq Chishti. Sampling contingency tables. *AKCE International Journal of Graphs and Combinatorics*, 15, October 2017.

[13] Donald E. Knuth. Permutations, matrices, and generalized Young tableaux. *Pacific Journal of Mathematics*, 34(3):709–727, 1970.

[14] Carl Kostka. Über den zusammenhang zwischen einigen formen von symmetrischen functionen. In *Journal für die reine und angewandte Mathematik Band 93*, pages 89–123. De Gruyter, 2020.

[15] Ian Grant Macdonald. *Symmetric functions and Hall polynomials*. Oxford university press, 1998.

[16] Ben Morris. Improved bounds for sampling contingency tables. *Random Structures & Algorithms*, 21:135 – 146, September 2002.

[17] Hariharan Narayanan. On the complexity of computing kostka numbers and littlewood-richardson coefficients. *Journal of Algebraic Combinatorics*, 24(3):347–354, 2006.

[18] Etienne Rassart. *Geometric approaches to computing Kostka numbers and Littlewood-Richardson coefficients*. PhD thesis, Massachusetts Institute of Technology, 2004.

[19] Dan Romik. *The surprising mathematics of longest increasing subsequences*, volume 4. Cambridge University Press, 2015.

[20] Richard P Stanley. Enumerative combinatorics. vol. 2, volume 62 of. *Cambridge Studies in Advanced Mathematics*, 1999.

[21] Alfred Young. On quantitative substitutional analysis. *Proceedings of the London Mathematical Society*, pages 361–397, 1902.

6 Appendix

The following Mathematica code was used to search for a counterexample in Section 3.3:

```
#!/usr/bin/env wolframscript
(* ::Package:: *)

Block[{$ContextPath}, Needs["Combinatorica`"]];
Swapped[e_] := Which[
                    Count[yts[[e[[1]]]] - yts[[e[[2]]]],
                                Except[0],{2}]==2,
                    True, True, False]
shape = {3,2,1};
yts = Combinatorica`Tableaux[shape];
```

```
n = Total[shape];
nv = Length[yts];
vertices = Range[nv];
edges = Select[Subsets[vertices, {2}], Swapped];
G = Graph[vertices,edges];
deg[v_] := VertexDegree[G,v]
d[v1_,v2_] := GraphDistance[G,v1,v2]
p[v1_,v2_] := Which[v1==v2, 1-2deg[v1]/n^2,
                    MemberQ[edges,MinMax[{v1,v2}]],
                      2/n^2, True, 0]
mu[v_] := Table[p[v,w],{w,nv}]
x[v1_,v2_] := Symbol[StringJoin["x$",ToString@v1,"$",
                      ToString@v2]]
varlist = Flatten[Table[Table[x[a,b],{a,nv}],{b,nv}]];
objective = Sum[d[a,b]x[a,b],{a,1,nv},{b,1,nv}];
rowconst[i_] := Table[Sum[x[a,b],{b,nv}]==mu[i][[a]], {a,nv}];
colconst[j_] := Table[Sum[x[a,b],{a,nv}]==mu[j][[b]], {b,nv}];
posconst = Table[varlist[[v]]>=0, {v,Length[varlist]}];
const[i_,j_] := Join[{objective},rowconst[i],colconst[j],
                     posconst];
opt[i_,j_] := d[i,j]-Minimize[const[i,j],varlist][[1]];

(*A table of all the opt() values for the pairs of vertices on
the edges of the graph. If any of these are negative, then
using the Path Coupling Theorem cannot work since the theorem
requires the expected distance to decrease (i.e. all values
need to be non-negative)*)
Table[opt[e[[1]],e[[2]]],{e,edges}]
```