ON THE LIMITING SHAPE OF RANDOM YOUNG TABLEAUX FOR MARKOVIAN WORDS

A Thesis
Presented to
The Academic Faculty

by

Trevis J. Litherland

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology
December 2008

ON THE LIMITING SHAPE OF RANDOM YOUNG TABLEAUX FOR MARKOVIAN WORDS

Approved by:

Professor Christian Houdré, Advisor
Department of Mathematics
Georgia Institute of Technology

Professor Vladimir Koltchinskii
Department of Mathematics
Georgia Institute of Technology

Professor Ionel Popescu
Department of Mathematics
Georgia Institute of Technology

Professor Robert Foley
Department of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Yuri Bakhtin
Department of Mathematics
Georgia Institute of Technology

Professor Mikhail Lifshitz
Mathematical Mechanics Faculty
St. Petersburg State University

Professor Heinrich Matzinger
Department of Mathematics
Georgia Institute of Technology

Date Approved: November 12, 2008

To my loving wife Sunchin,

Without whose support

This thesis would have been,

Almost surely,

Nonexistent.

ACKNOWLEDGEMENTS

"Good definitions make for easy theorems" is by now a piece of mathematical folklore, one which has the additional property of actually being true. A corollary might well be that good problems make for, if not easy, then at least interesting thesis research. And so I would like to first of all thank my advisor, Dr. Christian Houdré, for introducing me to the subject of longest increasing subsequences. Life is not all fun and games, however, and so I would also like to express my gratitude to the committee members, especially my advisor, for their perusal of this thesis and the helpful comments such efforts produced.

As one whose calculator is older than most students I teach, I must also thank the School of Mathematics for allowing me to return to the program with minimal academic disruption.

"Peer review" has a technical meaning, but another good definition certainly must include everyday interactions with fellow graduate students. In particular, I would like to thank Alex Grigo for asking me hundreds of probability questions for which I could often confirm the existence of answers, to my fellow advisee Xu Hua for his ongoing input regarding random matrices, and to Mitch Keller, for serving the public (and my personal!) good in creating the standard formatting paradigm for Georgia Tech theses. I would also like to thank the organizers of the Special Program on High-Dimensional Inference and Random Matrices at SAMSI in 2006. Their hospitality and support, through the grant DMS-0112069, greatly facilitated the research connected with this thesis. Finally, I would like to thank Dr. Mark Meckes, whose talk in Marseille in May of this year provided the final insight I needed to completely answer Kuperberg's Conjecture.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The limiting law of the length of the longest increasing subsequence, LI_n, for sequences (words) of length n arising from iid letters drawn from finite, ordered alphabets is studied using a straightforward Brownian functional approach. Building on the insights gained in both the uniform and non-uniform iid cases, this approach is then applied to iid countable alphabets. Some partial results associated with the extension to independent, growing alphabets are also given. Returning again to the finite setting, and keeping with the same Brownian formalism, a generalization is then made to words arising from irreducible, aperiodic, time-homogeneous Markov chains on a finite, ordered alphabet. At the same time, the probabilistic object, LI_n, is simultaneously generalized to the shape of the associated Young tableau given by the well-known RSK-correspondence. Our results on this limiting shape describe, in detail, precisely when the limiting shape of the Young tableau is (up to scaling) that of the iid case, thereby answering a conjecture of Kuperberg. These results are based heavily on an analysis of the covariance structure of an m-dimensional Brownian motion and the precise form of the Brownian functionals. Finally, in both the iid and more general Markovian cases, connections to the limiting laws of the spectrum of certain random matrices associated with the Gaussian Unitary Ensemble (GUE) are explored.

CHAPTER I

INTRODUCTION

A substantial portion of probability theory is concerned with the properties of sequences of random objects. Indeed, for sequences of real-valued random variables, the most fundamental questions include Laws of Large Numbers, Central Limit Theorems, Large Deviation Principles, and Invariance Principles (Functional Central Limit Theorems), all of which describe certain asymptotic properties of the sequence.

In this thesis we will be concerned with certain asymptotic properties of longest increasing subsequences, which we define as follows. Let $(X_k)_{1\le k\le n}$ be a sequence taken from an ordered alphabet $\mathcal{A}$ (usually finite, but possibly even uncountable). A *strictly increasing subsequence* of $(X_k)_{1\le k\le n}$ is a subsequence (X_{k_j}) such that $X_{k_j} < X_{k_{j+1}}$, for each j. Similarly, a *weakly increasing subsequence* of $(X_n)_{1\le k\le n}$ is a subsequence (X_{k_j}) such that $X_{k_j} \le X_{k_{j+1}}$, for each j. We will be primarily concerned with the latter type of increasing subsequence, and will refer to it as simply an *increasing subsequence*. A *longest increasing subsequence* of $(X_k)_{1\le k\le n}$, is then defined to be an increasing subsequence of maximal length, a length which we designate by LI_n.

Motivating our investigation of LI_n in various probabilistic contexts is the classical problem of describing the length of the longest (necessarily strictly) increasing subsequence of a *random permutation* of the first n positive integers. The study of the asymptotic behavior of this quantity, $L\sigma_n$, has enjoyed a rich history as "Ulam's Problem". The determination of its first-order asymptotics was accomplished by the work of Logan and Shepp [34], and Vershik and Kerov [45], who showed that $L\sigma_n/\sqrt{n} \to 2$ a.s. and in L^1. Newer methods making use of interacting particle processes and "hydrodynamical arguments" have brought new insights. In particular, Aldous and

Diaconis [1] and Seppäläinen [41] use such methods to show that $L\sigma_n/\sqrt{n} \to 2$ in expectation and in probability. Groeneboom [24] proves such convergence results using only the convergence of random signed measures, while Cator and Groeneboom [11] prove that $\mathbb{E}L\sigma_n/\sqrt{n} \to 2$ in a way that avoids both ergodic decomposition arguments and the subadditive ergodic theorem. Making further connections to other fields, Aldous and Diaconis [2] also connect these particle process concepts to the card game solitaire, while Seppäläinen [42] employs these particle processes to a verify an open asymptotics problem in Queuing Theory.

The far more challenging problem of finding the limiting behavior of $L\sigma_n$, once suitably centered and normalized, was solved by Baik, Deift, and Johansson, in their landmark paper [5]. In particular, they showed that $(L\sigma_n - 2\sqrt{n})/n^{1/6}$ converges in distribution to a non-trivial limiting distribution known as the Tracy-Widom distribution. Even more remarkable than the unusual scaling factor in this result is the fact that the Tracy-Widom distribution first arose in the study of the asymptotics of the largest eigenvalues of certain random matrices. Cator and Groeneboom [12] use particle processes to directly obtain the cube-root asymptotics of the variance of $L\sigma_n$. Further non-asymptotic results for $L\sigma_n$ are found in [25].

In this thesis, we will be concerned primarily with the asymptotics of LI_n for weakly increasing subsequences when the alphabet is finite or countably infinite. In the case that $(X_n)_{n\geq 1}$ is a sequence (often called a *word* in this context) of iid random variables taken from a finite ordered alphabet of size m, Tracy and Widom [44], as well as Johannson [31], have shown, in the uniform case, that the limiting distribution is that of the largest eigenvalue of an $m \times m$ matrix of the Gaussian Unitary Ensemble (GUE), subject to a zero-trace condition. Its, Tracy, and Widom [28, 29] have further examined this problem in the non-uniform iid case, relating the limiting distribution to certain direct sums of GUE matrices. (For a general overview of the subject of random matrices, refer to the standard text of Mehta [35].)

In this iid setting, we will investigate the limiting distribution of LI_n using a Brownian functional approach, which we will extend to the countably-infinite iid case as well. In the context of random growth processes, Gravner, Tracy, and Widom [22] have already obtained a Brownian functional of the form we derive. This functional appeared first in the work of Glynn and Whitt [20], in Queuing Theory, and its relation to the eigenvalues of the GUE has also been studied by Baryshnikov [6]. It is, moreover, remarked in [22] that the longest increasing subsequence problem could also be studied using a Brownian functional formulation.

To generalize beyond the iid setting, we then consider sequences generated by a time-homogeneous, irreducible, aperiodic Markov chain on a finite alphabet of size m. Moreover, we generalize the object of our study, LI_n, to that of the shape of the *Young tableau* generated by $(X_k)_{1\leq k\leq n}$ via the Robinson-Schensted-Knuth (RSK) correspondence. The shape of the Young tableau, which in this context consists of n left-aligned boxes arranged in at most m rows such that each row is no greater in length than the row above it, indeed generalizes LI_n: the length of the top row is simply LI_n. We confine our attention to irreducible, aperiodic Markov chains so as to ensure that the stationary distribution is unique.

In the particular case that the Markov chain generates a uniform iid sequence, Tracy and Widom [44] conjectured that the Young tableau has a limiting shape given by the *joint* distribution of the eigenvalues of a traceless $m \times m$ element of the GUE. Johansson [31] proved this conjecture using orthogonal polynomial methods. Further, Okounkov [38], and Borodin, Okounkov, and Olshankii [8], as well as Johansson [31], also answered a conjecture of Baik, Deift, and Johansson [4, 3] regarding the limiting shape of the Young tableau associated with a random permutation of the first n positive integers. In particular, as n grows without bound, the lengths $R_n^1, R_n^2, \ldots, R_n^k$ of the first k rows of the Young tableau, appropriately centered and scaled, have, asymptotically, the same limiting law as the k largest eigenvalues of an $n \times n$ element of the GUE, a result first proved, for $k = 2$, in [4, 3].

The non-uniform iid case was also addressed to some degree in Its, Tracy, and Widom [28, 29], who focused primarily on LI_n. Here the obvious conjecture is that the limiting shape has rows whose suitably centered and normalized lengths have a joint distribution which is that of the whole spectrum of the direct sum of certain GUE matrices, a result that was shown in the thesis of Xu [46].

However, the primary purpose of the Markovian framework is to move beyond the iid setting. Inspired by questions in statistical physics, Kuperberg [32] conjectured that if the sequence is generated by a more specific type of Markov chain, namely, an irreducible, aperiodic, *cyclic* one, then the limiting distribution of the shape is still that of the joint distribution of the eigenvalues of a traceless $m \times m$ element of the GUE. The cyclic criterion, *i.e.*, the Markov transition matrix P has entries satisfying $p_{i,j} = p_{i+1,j+1}$, for $1 \leq i, j \leq m$ (where $m + 1 = 1$), implies, but is not equivalent to, P being doubly stochastic, *i.e.*, having a uniform stationary distribution.

For $m = 2$, this was shown to be true by Chistyakov and Götze [13]. For $m = 3$, simulations by Kuperberg [32] indicated that it was true as well, and we show that, for $m = 3$, his conjecture is indeed true. However, for $m \geq 4$, this is no longer the case, as was also suggested by further simulations by Chistyakov and Götze [13]. Indeed, some, but not all, cyclic Markov chains lead to the same limiting law as in the iid uniform case already obtained by Johansson [31]. We obtain a precise description of the class of cyclic transition matrices generating the iid limiting shape.

Recall again that LI_n is the length of longest row of the associated Young tableau, and that an iid sequence may be viewed as a special case of a Markovian sequence. In this more specialized setting, we begin, in Section 2.1 of Chapter II, by writing LI_n as a simple algebraic expression. Using this simple characterization, we then investigate the m-letter iid case. In Section 2.2, we obtain the the limiting distribution of LI_n (properly centered and normalized) when the letters are chosen uniformly. Our result is expressed as a functional of an $(m-1)$-dimensional Brownian motion with correlated coordinates. Using certain natural symmetries, this limiting distribution is further expressed as various functionals of a (standard) Brownian motion. We then extend this development to the non-uniform iid case. In Section 2.3, connections with the Brownian functional originating with the work of Glynn and Whitt in Queuing Theory are investigated. This allows us to investigate the asymptotics of the limiting law of LI_n as the alphabet size m grows.

Next, in Chapter III, we extend our results to the iid case for countably infinite alphabets by reducing the problem to an effectively finite-alphabet one.

We then discuss, briefly, in Chapter IV, a time-*inhomogeneous* setting, wherein the sequence is chosen uniformly from independent, but growing, alphabets. As the results in this direction are partial, we prove only a first-order result which nonetheless bridges, in some sense, the linear asymptotics of LI_n in the iid finite-alphabet case and the $\sqrt{n}$ asymptotics of $L\sigma_n$.

Chapter V begins our study of the general Markovian framework for Young tableaux. In Section 5.1, we first use our combinatorial expression for LI_n developed in Section 2.1, to rederive the two-letter Markov case first studied by Chistyakov and Götze [13]. Then, in order to extend these results to alphabets of size $m \geq 3$, we introduce, in Section 5.2, a slight modification of our original combinatorial development, and so obtain a functional of combinatorial quantities which describes the shape of the entire Young tableau, along with a concise expression for the associated asymptotic covariance structure. Next, in Section 5.3, we apply Markovian Invariance Principles to express the limiting shape of the Young tableau as a Brownian functional for all irreducible, aperiodic, homogeneous Markov chains (without the cyclic or even the doubly-stochastic constraint.) Using this functional we are then able to answer Kuperberg's conjecture. In Section 5.4, we investigate, in further detail, various symmetries exhibited by the Brownian functional. In particular, we clarify the asymptotic covariance structure in the cyclic case, and obtain, for m arbitrary, a precise description of the class of cyclic Markov chains having the same limiting law as in the uniform iid case. In Section 5.5, we further explore connections between the various Brownian functionals obtained as limiting laws and eigenvalues of random matrices.

We conclude, in Chapter VI, with a brief discussion of natural extensions and complements (such as Queuing Theory) to some of the ideas and results presented in the thesis, and indicate promising directions for further research.

CHAPTER II

FINITE IID ALPHABETS

2.1 Combinatorics

Let $(X_n)_{n\geq 1}$ consist of a sequence of values taken from an m-letter ordered alphabet, $\mathcal{A} = \{\alpha_1 < \alpha_2 < \cdots < \alpha_m\}$. Let a_k^r be the number of occurrences of α_r among $X_1, X_2, \ldots, X_k$, $1 \leq k \leq n$. Each increasing subsequence of $(X_n)_{n\geq 1}$ consists simply of runs of identical values, with the values of each successive run forming an increasing subsequence of α_r. Moreover, the number of occurrences of α_r among $X_{k+1}, \ldots, X_\ell$, where $1 \leq k < \ell \leq n$, is simply $a_\ell^r - a_k^r$. The length of the longest increasing subsequence of $(X_n)_{n\geq 1}$ is then given by

$$LI_n = \max_{\substack{0\leq k_1 \leq \cdots \\ \leq k_{m-1} \leq n}} [(a_{k_1}^1 - a_0^1) + (a_{k_2}^2 - a_{k_1}^2) + \cdots + (a_n^m - a_{k_{m-1}}^m)], \tag{2.1.1}$$

i.e.,

$$LI_n = \max_{\substack{0\leq k_1 \leq \cdots \\ \leq k_{m-1} \leq n}} [(a_{k_1}^1 - a_{k_1}^2) + (a_{k_2}^2 - a_{k_2}^3) + \cdots + (a_{k_{m-1}}^{m-1} - a_{k_{m-1}}^m) + a_n^m], \tag{2.1.2}$$

where $a_0^r = 0$. For $i = 1, \ldots, n$ and $r = 1, \ldots, m-1$, let

$$Z_i^r = \begin{cases} 1, & \text{if } X_i = \alpha_r, \\ -1, & \text{if } X_i = \alpha_{r+1}, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1.3}$$

and let $S_k^r = \sum_{i=1}^k Z_i^r$, $k = 1, \ldots, n$, and also $S_0^r = 0$. Then clearly $S_k^r = a_k^r - a_k^{r+1}$. Hence,

$$LI_n = \max_{\substack{0 \le k_1 \le \cdots \\ \le k_{m-1} \le n}} \{S^1_{k_1} + S^2_{k_2} + \cdots + S^{m-1}_{k_{m-1}} + a^m_n\}. \tag{2.1.4}$$

Since $a^1_k, \dots, a^m_k$ must evidently sum to k, we have

$$\begin{aligned} n &= \sum_{r=1}^{m} a^r_n \\ &= \sum_{r=1}^{m-1} \left(a^m_n + \sum_{j=r}^{m-1} S^j_n \right) + a^m_n \\ &= \sum_{r=1}^{m-1} r S^r_n + m a^m_n. \end{aligned}$$

Solving for a^m_n gives us

$$a^m_n = \frac{n}{m} - \frac{1}{m} \sum_{r=1}^{m-1} r S^r_n.$$

Substituting into (2.1.4), we finally obtain

$$LI_n = \frac{n}{m} - \frac{1}{m} \sum_{r=1}^{m-1} r S^r_n + \max_{\substack{0 \le k_1 \le \cdots \\ \le k_{m-1} \le n}} \{S^1_{k_1} + S^2_{k_2} + \cdots + S^{m-1}_{k_{m-1}}\}. \tag{2.1.5}$$

The expression (2.1.5) is of a *purely combinatorial nature or, in more probabilistic terms, is of a pathwise nature*. We now analyze (2.1.5) in light of the probabilistic nature of the sequence $X_1, X_2, \dots, X_n$.

2.2 Probabilistic Development

Throughout the sequel, Brownian functionals will play a central rôle. By a *Brownian motion* we shall mean an a.s. continuous, centered Gaussian process $B(t)$, $0 \le t \le 1$, with $B(0) = 0$, having stationary, independent increments. By a *standard Brownian motion* we shall mean that $\operatorname{Var} B(t) = t$, $0 \le t \le 1$, *i.e.*, we endow $C[0,1]$ with the Wiener measure. A *standard m-dimensional Brownian motion* will be defined to be a vector-valued process consisting of m independent standard Brownian motions. More generally, an *m-dimensional Brownian motion* shall refer to a linear transformation of a standard m-dimensional Brownian motion. Throughout this thesis, we assume that our underlying probability space is rich enough so that all the Brownian motions and sequences we study can be defined on it.

We consider first the case in which $(X_n)_{n\geq 1}$ are iid, with each letter drawn uniformly from $\mathcal{A} = \{\alpha_1, \dots, \alpha_m\}$. Then, for each fixed letter r, the sequence $(Z_n^r)_{n\geq 1}$ is also formed of iid random variables with $\mathbb{P}(Z_1^r = 1) = \mathbb{P}(Z_1^r = -1) = 1/m$, and $\mathbb{P}(Z_1^r = 0) = 1 - 2/m$.

Thus $\mathbb{E}Z_1^r = 0$, and $\mathbb{E}(Z_1^r)^2 = 2/m$, and so, $\operatorname{Var} S_n^r = 2n/m$, for $r = 1, 2, \dots, m-1$. Defining $\hat{B}_n^r(t) = \frac{1}{\sqrt{2n/m}} S_{[nt]}^r + \frac{1}{\sqrt{2n/m}}(nt - [nt])Z_{[nt]+1}^r$, for $0 \leq t \leq 1$, and noting that the local maxima of $\hat{B}_n^i(t)$ occur at $t = k/n$, $k = 0, \dots, n$, we have from (2.1.5) that

$$\frac{LI_n - n/m}{\sqrt{2n/m}} = -\frac{1}{m}\sum_{i=1}^{m-1} i\hat{B}_n^i(1) + \max_{\substack{0\leq t_1 \leq \cdots \\ \leq t_{m-1} \leq 1}} [\hat{B}_n^1(t_1) + \cdots + \hat{B}_n^{m-1}(t_{m-1})]. \tag{2.2.1}$$

We can now invoke Donsker's Theorem since the measures $\mathbb{P}_n$ generated by $(\hat{B}_n^1(t), \dots, \hat{B}_n^{m-1}(t))$ satisfy $\mathbb{P}_n(A) \to \mathbb{P}_\infty(A)$, for all Borel subsets A of the space of continuous functions $C([0,1]^{m-1})$ for which $\mathbb{P}_\infty(\partial A) = 0$, where $\mathbb{P}_\infty$ is the $(m-1)$-dimensional Wiener measure. Thus, by Donsker's Theorem and the Continuous Mapping Theorem we have that $(\hat{B}_n^1(t), \dots, \hat{B}_n^{m-1}(t)) \Rightarrow (\tilde{B}^1(t), \dots, \tilde{B}^{m-1}(t))$, where the Brownian motion on the right has a covariance structure which we now describe. First, $\operatorname{Cov}(Z_1^r, Z_1^s) = \mathbb{E}Z_1^r Z_1^s = 0$, for $|r - s| \geq 2$, and $\operatorname{Cov}(Z_1^r, Z_1^{r+1}) = \mathbb{E}Z_1^r Z_1^{r+1} = -1/m$, for $r = 1, 2, \dots, m-1$. Then, as already noted, for each fixed r, $Z_1^r, Z_2^r, \dots Z_n^r, \dots$ are iid, and for fixed k, $Z_k^1, Z_k^2, \dots, Z_k^{m-1}$ are dependent but identically distributed random variables. Moreover, it is equally clear that for any r and s, $1 \leq r < s \leq m-1$, the sequences $(Z_k^r)_{k\geq 1}$ and $(Z_\ell^s)_{\ell\geq 1}$ are also identical distributions of the Z_k^r and that Z_k^r and Z_ℓ^s are independent for $k \neq \ell$. Thus, $\operatorname{Cov}(S_n^r, S_n^s) = n\operatorname{Cov}(Z_1^r, Z_1^s)$. This result, together with our $2n/m$ normalization factor gives the following covariance matrix for $(\tilde{B}^1(t), \ldots, \tilde{B}^{m-1}(t))$:

$$t \begin{pmatrix} 1 & -1/2 & & & \bigcirc \\ -1/2 & 1 & -1/2 & & \\ & \ddots & \ddots & \ddots & \\ & & -1/2 & 1 & -1/2 \\ \bigcirc & & & -1/2 & 1 \end{pmatrix}. \tag{2.2.2}$$

We remark here that the functional in (2.2.1) is a bounded continuous functional on $C(0,1)^{m-1}$. (This fact will be used throughout this thesis.) Hence, by a final application of the Continuous Mapping Theorem,

$$\frac{LI_n - n/m}{\sqrt{2n/m}} \Rightarrow -\frac{1}{m}\sum_{i=1}^{m-1} i\tilde{B}^i(1) + \max_{\substack{0\le t_1 \le \cdots \\ \le t_{m-1} \le 1}} \sum_{i=1}^{m-1} \tilde{B}^i(t_i). \tag{2.2.3}$$

We have thus obtained the limiting distribution of LI_n as a Brownian functional. Tracy and Widom [44] already obtained the limiting distribution of LI_n in terms of the distribution of the largest eigenvalue of the Gaussian Unitary Ensemble (GUE) of $m \times m$ Hermitian matrices having trace zero. Johansson [31] generalized this work to encompass all m eigenvalues. Gravner, Tracy, and Widom [22] in their study of random growth processes make a connection between the distribution of the largest eigenvalue in the $m \times m$ GUE and a Brownian functional essentially equivalent, up to a normal random variable, to the right hand side of (2.2.3). (This will become clear as we refine our understanding of (2.2.3) in the sequel.) For completeness, we now state our result.

Proposition 2.2.1 *Let $(X_n)_{n\ge1}$ be a sequence of iid random variables drawn uniformly from the ordered finite alphabet $\mathcal{A} = \{\alpha_1, \ldots, \alpha_m\}$. Then*

$$\frac{LI_n - n/m}{\sqrt{2n/m}} \Rightarrow -\frac{1}{m}\sum_{i=1}^{m-1} i\tilde{B}^i(1) + \max_{\substack{0\le t_1 \le \cdots \\ \le t_{m-1} \le 1}} \sum_{i=1}^{m-1} \tilde{B}^i(t_i), \tag{2.2.4}$$

where $(\tilde{B}^1(t), \ldots, \tilde{B}^{m-1}(t))$ *is an* $(m-1)$*-dimensional Brownian with covariance matrix given by* (2.2.2).

For $m = 2$, (2.2.4) simply becomes

$$\frac{LI_n - n/2}{\sqrt{n}} \Rightarrow -\frac{1}{2}B(1) + \max_{0 \le t \le 1} B(t), \tag{2.2.5}$$

where B is standard one-dimensional Brownian motion. A well-known result of Pitman [39] implies that, up to a factor of 2, the functional in (2.2.5) is identical in law to the radial part of a three-dimensional standard Brownian motion at time $t = 1$. Specifically, Pitman shows that the process $(2\max_{0 \le s \le t} B(s) - B(t))_{t \ge 0}$ is identical in law to $\left(\sqrt{(B^1(t))^2 + (B^2(t))^2 + (B^3(t))^2}\right)_{t \ge 0}$, where $(B^1(t), B^2(t), B^3(t))_{t \ge 0}$ is a standard 3-dimensional Brownian motion.

Let us now show that the functional in (2.2.5) does indeed have the same distribution as that of the largest eigenvalue of a 2×2 zero-trace matrix of the form

$$\begin{pmatrix} X & Y + iZ \\ Y - iZ & -X \end{pmatrix},$$

where X, Y, and Z are centered independent normal random variables, all with variance $1/4$. These random variables have a joint density given by

$$f_3(x, y, z) = \left(\frac{2}{\pi}\right)^{3/2} e^{-2x^2 - 2y^2 - 2z^2}, \quad (x, y, z) \in \mathbb{R}^3.$$

It is straightforward to show that the largest eigenvalue of our matrix is given by $\lambda_1 = \sqrt{X^2 + Y^2 + Z^2}$. Thus, up to a scaling factor of 2, λ_1 is equal in law to the radial Brownian motion expression given by Pitman at $t = 1$. Explicitly, since $4\lambda_1^2 = 4X^2 + 4Y^2 + 4Z^2$ consists of the sum of the squares of three iid standard normal random variables, $4\lambda_1^2$ must have a χ^2 distribution with 3 degrees of freedom. Since this distribution has a density of $h(x) = (1/\sqrt{2\pi})x^{1/2}e^{-x/2}$, we immediately find that λ_1 has density

$$\begin{aligned} g(\lambda_1) &= \frac{1}{\sqrt{2\pi}}(4\lambda_1^2)^{1/2}e^{-(4\lambda_1^2)/2}(8\lambda_1) \\ &= \frac{16}{\sqrt{2\pi}}\lambda_1^2 e^{-2\lambda_1^2}, \quad \lambda_1 > 0. \end{aligned}$$

Let us look now at the connection between the 2×2 GUE and the traceless matrix we have just analyzed. Consider the 2×2 matrix

$$\begin{pmatrix} X_1 & Y + iZ \\ Y - iZ & X_2 \end{pmatrix},$$

where X_1, X_2, Y, and Z are independent normal random variables, with $\operatorname{Var} X_1 = \operatorname{Var} X_2 = 1/2$, and with $\operatorname{Var} Y = \operatorname{Var} Z = 1/4$. Since these random variables have a joint density given by

$$f_4(x_1, x_2, y, z) = \frac{2}{\pi^2}e^{-x_1^2-x_2^2-2y^2-2z^2}, \quad (x_1, x_2, y, z) \in \mathbb{R}^4,$$

conditioning on the zero-trace subspace $\{X_1 + X_2 = 0\}$, and using the transformation $X_1' = (X_1 - X_2)/\sqrt{2}$ and $X_2' = (X_1 + X_2)/\sqrt{2}$, we obtain the conditional density

$$f_3(x_1', y, z) = \left(\frac{2}{\pi}\right)^{3/2} e^{-2(x_1')^2-2y^2-2z^2},$$

which is also the joint density of three iid centered normal random variables X_1', Y, and Z with common variance $1/4$. Note also that the traceless GUE model may be obtained from the GUE by simply subtracting the trace of the GUE from each diagonal. (See Xu [46] for further developments of this sort for more general random matrices.)

Let us finally note that one can directly evaluate (2.2.5) in a classical manner using the Reflection Principle to obtain the corresponding density (see, *e.g.* [22, 27]).

It is instructive to express (2.2.4) in terms of an $(m-1)$-dimensional standard Brownian motion $(B^1(t), \ldots, B^{m-1}(t))$. It is not hard to check that we can express $\tilde{B}^i(t)$, $i = 1, \ldots, m-1$, in terms of the $B^i(t)$ as follows:

$$\tilde{B}^i(t) = \begin{cases} B^1(t), & i = 1, \\ \sqrt{\frac{i+1}{2i}} B^i(t) - \sqrt{\frac{i-1}{2i}} B^{i-1}(t), & 2 \le i \le m-1. \end{cases} \tag{2.2.6}$$

Substituting (2.2.6) back into (2.2.4), we obtain a more symmetric expression for our limiting distribution:

$$\frac{LI_n - n/m}{\sqrt{n}} \Rightarrow \frac{1}{\sqrt{m}} \max_{\substack{0 \le t_1 \le \cdots \\ \le t_{m-1} \le t_m = 1}} \sum_{i=1}^{m-1} \left[-\sqrt{\frac{i}{i+1}} B^i(t_{i+1}) + \sqrt{\frac{i+1}{i}} B^i(t_i) \right]. \tag{2.2.7}$$

The above Brownian functional is similar to one introduced by Glynn and Whitt [20], in the context of a queuing problem:

$$D_m = \max_{\substack{0 = t_0 \le t_1 \le \cdots \\ \le t_{m-1} \le t_m = 1}} \sum_{i=1}^{m} \left[B^i(t_i) - B^i(t_{i-1}) \right], \tag{2.2.8}$$

where $(B^1(t), \ldots, B^m(t))$ is an m-dimensional standard Brownian motion. Gravner, Tracy, and Widom [22], in studying a one-dimensional discrete space and discrete time process, have shown that its limiting distribution is equal to both that of D_m and also that of the largest eigenvalue $\lambda_1^{(m)}$ of an $m \times m$ Hermitian matrix taken from a GUE. That is, D_m and $\lambda_1^{(m)}$ are in fact identical in law. Independently, Baryshnikov [6], studying closely related problems of Queuing Theory and of monotonous paths on the integer lattice, has shown that the *process* $(D_m)_{m \ge 1}$ has the same law as the *process* $(\lambda_1^{(m)})_{m \ge 1}$, where $\lambda_1^{(m)}$ is the largest eigenvalue of the matrix consisting of the first m rows and m columns of an infinite matrix in the Gaussian Unitary Ensemble.

Remark 2.2.1 *It is quite clear that $LI_n \geq n/m$, since at least one of the m letters must lie on a substring of length at least n/m. Hence, the limiting functional in* (2.2.4) *must be supported on the positive real line. We can also see directly that the functional on the right hand side of* (2.2.7) *is non-negative. Indeed, for consider the more general Brownian functional of the form*

$$\max_{\substack{0 \leq t_1 \leq \cdots \\ \leq t_{m-1} \leq t_m = 1}} \sum_{i=1}^{m-1} \left[\beta_i B^i(t_{i+1}) - \eta_i B^i(t_i)\right],$$

where $0 \leq \beta_i \leq \eta_i$, for $i = 1, 2, \ldots, m-1$. Now for any fixed $t_{i+1} \in (0, 1]$, $i = 1, \ldots, m-1$, $\max_{0 \leq t_i \leq t_{i+1}} [\beta_i B^i(t_{i+1}) - \eta_i B^i(t_i)]$ is at least as large as the maximum value at the two extremes, that is, when $t_i = 0$ or $t_i = t_{i+1}$. These two values are simply $\beta_i B^i(t_{i+1})$ and $(\beta_i - \eta_i)B^i(t_{i+1})$. Since $0 \leq \beta_i \leq \eta_i$, at least one of these two values is non-negative. Hence, we can successively find $t_{m-1}, t_{m-2}, \ldots, t_1$ such that each term of the functional is non-negative. Thus the whole functional must be non-negative. Taking $\beta_i = \sqrt{i/(i+1)}$ and $\eta_i = \sqrt{(i+1)/i}$, the result holds for (2.2.7). *The functional of Glynn and Whitt in* (2.2.8) *does not succumb to the same analysis since the $i = 1$ term demands that $t_0 = 0$.*

Let us now turn our attention to the m-letter case wherein each letter α_r occurs with probability $0 < p_r < 1$, independently, and the p_r need not be equal as in the previous uniform case. For the non-uniform case, Its, Tracy, and Widom in [28] and [29] obtained the limiting distribution of LI_n. Reordering the probabilities such that $p_1 \geq p_2 \geq \cdots \geq p_m$, and grouping those probabilities having identical values $p_{(j)}$ of multiplicity k_j, $j = 1, \ldots, d$, (so that $\sum_{j=1}^d k_j = m$ and $\sum_{j=1}^d p_{(j)} k_j = 1$), they show that the limiting distribution is identical to the distribution of the largest eigenvalue associated with the $k_1 \times k_1$ block of a direct sum of d mutually independent $k_j \times k_j$ GUEs, whose eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_m) = (\lambda_1^{k_1}, \lambda_2^{k_1}, \ldots, \lambda_{k_1}^{k_1}, \ldots, \lambda_1^{k_d}, \lambda_2^{k_d}, \ldots, \lambda_{k_d}^{k_d})$ satisfy $\sum_{i=1}^m \sqrt{p_i}\lambda_i = 0$. With the above ordering of the probabilities, the limiting distribution simplifies to an integral involving only p_1 and k_1. (See Remark 2.3.4 for some explicit expressions and more details.) We now state our own result in terms of functionals of Brownian motion.

Theorem 2.2.1 *Let* $(X_n)_{n\geq 1}$ *be a sequence of iid random variables taking values in an ordered finite alphabet* $\mathcal{A} = \{\alpha_1, \ldots, \alpha_m\}$*, such that* $\mathbb{P}(X_1 = \alpha_r) = p_r$*, for* $r = 1, \ldots, m$*, where* $0 < p_r < 1$ *and* $\sum_{r=1}^m p_r = 1$*. Then*

$$\frac{LI_n - p_{max}n}{\sqrt{n}} \Rightarrow -\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \tilde{B}^i(1) + \max_{\substack{0=t_0\leq t_1\leq\cdots\\ \leq t_{m-1}\leq t_m=1\\ t_i=t_{i-1},\ i\in I^*}} \sum_{i=1}^{m-1} \sigma_i \tilde{B}^i(t_i), \tag{2.2.9}$$

where $p_{max} = \max_{1\leq r\leq m} p_r$*,* $\sigma_r^2 = p_r + p_{r+1} - (p_r - p_{r+1})^2$*,* $I^* = \{r \in \{1, \ldots, m\} : p_r < p_{max}\}$*, and where* $(\tilde{B}^1(t), \ldots, \tilde{B}^{m-1}(t))$ *is an* $(m-1)$*-dimensional Brownian motion with covariance matrix given by*

$$t\begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \cdots & \rho_{1,m-1} \\ \rho_{2,1} & 1 & \rho_{2,3} & \cdots & \rho_{2,m-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & 1 & \rho_{m-2,m-1} \\ \rho_{m-1,1} & \rho_{m-1,2} & \cdots & \rho_{m-1,m-2} & 1 \end{pmatrix},$$

with

$$\rho_{r,s} = \begin{cases} -\frac{p_r+\mu_r\mu_s}{\sigma_r\sigma_s}, & s = r-1, \\ -\frac{p_s+\mu_r\mu_s}{\sigma_r\sigma_s}, & s = r+1, \\ -\frac{\mu_r\mu_s}{\sigma_r\sigma_s}, & |r-s| > 1, \quad 1 \leq r, s \leq m-1, \end{cases}$$

and with $\mu_r = p_r - p_{r+1}, 1 \leq r \leq m-1$*.*

Proof. As before, we begin with the expression for LI_n displayed in (2.1.5), noting that for each letter α_r, $1 \leq r \leq m-1$, $(Z_k^r)_{k\geq 1}$ forms a sequence of iid random variables, and that, moreover, Z_k^r and Z_ℓ^s are independent for $k \neq \ell$, and for any r

and s. Now, however, for each fixed k, the Z_k^r are no longer identically distributed; indeed,

$$\begin{cases} \mu_r := \mathbb{E} Z_1^r = p_r - p_{r+1}, & 1 \leq r \leq m-1, \\ \sigma_r^2 := \operatorname{Var} Z_1^r = p_r + p_{r+1} - (p_r - p_{r+1})^2, & 1 \leq r \leq m-1. \end{cases} \tag{2.2.10}$$

Since $0 < p_r < 1$, we have $\sigma_r^2 > 0$ for all $1 \leq r \leq m-1$. We are thus led to define our Brownian approximation by

$$\hat{B}_n^r(t) := \frac{S_{[nt]}^r - \mu_r [nt]}{\sigma_r \sqrt{n}} + (nt - [nt]) \frac{Z_{[nt]+1}^r - \mu_r}{\sigma_r \sqrt{n}}, \qquad 0 \leq t \leq 1, \quad 1 \leq r \leq m-1.$$

Again noting that the local maxima of $\hat{B}_n^i(t)$ occur on the set $\{t : t = k/n, k = 0, \ldots, n\}$, (2.1.5) becomes

$$\begin{aligned} LI_n = \frac{n}{m} - \frac{1}{m} \sum_{i=1}^{m-1} i \left[\sigma_i \hat{B}_n^i(1) \sqrt{n} + \mu_i n \right] \\ + \max_{\substack{0 = t_0 \leq t_1 \leq \cdots \\ \leq t_{m-1} \leq t_m = 1}} \left\{ \sum_{i=1}^{m-1} \left[\sigma_i \hat{B}_n^i(t_i) \sqrt{n} + \mu_i t_i n \right] \right\}. \end{aligned} \tag{2.2.11}$$

Next,

$$\begin{aligned} \sum_{i=1}^{m-1} i \mu_i &= \sum_{i=1}^{m-1} \sum_{j=i}^{m-1} \mu_j = \sum_{i=1}^{m-1} \sum_{j=i}^{m-1} (p_j - p_{j+1}) \\ &= \sum_{i=1}^{m-1} (p_i - p_m) = (1 - p_m) - (m-1) p_m \\ &= 1 - m p_m. \end{aligned}$$

Hence, (2.2.11) becomes

$$\begin{aligned} LI_n = \frac{n}{m} - \frac{(1 - m p_m) n}{m} - \frac{1}{m} \sum_{i=1}^{m-1} i \sigma_i \hat{B}_n^i(1) \sqrt{n} \\ + \max_{\substack{0 = t_0 \leq t_1 \leq \cdots \\ \leq t_{m-1} \leq t_m = 1}} \sum_{i=1}^{m-1} \left[\sigma_i \hat{B}_n^i(t_i) \sqrt{n} + \mu_i t_i n \right], \end{aligned} \tag{2.2.12}$$

and, dividing through by $\sqrt{n}$, we obtain

$$\frac{LI_n}{\sqrt{n}} = p_m\sqrt{n} - \frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}_n^i(1) + \max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-1}\le t_m=1}} \sum_{i=1}^{m-1}\left[\sigma_i\hat{B}_n^i(t_i) + \mu_i t_i\sqrt{n}\right]. \tag{2.2.13}$$

Let $t_0 = 0$, and let $\Delta_i = t_i - t_{i-1}$, $i = 1, \ldots, m-1$. Since

$$\sum_{i=1}^{m-1}\mu_i t_i = \sum_{i=1}^{m-1}\mu_i\sum_{j=1}^{i}\Delta_j = \sum_{i=1}^{m-1}\Delta_i\sum_{j=i}^{m-1}\mu_j = \sum_{i=1}^{m-1}\Delta_i(p_i - p_m),$$

(2.2.13) becomes

$$\frac{LI_n}{\sqrt{n}} = p_m\sqrt{n} - \frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}_n^i(1) + \max_{\substack{\Delta_i\ge 0\\ \sum_{i=1}^{m-1}\Delta_i\le 1}} \left\{\sum_{i=1}^{m-1}\sigma_i\hat{B}_n^i(t_i) + \sqrt{n}\sum_{i=1}^{m-1}\Delta_i(p_i - p_m)\right\}, \tag{2.2.14}$$

where $t_i = \sum_{j=1}^{i}\Delta_j$.

Recalling that $t_m := 1$, and setting $\Delta_m = 1 - t_{m-1}$, (2.2.14) enjoys a more symmetric representation as

$$\frac{LI_n}{\sqrt{n}} = -\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}_n^i(1) + \max_{\substack{\Delta_i\ge 0\\ \sum_{i=1}^{m}\Delta_i=1}} \left[\sum_{i=1}^{m-1}\sigma_i\hat{B}_n^i(t_i) + \sqrt{n}\sum_{i=1}^{m}\Delta_i p_i\right]. \tag{2.2.15}$$

Next,

$$\frac{LI_n - p_{max}n}{\sqrt{n}} = -\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}_n^i(1) + \max_{\substack{\Delta_i\ge 0\\ \sum_{i=1}^{m}\Delta_i=1}} \left[\sum_{i=1}^{m-1}\sigma_i\hat{B}_n^i(t_i) + \sqrt{n}\sum_{i=1}^{m}\Delta_i(p_i - p_{max})\right], \tag{2.2.16}$$

where $p_{max} = \max_{1 \le i \le m} p_i$. Clearly, if $\Delta_i > 0$ for any i such that $p_i < p_{max}$, then

$$\sqrt{n} \sum_{i=1}^{m} \Delta_i (p_i - p_{max}) \xrightarrow{a.s.} -\infty.$$

Intuitively, then, we should demand that $\Delta_i = 0$ for $i \in I^* := \{i \in \{1, 2, \ldots, m\} : p_i < p_{max}\}$. Indeed, we now show that in fact

$$\frac{LI_n - p_{max} n}{\sqrt{n}} = -\frac{1}{m} \sum_{i=1}^{m-1} i \sigma_i \hat{B}_n^i(1) + \max_{\substack{0 = t_0 \le t_1 \le \cdots \\ \le t_{m-1} \le t_m = 1 \\ t_i = t_{i-1},\ i \in I^*}} \sum_{i=1}^{m-1} \sigma_i \hat{B}_n^i(t_i) + E_n, \tag{2.2.17}$$

where the remainder term E_n is a random variable converging to zero in probability as $n \to \infty$.

To see this, let us introduce the following notation. Writing $t = (t_1, t_2, \ldots, t_m)$, let $T = \{t : 0 \le t_1 \le \cdots \le t_{m-1} \le t_m = 1\}$ and let $T^* = \{t \in T : t_i = t_{i-1}, i \in I^*\}$. Setting $C_n(t) = \sum_{i=1}^{m-1} \sigma_i \hat{B}_n^i(t_i)$ and $R(t) = \sum_{i=1}^{m} (t_i - t_{i-1})(p_{max} - p_i)$, we can rewrite the terms involving max in (2.2.16) and (2.2.17) as

$$\max_{t \in T} \left[C_n(t) - \sqrt{n} R(t) \right]$$

and

$$\max_{t \in T^*} C_n(t).$$

By the compactness of T and T^* and the continuity of $C_n(t)$ and $R(t)$, we see that for each n and each $\omega \in \Omega$, there is a $\tau^n \in T$ and a $\tau_*^n \in T^*$ such that

$$C_n(\tau^n) - \sqrt{n} R(\tau^n) = \max_{t \in T} \left[C_n(t) - \sqrt{n} R(t) \right],$$

and

$$C_n(\tau_*^n) = \max_{t \in T^*} C_n(t).$$

(Note that the piecewise-linear nature of $C_n(t)$ and the linear nature of $R(t)$ imply that the arguments maximizing the above must lie on a finite set; thus, the measurablility of τ^n and τ_*^n is trivial.)

Now we first claim that the set of optimizing arguments $\{\tau^n\}_{n=1}^{\infty}$ a.s. does not have an accumulation point lying outside of T^*. Suppose the contrary, namely that for each ω in a set A of positive measure, there is a subsequence $(\tau^{n_k})_{k=1}^{\infty}$ of $(\tau^n)_{n=1}^{\infty}$ such that $d(\tau^{n_k}, T^*) > \epsilon$, for some $\epsilon > 0$, where the metric d is the one induced by the L_∞-norm over T, *i.e.*, by $\|t\|_\infty = \max_{1\le i\le m} |t_i|$.

Then, since $T^* \subset T$, it follows that, for all n,

$$C_n(\tau^n) - \sqrt{n}R(\tau^n) \ge C_n(\tau^n_*).$$

Now if $p_{max} = p_m$, then $t = (0, \dots, 0, 1) \in T^*$, and if for some $1 \le j \le m-1$ we have $p_{max} = p_j > \max_{j+1\le i\le m} p_i$, then $t = (0, \dots, 0, 1, \dots, 1) \in T^*$, where there are $j-1$ zeros in t. Hence $C_{n_k}(\tau^{n_k}_*) \ge C_{n_k}(0, \dots, 0, 1, \dots, 1) = \sum_{i=j}^{m-1} \sigma_i \hat{B}^i_{n_k}(1)$, where the sum is taken to be zero for $j = m$. Given $0 < \delta < 1$, by the Central Limit Theorem, we can find a sufficiently negative real α such that

$$\begin{aligned}
\mathbb{P}\left(C_{n_k}(\tau^{n_k}) - \sqrt{n_k}R(\tau^{n_k}) \ge \alpha\right) &\ge \mathbb{P}\left(C_{n_k}(\tau^{n_k}_*) \ge \alpha\right) \\
&\ge \mathbb{P}\left(\sum_{i=j}^{m-1} \sigma_i \hat{B}^i_{n_k}(1) \ge \alpha\right) \\
&> 1 - \delta,
\end{aligned}$$

for n_k large enough. In particular, this implies that

$$\mathbb{P}\left(A \cap \{C_{n_k}(\tau^{n_k}) - \sqrt{n_k}R(\tau^{n_k}) \ge \alpha\}\right) > \frac{1}{2}\mathbb{P}(A), \tag{2.2.18}$$

for n_k large enough.

Next, note that for any $t \in T$, we can modify its components t_i to obtain an element of T^*, by collapsing certain consecutive t_is to single values, where $i \in \{j-1, j, \dots, \ell\}$ and $\{j, j+1, \dots, \ell\} \subset I^*$. With this observation, it is not hard to see that by replacing such maximal consecutive sets of components $\{t_i\}_{i=j-1}^{\ell}$ with their median values, we must have

$$d(\tau^{n_k}, T^*) = \max_{\{(j,\ell):\{j,j+1,\ldots,\ell\}\subset I^*\}} \frac{(\tau_\ell^{n_k} - \tau_{j-1}^{n_k})}{2}.$$

Writing $p_{(2)}$ for the largest of the $p_i < p_{max}$, we see that for all k, and for almost all $\omega \in A$,

$$\begin{aligned}
R(\tau^{n_k}) &= \sum_{i=1}^{m} (\tau_i^{n_k} - \tau_{i-1}^{n_k})(p_{max} - p_i) \\
&= \sum_{i\in I^*} (\tau_i^{n_k} - \tau_{i-1}^{n_k})(p_{max} - p_i) \\
&\geq (p_{max} - p_{(2)}) \sum_{i\in i^*} (\tau_i^{n_k} - \tau_{i-1}^{n_k}) \\
&\geq 2(p_{max} - p_{(2)}) d(\tau^{n_k}, T^*) \geq 2(p_{max} - p_{(2)})\epsilon.
\end{aligned}$$

Now by Donsker's Theorem and the Continuous Mapping Theorem, we have that

$$\max_{t\in T} C_n(t) \Rightarrow \max_{t\in T} \sum_{i=1}^{m-1} \sigma_i \tilde{B}^i(t_i),$$

as $n_k \to \infty$, where $(\tilde{B}^1(t), \ldots, \tilde{B}^{m-1}(t))$ is an $(m-1)$-dimensional Brownian motion described in greater detail below. The point here is simply that this limiting functional exists. Moreover,

$$\max_{t\in T} C_n(t) \geq C_n(\tau^n),$$

hence, given $0 < \delta < 1$, if M is chosen large enough, then

$$\begin{aligned}
\mathbb{P}\left(C_{n_k}(\tau^{n_k}) \leq M\right) &\geq \mathbb{P}\left(\max_{t\in T} C_{n_k}(t) \leq M\right) \\
&> 1 - \delta,
\end{aligned}$$

for n_k large enough.

We can next see how the boundedness of $R(\tau^{n_k})$ on A influences that of the whole expression $C_{n_k}(\tau^{n_k}) - \sqrt{n_k}R(\tau^{n_k})$ via the following estimates. Given $M > 0$ as above, if k is large enough, then

$$n_k \geq ((M - \alpha + 1)/(2(p_{max} - p_{(2)})\epsilon))^2,$$

and also

$$\begin{aligned}
&\mathbb{P}\left(A \cap \{C_{n_k}(\tau^{n_k}) - \sqrt{n_k}R(\tau^{n_k}) \leq \alpha - 1\}\right) \\
&\quad = \mathbb{P}\left(A \cap \{C_{n_k}(\tau^{n_k}) \leq \alpha - 1 + \sqrt{n_k}R(\tau^{n_k})\}\right) \\
&\quad \geq \mathbb{P}\left(A \cap \{C_{n_k}(\tau^{n_k}) \leq \alpha - 1 + \sqrt{n_k}(2(p_{max} - p_{(2)})\epsilon)\}\right) \\
&\quad \geq \mathbb{P}\left(A \cap \{C_{n_k}(\tau^{n_k}) \leq M\}\right) \\
&\quad > \frac{1}{2}\mathbb{P}(A).
\end{aligned}$$

But this contradicts (2.2.18); thus, our optimal parameter sequences $(\tau^n)_{n=1}^{\infty}$ must a.s. have their accumulation points in T^*.

Thus, given $\epsilon > 0$, there is an integer N_ϵ such that the set $A_{n,\epsilon} = \{d(\tau^k, T^*) < \epsilon^3, k \geq n\}$ satisfies $\mathbb{P}(A_{n,\epsilon}) \geq 1 - \epsilon$, for all $n \geq N_\epsilon$. Now, for each τ^n, define $\hat{\tau}^n \in T^*$ to be the (not necessarily unique) point of T^* which is closest in the L^∞-distance to τ^n. Recalling that

$$E_n = C_n(\tau^n) - \sqrt{n}R(\tau^n) - C_n(\tau_*^n) \geq 0,$$

and noting that $R(t) \geq 0$, for all $t \in T$, we can estimate the remainder term E_n as follows: for $n \geq N_\epsilon$,

$$\begin{aligned}
\mathbb{P}(E_n \geq \epsilon) &= \mathbb{P}\left(\{E_n \geq \epsilon\} \cap A_{n,\epsilon}\right) + \mathbb{P}\left(\{E_n \geq \epsilon\} \cap A_{n,\epsilon}^c\right) \\
&\leq \mathbb{P}\left(\{E_n \geq \epsilon\} \cap A_{n,\epsilon}\right) + \mathbb{P}\left(A_{n,\epsilon}^c\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(\{E_n \geq \epsilon\} \cap A_{n,\epsilon}\right) + \epsilon \\
&= \mathbb{P}\left(\{C_n(\tau^n) - \sqrt{n}R(\tau^n) - C_n(\tau^{n*}) \geq \epsilon\} \cap A_{n,\epsilon}\right) + \epsilon \\
&\leq \mathbb{P}\left(\{C_n(\tau^n) - \sqrt{n}R(\tau^n) - C_n(\hat{\tau}^n) \geq \epsilon\} \cap A_{n,\epsilon}\right) + \epsilon \\
&\leq \mathbb{P}\left(\{C_n(\tau^n) - C_n(\hat{\tau}^n) \geq \epsilon\} \cap A_{n,\epsilon}\right) + \epsilon \\
&\leq \mathbb{P}\left(\left|\sum_{i=1}^{m-1} \sigma_i(\hat{B}_n^i(\tau_i^n) - \hat{B}_n^i(\hat{\tau}_i^n))\right| \geq \epsilon\right) + \epsilon. \qquad (2.2.19)
\end{aligned}
$$

To further bound the right-hand side of (2.2.19), note that for all $n \geq 1$ and all $1 \leq i \leq m-1$, we have $\mathrm{Var}(\hat{B}_n^i(t_i) - \hat{B}_n^i(s_i)) = |t_i - s_i|$. Then, let $(s,t) \in T \times T$ be such that $\|t-s\|_\infty \leq \epsilon^3$. Using the Bienaymé-Chebyshev inequality, we find that for n large enough,

$$
\begin{aligned}
\mathbb{P}\left(\left|\sum_{i=1}^{m-1} \sigma_i(\hat{B}_n^i(t_i) - \hat{B}_n^i(s_i))\right| \geq \epsilon\right) &\leq \epsilon^{-2}(m-1)^2 \max_{1\leq i\leq m-1} \sigma_i^2 \|t-s\|_\infty \\
&\leq \epsilon^{-2}(m-1)^2 \max_{1\leq i\leq m-1} \sigma_i^2 \epsilon^3 \\
&= \epsilon(m-1)^2 \max_{1\leq i\leq m-1} \sigma_i^2.
\end{aligned}
$$

Since $\|\tau^n - \hat{\tau}^n\| < \epsilon^3$, for $n \geq N_\epsilon$, this can be used to bound (2.2.19):

$$
\begin{aligned}
\mathbb{P}\left(|E_n| \geq \epsilon\right) &< \mathbb{P}\left(\left|\sum_{i=1}^{m-1} \sigma_i(\hat{B}_n^i(\tau_i^n) - \hat{B}_n^i(\hat{\tau}_i^n))\right| \geq \epsilon\right) + \epsilon \\
&\leq \epsilon\left\{(m-1)^2 \max_{1\leq i\leq m-1} \sigma_i^2 + 1\right\}.
\end{aligned}
$$

Finally, ϵ being arbitrary, we have indeed shown that $E_n \to 0$ in probability.

Applying Donsker's Theorem, the Continuous Mapping Theorem, and Slutsky's (or the converging-together) Theorem [7, 17] to (2.2.17) we finally have:

$$
\frac{LI_n - p_{max}n}{\sqrt{n}} \Rightarrow -\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \tilde{B}^i(1) + \max_{\substack{0=t_0\leq t_1\leq\cdots\\ \leq t_{m-1}\leq t_m=1\\ t_i=t_{i-1},\ i\in I^*}} \sum_{i=1}^{m-1} \sigma_i \tilde{B}^i(t_i), \qquad (2.2.20)
$$

where $(\tilde{B}^1(t), \ldots, \tilde{B}^{m-1}(t))$ is an $(m-1)$-dimensional Brownian motion covariance matrix, $t(\rho_{r,s})_{r,s}$, where

$$\rho_{r,s} = \begin{cases} 1, & r = s, \\ -\frac{p_r + \mu_r \mu_s}{\sigma_r \sigma_s}, & s = r-1, \\ -\frac{p_s + \mu_r \mu_s}{\sigma_r \sigma_s}, & s = r+1, \\ -\frac{\mu_r \mu_s}{\sigma_r \sigma_s}, & |r-s| > 1, \quad 1 \leq r, s \leq m-1. \end{cases}$$

Now for $t = \ell/n$, and $1 \leq r \leq s \leq m-1$, the covariance structure above is computed as follows:

$$\begin{aligned}
\operatorname{Cov}(\hat{B}_n^r(t), \hat{B}_n^s(t)) &= \operatorname{Cov}\left(\sum_{i=1}^{\ell} \frac{Z_i^r - \mu_r}{\sigma_r \sqrt{n}}, \sum_{i=1}^{\ell} \frac{Z_i^s - \mu_s}{\sigma_s \sqrt{n}}\right) \\
&= \frac{1}{n\sigma_r\sigma_s} \operatorname{Cov}\left(\sum_{i=1}^{\ell}(Z_i^r - \mu_r), \sum_{i=1}^{\ell}(Z_i^s - \mu_s)\right) \\
&= \frac{1}{n\sigma_r\sigma_s} \sum_{i=1}^{\ell} \operatorname{Cov}(Z_i^r - \mu_r, Z_i^s - \mu_s) \\
&= \frac{\ell}{n\sigma_r\sigma_s} \operatorname{Cov}(Z_1^r - \mu_r, Z_1^s - \mu_s) \\
&= t \begin{cases} \frac{1}{\sigma_r\sigma_s}\sigma_r\sigma_s, & s = r, \\ \frac{1}{\sigma_r\sigma_s}(0 - \mu_r\mu_s - \mu_r\mu_s + \mu_r\mu_s), & s > r+1, \\ \frac{1}{\sigma_r\sigma_s}(-p_s - \mu_r\mu_s - \mu_r\mu_s + \mu_r\mu_s), & s = r+1, \end{cases} \\
&= t \begin{cases} 1 & s = r, \\ -\frac{\mu_r\mu_s}{\sigma_r\sigma_s} & s > r+1, \\ -\frac{(p_s + \mu_r\mu_s)}{\sigma_r\sigma_s} & s = r+1, \end{cases}
\end{aligned}$$

using the properties of the Z_k^r noted at the beginning of the proof. ∎

We now study (2.2.9) on a case-by-case basis. First, let $I^* = \emptyset$, that is, let

$p_i = 1/m$, for $i = 1, \ldots, m$. Then $\sigma_i^2 = 2p_i = 2/m$, for all $i \in \{1, 2, \ldots, m\}$. Hence, simply rescaling (2.2.9) by $\sqrt{2/m}$ recovers the uniform result in (2.2.4).

Next, consider the case where $p_{max} = p_j$, for *precisely one* $j \in \{1, \ldots, m\}$. We then have $I^* = \{1, 2, \ldots, m\} \setminus \{j\}$. This forces us to set $0 = t_0 = t_1 = \cdots = t_{j-1}$ and $t_j = t_{j+1} = \cdots = t_{m-1} = t_m = 1$, in the maximizing term in (2.2.9) . This leads to the following result, where, below, $(LI_n - p_{max}n)/\sqrt{n}$ converges to a centered normal random variable. Intuitively, this result is not surprising since the longest increasing subsequence is, asymptotically, a string consisting primarily of the most frequently occurring letter, a string whose length is approximated by a binomial random variable with parameters n and p_{max}. We show below that the variance of the limiting normal distribution is, in fact, equal to $p_{max}(1 - p_{max})$.

Corollary 2.2.1 *If $p_{max} = p_j$ for precisely one $j \in \{1, \ldots, m\}$, then*

$$\frac{LI_n - p_{max}n}{\sqrt{n}} \Rightarrow -\frac{1}{m} \sum_{i=1}^{m-1} i\sigma_i \tilde{B}^i(1) + \sum_{i=j}^{m-1} \sigma_i \tilde{B}^i(1), \tag{2.2.21}$$

where the last term in (2.2.21) *is not present if $j = m$.*

Proof. One could compute the variance of the right hand side of (2.2.21) directly to verify that it is in fact $p_{max}(1-p_{max})$. However, the nature of the covariance structure of the Brownian motion makes the calculation somewhat cumbersome. Instead, we revisit the approximation to our Brownian motion in the first term on the right hand side of (2.2.21). In doing this, we not only recover the variance of the limiting distribution, but also see that our approximating functional does indeed take the form of the sum of a binomial random variable and of a term which converges to zero in probability.

From the very definition of the approximation, we have

$$-\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}_n^i(1) = -\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \left[\frac{S_n^i - \mu_i n}{\sigma_i\sqrt{n}}\right]$$
$$= \frac{1}{\sqrt{n}}\left[-\frac{1}{m}\sum_{i=1}^{m-1} iS_n^i + \frac{n}{m}\sum_{i=1}^{m-1} i\mu_i\right]. \qquad (2.2.22)$$

Recalling that $-\frac{1}{m}\sum_{i=1}^{m-1} iS_n^i = a_n^m - \frac{n}{m}$, and that $\sum_{i=1}^{m-1} i\mu_i = 1 - mp_m$, (2.2.22) becomes

$$\frac{1}{\sqrt{n}}\left[\left(a_n^m - \frac{n}{m}\right) + \frac{n}{m}(1 - mp_m)\right] = \frac{1}{\sqrt{n}}(a_n^m - np_m). \qquad (2.2.23)$$

Turning to the second term on the right hand side of (2.2.21) and noting that for $1 \le j < k \le m-1$, $\sum_{i=j}^{k} \mu_i = p_j - p_{k+1}$ and that $\sum_{i=j}^{k} S_r^i = a_r^j - a_r^{k+1}$, for $1 \le r \le n$, we then have

$$\sum_{i=j}^{m-1} \sigma_i \hat{B}_n^i(1) = \frac{1}{\sqrt{n}}\left[\sum_{i=j}^{m-1} S_n^i - n\sum_{i=j}^{m-1}\mu_i\right]$$
$$= \frac{1}{\sqrt{n}}\left[(a_n^j - a_n^m) - n(p_j - p_m)\right]$$
$$= \frac{1}{\sqrt{n}}\left[(a_n^j - np_j) - (a_n^m - np_m)\right]. \qquad (2.2.24)$$

We saw in (2.2.17) that we could write $(LI_n - p_{max}n)/\sqrt{n}$, as the sum of a functional approximating a Brownian motion and of an error term E_n converging to zero in probability. In the present case, this expression simplifies to

$$-\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}^i(1) + \sum_{i=j}^{m-1}\sigma_i \hat{B}^i(1) + E_n = \frac{a_n^j - np_j}{\sqrt{n}} + E_n, \qquad (2.2.25)$$

using (2.2.22)–(2.2.24).

Now a_n^j is a binomial random variable with parameters n and $p = p_j = p_{max}$. By the Central Limit Theorem and the converging together lemma, the right hand side of (2.2.25) converges to a $N(0, p_{max}(1 - p_{max}))$ distribution, while by Donsker's

Theorem, the left hand side converges to the Brownian functional obtained in (2.2.21). Hence, $(LI_n - p_{max}n)/\sqrt{n} \Rightarrow N(0, p_{max}(1 - p_{max}))$, as claimed. ■

Let us now study what happens when $p_{max} = p_j = p_k$, $1 \leq j < k \leq m$, and $p_i < p_{max}$ otherwise, that is, when *precisely two letters* have the maximal probability. We then have $I^* = \{1, \dots, m\} \backslash \{j, k\}$. This requires that

$$
\begin{aligned}
&0 = t_0 = t_1 = \cdots = t_{j-1}, \\
&t_j = t_{j+1} = \cdots = t_{k-1}, \\
&t_k = t_{k+1} = \cdots = t_m = 1.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\max_{\substack{0 = t_0 \leq t_1 \leq \cdots \\ \leq t_{m-1} \leq t_m = 1}} \sum_{i=1}^{m-1} \sigma_i \tilde{B}^i(t_i) &= \max_{0 \leq t \leq 1} \left[\sum_{i=j}^{k-1} \sigma_i \tilde{B}^i(t) + \sum_{i=k}^{m-1} \sigma_i \tilde{B}^i(1) \right] \\
&= \sum_{i=k}^{m-1} \sigma_i \tilde{B}^i(1) + \max_{0 \leq t \leq 1} \sum_{i=j}^{k-1} \sigma_i \tilde{B}^i(t).
\end{aligned}
$$

Thus the limiting law is

$$
-\frac{1}{m} \sum_{i=1}^{m-1} i \sigma_i \tilde{B}_n^i(1) + \sum_{i=k}^{m-1} \sigma_i \tilde{B}^i(1) + \max_{0 \leq t \leq 1} \sum_{i=j}^{k-1} \sigma_i \tilde{B}^i(t). \tag{2.2.26}
$$

To consolidate our analysis, we treat the general case for which p_{max} occurs exactly k times among $\{p_1, p_2, \dots, p_m\}$, where $2 \leq k \leq m-1$. Not only will we recover the natural analogues of (2.2.26), but we will also express our results in terms of another functional of Brownian motion which is more symmetric. Combining the $2 \leq k \leq m-1$ case at hand with the $k = 1$ case previously examined, we have the following:

Corollary 2.2.2 *Let $p_{max} = p_{j_1} = p_{j_2} = \cdots = p_{j_k}$ for $1 \leq j_1 < j_2 < \cdots < j_k \leq m$, for some $1 \leq k \leq m-1$, and let $p_i < p_{max}$, otherwise. Then*

$$\frac{LI_n - p_{max}n}{\sqrt{n}} \Rightarrow \sqrt{p_{max}(1-p_{max})} \max_{\substack{0=t_0\leq t_1\leq\cdots \\ \leq t_{k-1}\leq t_k=1}} \sum_{\ell=1}^{k} \left[\tilde{B}^{\ell}(t_\ell) - \tilde{B}^{\ell}(t_{\ell-1})\right], \tag{2.2.27}$$

where the k-dimensional Brownian motion $(\tilde{B}^1(t), \tilde{B}^1(t), \ldots, \tilde{B}^k(t))$ has the covariance matrix

$$t \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \rho & 1 & \rho \\ \rho & \cdots & \cdots & \rho & 1 \end{pmatrix}, \tag{2.2.28}$$

with $\rho = -p_{max}/(1-p_{max})$.

Proof. Let $p_{max} = p_{j_1} = p_{j_1} = \cdots = p_{j_k}$, with $1 \leq j_1 < j_2 < \cdots < j_k \leq m$ and $2 \leq k \leq m-1$, *i.e.*, let $I^* = \{1, 2, \ldots, m\} \backslash \{j_1, j_2, \ldots, j_k\}$. Set $j_0 = 1$ and $j_{k+1} = m$. Then (2.2.17) becomes

$$\begin{aligned} \frac{LI_n - p_{max}n}{\sqrt{n}} &= -\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}_n^i(1) \\ &\quad + \max_{\substack{0=t_0\leq t_1\leq\cdots \\ \leq t_{m-1}\leq t_m=1 \\ t_i=t_{i-1},\ i\in I^*}} \sum_{i=1}^{m-1} \sigma_i \hat{B}_n^i(t_i) + E_n \\ &= -\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}_n^i(1) \\ &\quad + \max_{\substack{0=t_{j_0}\leq t_{j_1}\leq\cdots \\ \leq t_{j_k}\leq t_{j_{k+1}}=1}} \sum_{\ell=0}^{k} \sum_{i=j_\ell}^{j_{\ell+1}-1} \sigma_i \hat{B}_n^i(t_{j_l}) + E_n \\ &= -\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i \hat{B}_n^i(1) + \end{aligned}$$

$$
+ \max_{\substack{0=t_{j_0}\le t_{j_1}\le\cdots\\ \le t_{j_k}\le t_{j_{k+1}}=1}} \left[\sum_{\ell=1}^{k-1}\sum_{i=j_\ell}^{j_{\ell+1}-1}\sigma_i\hat{B}_n^i(t_{j_\ell}) + \sum_{i=j_k}^{m-1}\sigma_i\hat{B}_n^i(1)\right] + E_n
$$

$$
= \left[-\frac{1}{m}\sum_{i=1}^{m-1} i\sigma_i\hat{B}_n^i(1) + \sum_{i=j_k}^{m-1}\sigma_i\hat{B}_n^i(1)\right]
$$

$$
+ \max_{\substack{0=t_{j_0}\le t_{j_1}\le\cdots\\ \le t_{j_k}\le t_{j_{k+1}}=1}} \sum_{\ell=1}^{k-1}\sum_{i=j_\ell}^{j_{\ell+1}-1}\sigma_i\hat{B}_n^i(t_{j_\ell}) + E_n. \tag{2.2.29}
$$

We immediately recognize the first term on the right hand side of (2.2.29) as what we encountered for $k=1$. Using the definition of the $\hat{B}_n^i$, (2.2.29) can then be rewritten as

$$
\frac{a_n^{j_k} - np_{max}}{\sqrt{n}} + \max_{\substack{0=t_{j_0}\le t_{j_1}\le\cdots\\ \le t_{j_k}\le t_{j_{k+1}}=1}} \sum_{\ell=1}^{k-1}\sum_{i=j_\ell}^{j_{\ell+1}-1}\sigma_i\hat{B}_n^i(t_{j_\ell}) + E_n
$$

$$
= \frac{a_n^{j_k} - np_{max}}{\sqrt{n}} + \max_{\substack{0=t_{j_0}\le t_{j_1}\le\cdots\\ \le t_{j_k}\le t_{j_{k+1}}=1}} \sum_{\ell=1}^{k-1}\sum_{i=j_\ell}^{j_{\ell+1}-1}\sigma_i\left(\frac{S^i_{[nt_{j_\ell}]} - \mu_i\,[nt_{j_\ell}]}{\sigma_i\sqrt{n}}\right) + E_n
$$

$$
= \frac{a_n^{j_k} - np_{max}}{\sqrt{n}}
$$

$$
+ \frac{1}{\sqrt{n}} \max_{\substack{0=t_{j_0}\le t_{j_1}\le\cdots\\ \le t_{j_k}\le t_{j_{k+1}}=1}} \sum_{\ell=1}^{k-1}\left(\left(a^{j_\ell}_{[nt_{j_\ell}]} - a^{j_{\ell+1}}_{[nt_{j_\ell}]}\right) - [nt_{j_\ell}]\left(p_{j_\ell} - p_{j_{\ell+1}}\right)\right)
$$

$$
+ E_n
$$

$$
= \frac{a_n^{j_k} - np_{max}}{\sqrt{n}}
$$

$$
+ \frac{1}{\sqrt{n}} \max_{\substack{0=t_{j_0}\le t_{j_1}\le\cdots\\ \le t_{j_k}\le t_{j_{k+1}}=1}} \sum_{\ell=1}^{k-1}\left(\left(a^{j_\ell}_{[nt_{j_\ell}]} - [nt_{j_\ell}]\,p_{max}\right) - \left(a^{j_{\ell+1}}_{[nt_{j_\ell}]} - [nt_{j_\ell}]\,p_{max}\right)\right)
$$

$$
+ E_n. \tag{2.2.30}
$$

Setting $a_n^{j_{k+1}} = n - \sum_{\ell=1}^{k} a_n^{j_\ell}$, we note that the random vector $(a_n^{j_1}, a_n^{j_2}, \ldots,$ $a_n^{j_{k+1}})$ has a multinomial distribution with parameters n and $(p_{max}, p_{max}, \ldots, p_{max}, 1-$

kp_{max}). It is thus natural to introduce a new Brownian motion approximation as follows:

$$\check{B}_n^\ell(t) = \frac{a^{j_\ell}_{[nt_{j_\ell}]} - [nt_{j_\ell}]\, p_{max}}{\sqrt{np_{max}(1-p_{max})}}, \quad 1 \le \ell \le k. \tag{2.2.31}$$

Substituting (2.2.31) into (2.2.30) gives

$$\begin{aligned} &\sqrt{p_{max}(1-p_{max})}\left\{ \check{B}_n^k(1) + \max_{\substack{0=t_0\le t_1 \le \cdots \\ \le t_{k-1}\le t_k=1}} \sum_{\ell=1}^{k-1} \left[\check{B}_n^\ell(t_\ell) - \check{B}_n^{\ell+1}(t_\ell)\right] \right\} + E_n \\ &= \sqrt{p_{max}(1-p_{max})} \max_{\substack{0=t_0\le t_1 \le \cdots \\ \le t_{k-1}\le t_k=1}} \sum_{\ell=1}^{k} \left[\check{B}_n^\ell(t_\ell) - \check{B}_n^\ell(t_{\ell-1})\right] + E_n. \end{aligned} \tag{2.2.32}$$

By Donsker's Theorem, $(\check{B}_n^1(t), \check{B}_n^2(t), \ldots, \check{B}_n^k(t))$ converges jointly to a k-dimensional Brownian motion $(\tilde{B}^1(t), \tilde{B}^2(t), \ldots, \tilde{B}^k(t))$. This Brownian motion has the covariance structure

$$t \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \rho & 1 & \rho \\ \rho & \cdots & \cdots & \rho & 1 \end{pmatrix},$$

where $\rho = -p_{max}/(1-p_{max})$, a fact which follows immediately from the covariance of the multinomial distribution, where the covariance of any two distinct $a_r^{j_\ell}$ is simply $-rp_{max}^2$, for $1 \le r \le n$. This, together with our analysis of the unique p_{max} case, proves the corollary. ■

Remark 2.2.2 *The above results provide a Brownian functional equivalent to the GUE result of Its, Tracy, and Widom [28] (described in detail in the comments preceding Theorem 2.2.1 and with a law given in Remark 2.3.4). Note that the limiting* *distribution in* (2.2.27) *depends only on* k *and* p_{max}*; neither the specific values of* $j_1, j_2, \ldots, j_k$ *nor the remaining values of* p_i *are material, a fact already noted in [28]. Also, it follows from generic results on Brownian functionals that this limiting law has a density, which in the uniform case is supported on the positive real line, while supported on all of* $\mathbb{R}$ *in the non-uniform case.*

We have already seen in (2.2.7) that the limiting distribution for the uniform case has a nice representation as a functional of standard Brownian motion. We now also express the limiting distribution in (2.2.27) as a functional of standard Brownian motion. This new functional extends to the uniform case, although its form is different from that of (2.2.7). This limiting random variable can be viewed as the sum of a normal one and of a maximal eigenvalue type one.

Corollary 2.2.3 *Let* $p_{max} = p_{j_1} = p_{j_2} = \cdots = p_{j_k}$*, for* $1 \le j_1 < j_2 < \cdots < j_k \le m$*, and some* $1 \le k \le m$*, and let* $p_i < p_{max}$*, otherwise. Then*

$$\frac{LI_n - p_{max}n}{\sqrt{n}} \Rightarrow \sqrt{p_{max}}\Bigg\{\frac{\sqrt{1-kp_{max}}-1}{k}\sum_{j=1}^{k} B^j(1) + \max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{k-1}\le t_k=1}} \sum_{\ell=1}^{k}\left[B^\ell(t_\ell) - B^\ell(t_{\ell-1})\right]\Bigg\}. \tag{2.2.33}$$

where $(B^1(t), B^2(t), \ldots, B^k(t))$ *is a standard* k*-dimensional Brownian motion.*

Proof. Let us first examine the non-uniform case $1 \le k \le m-1$. Recall that $\rho = -p_{max}/(1-p_{max})$. Now the covariance matrix in (2.2.28) has eigenvalues $\lambda_1 = 1-\rho = 1/(1-p_{max})$ of multiplicity $k-1$ and $\lambda_2 = 1+(k-1)\rho = (1-kp_{max})/(1-p_{max}) < \lambda_1$ of multiplicity 1. From the symmetries of the covariance matrix, it is not hard to see that we can write each Brownian motion $\tilde{B}^i(t)$ as a linear combination of standard Brownian motions $(B^1(t), \ldots, B^k(t))$ as follows:

$$\tilde{B}^i(t) = \beta B^i(t) + \eta \sum_{j=1, j\neq i}^{k} B^j(t), \qquad i = 1, \ldots, k, \tag{2.2.34}$$

where

$$\beta = \frac{(k-1)\sqrt{\lambda_1} + \sqrt{\lambda_2}}{k}, \qquad \eta = \frac{-\sqrt{\lambda_1} + \sqrt{\lambda_2}}{k}. \tag{2.2.35}$$

Substituting (2.2.34) and (2.2.35) into (2.2.27), and noting that $\beta - \eta = \sqrt{\lambda_1} = 1/\sqrt{1 - p_{max}}$, we find that

$$\begin{aligned}
&\sqrt{p_{max}(1-p_{max})} \max_{\substack{0=t_0 \le t_1 \le \cdots \\ \le t_{k-1} \le t_k = 1}} \sum_{\ell=1}^{k} \left[\tilde{B}^\ell(t_\ell) - \tilde{B}^\ell(t_{\ell-1})\right] \\
&= \sqrt{p_{max}(1-p_{max})} \max_{\substack{0=t_0 \le t_1 \le \cdots \\ \le t_{k-1} \le t_k = 1}} \sum_{\ell=1}^{k} \Big\{ \beta \left[B^\ell(t_\ell) - B^\ell(t_{\ell-1})\right] \\
&\qquad + \eta \sum_{j=1, j\neq \ell}^{k} \left[B^j(t_\ell) - B^j(t_{\ell-1})\right] \Big\} \\
&= \sqrt{p_{max}(1-p_{max})} \max_{\substack{0=t_0 \le t_1 \le \cdots \\ \le t_{k-1} \le t_k = 1}} \sum_{\ell=1}^{k} \Big\{ (\beta - \eta) \left[B^\ell(t_\ell) - B^\ell(t_{\ell-1})\right] \\
&\qquad + \eta \sum_{j=1}^{k} \left[B^j(t_\ell) - B^j(t_{\ell-1})\right] \Big\} \\
&= \sqrt{p_{max}(1-p_{max})} \max_{\substack{0=t_0 \le t_1 \le \cdots \\ \le t_{k-1} \le t_k = 1}} \Big\{ \sum_{\ell=1}^{k} (\beta - \eta) \left[B^\ell(t_\ell) - B^\ell(t_{\ell-1})\right] \\
&\qquad + \eta \sum_{\ell=1}^{k} \sum_{j=1}^{k} \left[B^j(t_\ell) - B^j(t_{\ell-1})\right] \Big\} \\
&= \sqrt{p_{max}(1-p_{max})} \Big\{ \eta \sum_{j=1}^{k} B^j(1) \\
&\qquad + (\beta - \eta) \max_{\substack{0=t_0 \le t_1 \le \cdots \\ \le t_{k-1} \le t_k = 1}} \sum_{\ell=1}^{k} \left[B^\ell(t_\ell) - B^\ell(t_{\ell-1})\right] \Big\} \\
&= \sqrt{p_{max}} \Big\{ \frac{\sqrt{1 - kp_{max}} - 1}{k} \sum_{j=1}^{k} B^j(1)
\end{aligned}$$

$$+ \max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{k-1}\le t_k=1}} \sum_{\ell=1}^{k} \left[B^\ell(t_\ell) - B^\ell(t_{\ell-1})\right]\Bigg\}. \tag{2.2.36}$$

To complete the proof, we now examine the uniform case $k = m$, where necessarily $p_{max} = 1/m$. We saw in Proposition 2.2.1 that

$$\frac{LI_n - n/m}{\sqrt{n}} \Rightarrow \sqrt{\frac{2}{m}}\Bigg\{-\frac{1}{m}\sum_{i=1}^{m-1} i\tilde{B}^i(1) + \max_{\substack{0\le t_1\le\cdots\\ \le t_{m-1}\le 1}} \sum_{i=1}^{m-1} \tilde{B}^i(t_i)\Bigg\}, \tag{2.2.37}$$

where the $(m-1)$-dimensional Brownian motion $(\tilde{B}^1(t), \dots, \tilde{B}^{m-1}(t))$ had a tridiagonal covariance matrix given by (2.2.2). Now we can derive this Brownian motion from a standard m-dimensional Brownian motion $(B^1(t), \dots, B^m(t))$ via the a.s. transformations

$$\tilde{B}^i(t) = \frac{1}{\sqrt{2}}(B^i(t) - B^{i+1}(t)), \qquad 1 \le i \le m-1.$$

It is easily verified that the Brownian motion $(\tilde{B}^1(t), \dots, \tilde{B}^{m-1}(t))$ so obtained does indeed have the covariance structure given by (2.2.2). Substituting these independent Brownian motions into (2.2.37), we obtain the following a.s equalities:

$$\begin{aligned}
\frac{LI_n - n/m}{\sqrt{n}} &\Rightarrow \sqrt{\frac{2}{m}}\Bigg\{-\frac{1}{m}\sum_{i=1}^{m-1} i\tilde{B}^i(1) + \max_{\substack{0\le t_1\le\cdots\\ \le t_{m-1}\le 1}} \sum_{i=1}^{m-1} \tilde{B}^i(t_i)\Bigg\} \\
&= \sqrt{\frac{1}{m}}\Bigg\{-\frac{1}{m}\sum_{i=1}^{m-1} i[B^i(1) - B^{i+1}(1)] \\
&\qquad + \max_{\substack{0\le t_1\le\cdots\\ \le t_{m-1}\le 1}} \sum_{i=1}^{m-1}[B^i(t_i) - B^{i+1}(t_i)]\Bigg\} \\
&= \sqrt{\frac{1}{m}}\Bigg\{-\frac{1}{m}\sum_{i=1}^{m} B^i(1) + B^m(1) \\
&\qquad + \max_{\substack{0\le t_1\le\cdots\\ \le t_{m-1}\le 1}} \sum_{i=1}^{m}[B^i(t_i) - B^i(t_{i-1})] - B^m(1)\Bigg\}
\end{aligned}$$

$$= \sqrt{\frac{1}{m}} \left\{ -\frac{1}{m} \sum_{i=1}^{m} B^i(1) + \max_{\substack{0 \le t_1 \le \cdots \\ \le t_{m-1} \le 1}} \sum_{i=1}^{m} [B^i(t_i) - B^i(t_{i-1})] \right\}, \qquad (2.2.38)$$

which give (2.2.33), with $k = m$ and $p_{max} = 1/m$. ■

We have already seen several representations for the limiting law in the uniform case. Yet one more pleasing functional for the limiting distribution of LI_n is described in the following

Theorem 2.2.2 *Let $p_{max} = p_1 = p_2 = \cdots = p_m = 1/m$. Then*

$$\frac{LI_n - n/m}{\sqrt{n}} \Rightarrow \frac{\tilde{H}_m}{\sqrt{m}},$$

where

$$\tilde{H}_m = \sqrt{\frac{m-1}{m}} \max_{\substack{0 = t_0 \le t_1 \le \cdots \\ \le t_{m-1} \le t_m = 1}} \sum_{i=1}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1}) \right], \qquad (2.2.39)$$

and where $(\tilde{B}^1(t), \tilde{B}^2(t), \ldots, \tilde{B}^m(t))$ is an m-dimensional Brownian motion having covariance matrix (2.2.28)*, with $\rho = -1/(m-1)$, and thus such that $\sum_{i=1}^{m} \tilde{B}^i(t) = 0$, for all $0 \le t \le 1$.*

Proof. We show that the functional being maximized in (2.2.39) has the same covariance structure as the functional being maximized in (2.2.7), a result which we restate as:

$$\frac{LI_n - n/m}{\sqrt{n}} \Rightarrow \frac{1}{\sqrt{m}} \max_{\substack{0 \le t_1 \le \cdots \\ \le t_{m-1} \le t_m = 1}} \sum_{i=1}^{m-1} \left[\beta_i B^i(t_{i+1}) - \eta_i B^i(t_i) \right], \qquad (2.2.40)$$

where $\beta_i = \sqrt{i/(i+1)}$ and $\eta_i = \sqrt{(i+1)/i}$. From this it will immediately follow that the maxima, over all $0 \le t_1 \le t_2 \le \cdots \le t_{m-1} \le 1$, in both expressions have the same law, clinching the proof.

Let $(\tilde{B}^1(t), \tilde{B}^2(t), \dots, \tilde{B}^m(t))$ be an m-dimensional Brownian motion with a permutation-invariant covariance matrix described by

$$\mathrm{Cov}(\tilde{B}^i(t), \tilde{B}^j(t)) = -\frac{t}{m-1}, \quad i \neq j,$$

and

$$\mathrm{Var}\, \tilde{B}^i(t) = t.$$

Hence, $\mathbb{E}(\sum_{i=1}^m \tilde{B}^i(t))^2 = 0$, for all $0 \leq t \leq 1$, so that $\sum_{i=1}^m \tilde{B}^i(t)$ is identically equal to zero.

Let $t = (t_1, t_2, \dots, t_{m-1})$ be a fixed collection of t_i from the Weyl chamber $T = \{(t_1, t_2, \dots, t_{m-1}) : 0 \leq t_1 \leq t_2 \leq \cdots \leq t_{m-1} \leq 1\}$. Setting $t_m = 1$, and also setting

$$X_t = \sqrt{\frac{m-1}{m}} \sum_{i=1}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1}) \right], \tag{2.2.41}$$

we then have

$$\begin{aligned}
\mathrm{Cov}(X_t, X_s) &= \frac{m-1}{m} \sum_{1 \leq i,j \leq m} \mathrm{Cov}(\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1}), \tilde{B}^j(s_j) - \tilde{B}^j(s_{j-1})) \\
&= \frac{m-1}{m} \sum_{i=1}^{m} \left[t_i \wedge s_i - t_i \wedge s_{i-1} - t_{i-1} \wedge s_i + t_{i-1} \wedge s_{i-1} \right] \\
&\quad - \frac{1}{m} \sum_{i \neq j} \left[t_i \wedge s_j - t_i \wedge s_{j-1} - t_{i-1} \wedge s_j + t_{i-1} \wedge s_{j-1} \right].
\end{aligned} \tag{2.2.42}$$

We can rewrite (2.2.42) in a clear way by setting $T_1 = [0, t_1]$ and $T_i = (t_i, t_{i+1}]$, $i = 2, \dots, m-1$, and similarly $S_1 = [0, s_1]$ and $S_i = (s_i, s_{i+1}]$, $i = 2, \dots, m-1$. Letting *Leb* denote the Lebesgue measure on $[0, 1]$, a case-by-case analysis of the relative positions of t_i, t_{i-1}, s_i, and s_{i-1} quickly yields that

$$\begin{aligned}
\operatorname{Cov}(X_t, X_s) &= \frac{m-1}{m}\sum_{i=1}^{m} Leb(T_i \cap S_i) - \frac{1}{m}\sum_{i\neq j} Leb(T_i \cap S_j) \\
&= \frac{m-1}{m}\sum_{i=1}^{m} Leb(T_i \cap S_i) - \frac{1}{m}\left[1 - \sum_{i=1}^{m} Leb(T_i \cap S_i)\right] \\
&= -\frac{1}{m} + \sum_{i=1}^{m} Leb(T_i \cap S_i). \qquad (2.2.43)
\end{aligned}$$

To complete the proof, we now show that

$$Y_t = \sum_{i=1}^{m-1}\left[\beta_i B^i(t_{i+1}) - \eta_i B^i(t_i)\right], \qquad (2.2.44)$$

has the same covariance structure as X_t, where $\beta_i = \sqrt{i/(i+1)}$ and $\eta_i = \sqrt{(i+1)/i}$. Using the independence of the components of the Brownian motion, we also have

$$\begin{aligned}
\operatorname{Cov}(Y_t, Y_s) &= \sum_{i=1}^{m-1} \operatorname{Cov}\left(\beta_i B^i(t_{i+1}) - \eta_i B^i(t_i), \beta_i B^i(s_{i+1}) - \eta_i B^i(s_i)\right) \\
&= \sum_{i=1}^{m-1}\left[\frac{i}{i+1}(t_{i+1} \wedge s_{i+1}) - t_{i+1} \wedge s_i - t_i \wedge s_{i+1} + \frac{i+1}{i}(t_i \wedge s_i)\right] \\
&= \sum_{i=1}^{m}\frac{i-1}{i} t_i \wedge s_i - \sum_{i=1}^{m-1}\left[t_{i+1} \wedge s_i + t_i \wedge s_{i+1} - \frac{i+1}{i} t_i \wedge s_i\right] \\
&= \frac{m-1}{m} - \sum_{i=1}^{m-1}\left[t_{i+1} \wedge s_i + t_i \wedge s_{i+1} - 2(t_i \wedge s_i)\right]. \qquad (2.2.45)
\end{aligned}$$

As before, a simple case-by-case analysis of the summands in (2.2.45) reveals that

$$\begin{aligned}
\operatorname{Cov}(Y_t, Y_s) &= \frac{m-1}{m} - \left[1 - \sum_{i=1}^{m} Leb(T_i \cap S_i)\right] \\
&= -\frac{1}{m} + \sum_{i=1}^{m} Leb(T_i \cap S_i), \qquad (2.2.46)
\end{aligned}$$

completing the proof. ■

2.3 Large-m Asymptotics and Related Results

With the covariance structure of X_t now in hand, and the help of (2.2.43), we can compute the L^2-distance between any X_t and X_s:

$$\begin{aligned}\mathbb{E}(X_t - X_s)^2 &= \operatorname{Var} X_t + \operatorname{Var} X_s - 2\operatorname{Cov}(X_t, X_s) \\ &= 2(1 - 1/m) - 2\left[-1/m + \sum_{i=1}^{m} Leb(T_i \cap S_i)\right] \\ &= 2\left[1 - \sum_{i=1}^{m} Leb(T_i \cap S_i)\right]. \qquad (2.3.1)\end{aligned}$$

Such a metric is useful, for instance, in applying Dudley's Entropy Bound to show that $\limsup_{m\to\infty} \mathbb{E}\left(\max_{t\in T} X_t\right)/\sqrt{m}$ is bounded above by a constant. To obtain this constant, we argue as follows. Now for s and t in the Weyl chamber T, we can check that if $\max_{1\le i\le m} |t_i - s_i| < 1/k$, for some integer k, then

$$\begin{aligned}\mathbb{E}(X_t - X_s)^2 &= 2\left[1 - \sum_{i=1}^{m} Leb(T_i \cap S_i)\right] \\ &= 2\sum_{i=1}^{m} Leb(T_i^C \cap S_i) \\ &< \frac{2m}{k}.\end{aligned}$$

Now consider the points of $t \in T$ for which each t_i is of the form j/k for some $j = 0, 1, \ldots, k$. Let us try to cover T with balls of radius $\varepsilon > 0$ centered at each such point of T. Then, clearly, if $k > 2m/\varepsilon^2$, then these $N(\varepsilon)$ balls will indeed cover T. Since in that case $N(\varepsilon) < (2m/\varepsilon)^m/m!$, and since $\operatorname{Var} X_t = (m-1)/m$ for all $t \in T$, Dudley's Entropy Bound (see Theorem 1 on p. 179 of [33]) gives us that

$$\begin{aligned}\mathbb{E}\tilde{H}_m = \mathbb{E}\left(\max_{t\in T} X_t\right) &\le 4\sqrt{2}\int_0^{\frac{1}{2}\sqrt{\frac{m-1}{m}}} \sqrt{N(\varepsilon)} \quad d\varepsilon \\ &< 4\sqrt{2}\int_0^{\frac{1}{2}} \sqrt{-2m\log\varepsilon + m\log(2m) - \log m!} \quad d\varepsilon,\end{aligned}$$

so that, by Stirling's formula,

$$\limsup_{m\to\infty} \frac{\tilde{H}_m}{\sqrt{m}} < 8\int_0^{\frac{1}{2}} \sqrt{-\log\varepsilon + (\log 2 + 1)/2} \quad d\varepsilon$$
$$\approx 8(0.7843) = 6.2744.$$

One can also obtain a lower bound for $\liminf_{m\to\infty} \mathbb{E}\tilde{H}_m/\sqrt{m}$ using the following direct argument. First, note that, almost surely,

$$\begin{aligned}
&\max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-1}\le t_m=1}} \sum_{i=2}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right] \\
&\quad \ge \mathbf{1}_{\{\tilde{B}^1(1/m)>0\}} \left(\tilde{B}^1(1/m) + \max_{\substack{1/m=t_1\le t_2\le\cdots\\ \le t_{m-1}\le t_m=1}} \sum_{i=2}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right] \right) \\
&\qquad + \mathbf{1}_{\{\tilde{B}^1(1/m)\le 0\}} \left(\tilde{B}^2(1/m) + \max_{\substack{1/m=t_2\le t_3\le\cdots\\ \le t_{m-1}\le t_m=1}} \sum_{i=3}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right] \right).
\end{aligned} \tag{2.3.2}$$

Here the idea is that if $\tilde{B}^1(1/m) > 0$, then we "keep" $\tilde{B}^1$ for the interval $[0, 1/m]$; otherwise, we keep $\tilde{B}^2$ for $[0, 1/m]$. Then, taking expectations in (2.3.2), and using the scaling and Markovian properties of Brownian motion, and also, crucially, the fact that $\mathbb{E}W_2\mathbf{1}_{\{W_1\le 0\}} = -\rho\sigma_2/\sqrt{2\pi}$ for any centered Gaussian random variables W_1 and W_2 with correlation coefficient ρ and $\operatorname{Var} W_2 = \sigma_2^2$, we find that

$$\begin{aligned}
&\mathbb{E}\left(\max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-1}\le t_m=1}} \sum_{i=2}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right] \right) \\
&\quad \ge \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m}} + \frac{1}{2}\mathbb{E}\left(\max_{\substack{1/m=t_1\le t_2\le\cdots\\ \le t_{m-1}\le t_m=1}} \sum_{i=2}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right] \right) \\
&\qquad + \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m}}\left(\frac{-1}{m-1}\right) + \frac{1}{2}\mathbb{E}\left(\max_{\substack{1/m=t_2\le t_3\le\cdots\\ \le t_{m-1}\le t_m=1}} \sum_{i=3}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right] \right)
\end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m}}\left(\frac{m}{m-1}\right) + \frac{1}{2}\sqrt{\frac{m-1}{m}}\mathbb{E}\left(\max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-2}\le t_{m-1}=1}} \sum_{i=1}^{m-1}\left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right]\right)$$
$$+ \frac{1}{2}\sqrt{\frac{m-1}{m}}\mathbb{E}\left(\max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-3}\le t_{m-2}=1}} \sum_{i=1}^{m-2}\left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right]\right)$$
$$\ge \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m}} + \sqrt{\frac{m-1}{m}}\mathbb{E}\left(\max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-3}\le t_{m-2}=1}} \sum_{i=1}^{m-2}\left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right]\right). \tag{2.3.3}$$

Iterating the inequality (2.3.3), and assuming that m is even, we obtain

$$\mathbb{E}\left(\max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-1}\le t_{m}=1}} \sum_{i=2}^{m}\left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right]\right)$$
$$\ge \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m}} + \sqrt{\frac{m-1}{m}}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m-2}}$$
$$+ \sqrt{\frac{m-1}{m}}\sqrt{\frac{m-3}{m-2}}\mathbb{E}\left(\max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-5}\le t_{m-4}=1}} \sum_{i=1}^{m-4}\left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right]\right)$$
$$\ge 2\left(\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m}}\right) + \sqrt{\frac{m-3}{m}}\mathbb{E}\left(\max_{\substack{0=t_0\le t_1\le\cdots\\ \le t_{m-5}\le t_{m-4}=1}} \sum_{i=1}^{m-4}\left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right]\right)$$
$$\vdots$$
$$\ge \frac{m}{2}\left(\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m}}\right) = \left(\frac{1}{2\sqrt{2\pi}}\right)\sqrt{m} \tag{2.3.4}$$

or, in terms of $\tilde{H}_m$,

$$\mathbb{E}\tilde{H}_m \ge \sqrt{\frac{m-1}{m}}\left(\frac{1}{2\sqrt{2\pi}}\right)\sqrt{m}. \tag{2.3.5}$$

Since $\sqrt{m/(m-1)}\tilde{H}_m \ge \sqrt{(m-1)/(m-2)}\tilde{H}_{m-1}$, almost surely, we conclude that

$$\liminf_{m\to\infty} \frac{\mathbb{E}\tilde{H}_m}{\sqrt{m}} \ge \frac{1}{2\sqrt{2\pi}} \approx 0.1995. \tag{2.3.6}$$

We can also more clearly see the similarities between the functional D_m of Glynn and Whitt in (2.2.8) and that of (2.2.7), which we have shown to have the same law as $\tilde{H}_m$ in (2.2.39). Indeed, the only difference between the functionals is simply that in (2.2.8) the Brownian motions are independent, while in (2.2.39) they are subject to the zero-sum constraint. Gravner, Tracy, and Widom [22] have already remarked that random words could be studied via such Brownian functionals. In fact, a restatement of Corollary 2.2.3 shows that, in law, D_m and $\tilde{H}_m$ differ by a centered normal random variable, as indicated by the next theorem and corollary. This, in turn, will allow us to clearly state more precise asymptotic results for $\tilde{H}_m$ from the known corresponding results for D_m.

Theorem 2.3.1 *Let*

$$H_m = \sqrt{2}\left\{ -\frac{1}{m}\sum_{i=1}^{m-1} i\tilde{B}^i(1) + \max_{\substack{0\le t_1\le\cdots \\ \le t_{m-1}\le 1}} \sum_{i=1}^{m-1} \tilde{B}^i(t_i) \right\},$$

$m \geq 2$, and let $H_1 \equiv 0$ a.s., where $(\tilde{B}^1(t), \ldots, \tilde{B}^{m-1}(t))$ is an $(m-1)$-dimensional Brownian motion with tridiagonal covariance matrix given by (2.2.2). *Let*

$$D_m = \max_{\substack{0=t_0\le t_1\le\cdots \\ \le t_{m-1}\le t_m=1}} \sum_{i=1}^{m} \left[B^i(t_i) - B^i(t_{i-1}) \right],$$

where $(B^1(t), \ldots, B^m(t))$ is a standard m-dimensional Brownian motion, which is related to $(\tilde{B}^1(t), \ldots, \tilde{B}^{m-1}(t))$ by the almost sure identities

$$\tilde{B}^i(t) = \frac{1}{\sqrt{2}}(B^i(t) - B^{i+1}(t)), \qquad 1 \leq i \leq m-1.$$

Then $D_m = Z_m + H_m$ a.s., where Z_m is a centered normal random variable with variance $1/m$, which is given by $Z_m = (1/m)\sum_{i=1}^m B^i(1)$.

Proof. The $m = 1$ case is trivial. For $m \geq 2$, reformulating the proof of Corollary 2.2.3, for the uniform case, in terms of the functionals H_m and D_m shows that, almost surely,

$$\frac{H_m}{\sqrt{m}} = \frac{1}{\sqrt{m}} \left(-\frac{1}{m} \sum_{i=1}^{m} B^i(1) + D_m \right)$$
$$= \frac{1}{\sqrt{m}}(-Z_m + D_m).$$

■

Recalling the definition of $\tilde{H}_m$ from Theorem 2.2.2:

$$\tilde{H}_m := \sqrt{\frac{m-1}{m}} \max_{\substack{0=t_0 \le t_1 \le \cdots \\ \le t_{m-1} \le t_m = 1}} \sum_{i=1}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1}) \right],$$

where $(\tilde{B}^1(t), \tilde{B}^2(t), \ldots, \tilde{B}^m(t))$ is an m-dimensional Brownian motion having covariance matrix (2.2.28), with $\rho = -1/(m-1)$, implying that $\sum_{i=1}^m \tilde{B}^i(t) = 0$, for all $0 \le t \le 1$, we then have

Corollary 2.3.1 *For each $m \ge 1$, $\tilde{H}_m \stackrel{\mathcal{L}}{=} D_m - Z_m$, where $\mathcal{L}$ denotes equality in distribution.*

Proof. Proposition 2.2.1 asserts that

$$\frac{LI_n - n/m}{\sqrt{n}} \Rightarrow \frac{H_m}{\sqrt{m}},$$

as $n \to \infty$, while by Theorem 2.2.2

$$\frac{LI_n - n/m}{\sqrt{n}} \Rightarrow \frac{\tilde{H}_m}{\sqrt{m}},$$

as $n \to \infty$ as well. The conclusion follows from the previous theorem. ■

This relationship between $\tilde{H}_m$ (resp.,H_m) and D_m allows us to further express the limiting distribution in a rather compact form.

Proposition 2.3.1 *Let $p_{max} = p_{j_1} = p_{j_2} = \cdots = p_{j_k}$, for $1 \le j_1 < j_2 < \cdots < j_k \le m$, and some $1 \le k \le m$, and let $p_i < p_{max}$, otherwise. Then*

$$\frac{LI_n - p_{max}n}{\sqrt{n}} \Rightarrow \sqrt{p_{max}}\{\sqrt{1 - kp_{max}}Z_k + H_k\}.$$

Proof. For $k = m$, we have $p_{max} = 1/m$, and thus simply recover the limiting distribution $H_m/\sqrt{m} \stackrel{\mathcal{L}}{=} \tilde{H}_m/\sqrt{m}$ of the uniform case.

For $1 \le k \le m-1$, we saw in Corollary 2.2.3 that we could write the limiting law of $(LI_n - p_{max}n)/\sqrt{n}$ as

$$\begin{aligned} \sqrt{p_{max}} \Bigg\{ & \frac{\sqrt{1-kp_{max}}-1}{k} \sum_{j=1}^{k} B^j(1) \\ & + \max_{\substack{0=t_0 \le t_1 \le \cdots \\ \le t_{k-1} \le t_k = 1}} \sum_{\ell=1}^{k} \left[B^\ell(t_\ell) - B^\ell(t_{\ell-1}) \right] \Bigg\}, \end{aligned} \tag{2.3.7}$$

where $(B^1(t), B^2(t), \ldots, B^k(t))$ is a standard k-dimensional Brownian motion. But, recalling the definitions of D_k and Z_k, and the fact that $D_k = Z_k + H_k$ a.s., (2.3.7) becomes

$$\begin{aligned} & \sqrt{p_{max}} \left\{ \frac{\sqrt{1-kp_{\max}}-1}{k} (kZ_k) + D_k \right\} \\ & = \sqrt{p_{max}} \left\{ \left(\sqrt{1-kp_{max}} - 1 \right) Z_k + (Z_k + H_k) \right\} \\ & = \sqrt{p_{max}} \left\{ \sqrt{1-kp_{max}} Z_k + H_k \right\}. \end{aligned} \tag{2.3.8}$$

■

Remark 2.3.1 *One can also write the limiting law of Proposition 2.3.1 in terms of the functional D_k. Indeed, we have*

$$\frac{LI_n - p_{max}n}{\sqrt{p_{max}n}} \Rightarrow \left\{ \sqrt{1-kp_{max}} - 1 \right\} Z_k + D_k,$$

so that the limiting law is expressed as the sum of a centered normal random variable and of the maximal eigenvalue of a $k \times k$ element of the GUE.

The behavior of D_m has been well-studied. In particular, it has been shown that that $D_m/\sqrt{m} \to 2$ a.s. and in L^1, as $m \to \infty$ (see [6, 19, 20, 26, 36, 37, 42]), and that

$(D_m - 2\sqrt{m})m^{1/6} \Rightarrow F_2$, as $m \to \infty$, where F_2 is the Tracy-Widom distribution (see [6, 22, 44, 43]). From these results, the asymptotics of H_m follows.

Theorem 2.3.2 *We have that*

$$\frac{H_m}{\sqrt{m}} \to 2$$

a.s. and in L^1, as $m \to \infty$. Moreover,

$$\left(\frac{H_m}{\sqrt{m}} - 2\right) m^{2/3} \Rightarrow F_2, \tag{2.3.9}$$

where F_2 is the Tracy-Widom distribution. The same statements hold for $\tilde{H}_m$ in place of H_m.

Proof. From Theorem 2.3.1 we have $D_m = Z_m + H_m$ a.s., where $Z_m = (1/m)\sum_{i=1}^m B^i(1)$.

Clearly, $Z_m \to 0$ a.s. and in L^1. Thus, a.s. and in L^1,

$$\lim_{m\to\infty} \frac{H_m}{\sqrt{m}} = \lim_{m\to\infty} \frac{D_m}{\sqrt{m}}.$$

Since this last limit is 2, and since, for each $m \geq 1$, $H_m \stackrel{\mathcal{L}}{=} \tilde{H}_m$, it also follows that

$$\lim_{m\to\infty} \mathbb{E}\left|\frac{\tilde{H}_m}{\sqrt{m}} - 2\right| = 0.$$

We are thus left with proving the a.s. convergence to 2 of $\tilde{H}_m/\sqrt{m}$. Since the variance of the functional being maximized in the definition of $\tilde{H}_m$ equals $1 - 1/m$, the Gaussian concentration inequality implies that

$$\mathbb{P}(|\tilde{H}_m - \mathbb{E}\tilde{H}_m| > h) \leq 2e^{\frac{-h^2}{2(1-1/m)}} < 2e^{-\frac{h^2}{2}}$$

for all $h > 0$. Then since $\mathbb{E}\tilde{H}_m/\sqrt{m} \to 2$ as $m \to \infty$ we have for m large enough that

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{\tilde{H}_m}{\sqrt{m}} - 2\right| > h\right) &\leq \mathbb{P}\left(\left|\tilde{H}_m - \mathbb{E}\tilde{H}_m\right| > \sqrt{m}\left(h - \left|\frac{\mathbb{E}\tilde{H}_m}{\sqrt{m}} - 2\right|\right)\right) \\
&\leq \mathbb{P}\left(\left|\tilde{H}_m - \mathbb{E}\tilde{H}_m\right| > \frac{\sqrt{m}h}{2}\right) \\
&< 2e^{-\frac{mh^2}{8}}.
\end{aligned}$$

This concentration result implies that, for all $h > 0$,

$$\sum_{m=1}^{\infty} \mathbb{P}\left(\left|\frac{\tilde{H}_m}{\sqrt{m}} - 2\right| > h\right) < \infty,$$

and the Borel-Cantelli lemma allows us to conclude.

Turning to the limiting law, we know ([6, 22]) that D_m has the same distribution as the largest eigenvalue of the $m \times m$ GUE. Then the fundamental random matrix theory result of Tracy and Widom [43] implies that

$$\left(\frac{D_m}{\sqrt{m}} - 2\right) m^{2/3} \Rightarrow F_2.$$

Since, moreover, $D_m = Z_m + H_m$, and since Z_m has variance $1/m$, $Z_m m^{1/6} \Rightarrow 0$, and so

$$\left(\frac{H_m}{\sqrt{m}} - 2\right) m^{2/3} = \left(\frac{D_m}{\sqrt{m}} - 2\right) m^{2/3} - Z_m m^{1/6} \Rightarrow F_2.$$

Finally, $H_m \stackrel{\mathcal{L}}{=} \tilde{H}_m$, and the same result holds for $\tilde{H}_m$ in place of H_m. ■

Remark 2.3.2 *(i) In the conclusion to [44], Tracy and Widom already derived* (2.3.9) *by applying a scaling argument to the limiting distribution of the uniform alphabet case. In our case we can moreover assert that a.s. and in the mean,*

$$\lim_{k \to +\infty} \lim_{n \to +\infty} \frac{LI_n - p_{max} n}{\sqrt{k p_{max} n}} = 2,$$

and that

$$\left(\frac{LI_n - p_{max} n}{\sqrt{k p_{max} n}} - 2\right) k^{2/3} \Rightarrow F_2,$$

where the weak limit is first taken over n and then over k, and where p_{max} depends on k, and necessarily decreases to zero, as $k \to \infty$.

(ii) Using scaling, subadditivity, and concentration arguments found in Hambly, Martin, and O'Connell [26] and in O'Connell and Yor [36], one could prove directly that

$\tilde{H}_m/\sqrt{m} \to 2$ *a.s. This could be accomplished by studying, as do these authors, a process version of* $\tilde{H}_m$*, i.e.,*

$$\tilde{H}_m(\varepsilon) := \sqrt{\frac{m-1}{m}} \max_{\substack{0=t_0\le t_1\le\cdots \\ \le t_{m-1}\le t_m=\varepsilon}} \sum_{i=1}^{m} \left[\tilde{B}^i(t_i) - \tilde{B}^i(t_{i-1})\right],$$

for $\varepsilon > 0$*. With obvious notations, for all* $\varepsilon > 0$ *and* $m \ge 1$*,* $D_m(\varepsilon) = Z(\varepsilon) + H_m(\varepsilon)$*, a.s., where* $Z(\varepsilon) = (1/m)\sum_{i=1}^{m} B^i(\varepsilon)$*.*

To see in further detail how D_m and $\tilde{H}_m$ are related, first note that $D_m \le D_{m+1}$ a.s. for $m \ge 1$, since D_m can simply be obtained by restricting the right-most parameter t_m to be 1 in the definition of D_{m+1}. We now show a stochastic domination result between D_m and $\tilde{H}_m$.

Recall that a random variable X is said to *stochastically dominate* another random variable Y ($X \ge_{st} Y$) if, for all $x \in \mathbb{R}$, $\mathbb{P}(X \ge x) \ge \mathbb{P}(Y \ge x)$.

Proposition 2.3.2 $\tilde{H}_m \ge_{st} \sqrt{(1-1/m)}D_m$*, for* $m \ge 1$*. The same statement holds for* H_m *in place of* $\tilde{H}_m$*.*

Proof. Since the $m = 1$ case is trivial, let $m \ge 2$. We saw in (2.2.43) that the functional X_t being maximized in the definition of $\tilde{H}_m$ had a covariance structure given by $\mathrm{Cov}(X_t, X_s) = -1/m + \sum_{i=1}^{m} Leb(T_i \cap S_i)$. A similar argument shows that the functional $U_t = \sum_{i=1}^{m}[B^i(t_i) - B^i(t_{i-1})]$ which is being maximized in the definition of D_m has a covariance structure given by $\mathrm{Cov}(U_t, U_s) = \sum_{i=1}^{m} Leb(T_i \cap S_i)$. Therefore,

$$\mathrm{Var}(\sqrt{(1-1/m)}U_t) = \mathrm{Var}\, X_t = 1 - 1/m,$$

and

$$\begin{aligned}\mathrm{Cov}(\sqrt{(1-1/m)}U_t, \sqrt{(1-1/m)}U_s) &= (1-1/m)\sum_{i=1}^{m} Leb(T_i \cap S_i) \\ &\ge \mathrm{Cov}(X_t, X_s).\end{aligned}$$

By Slepian's Lemma we conclude that $\tilde{H}_m \geq_{st} \sqrt{(1-1/m)}D_m$. The final assertion follows from the equality in law between $\tilde{H}_m$ and H_m. ■

Remark 2.3.3 *Note that*

$$\begin{aligned}\mathbb{E}(X_t - X_s)^2 &= \mathbb{E}(\sqrt{(1-1/m)}U_t - \sqrt{(1-1/m)}U_s)^2 \\ &= 2\left(1 - \sum_{i=1}^{m} Leb(T_i \cap S_i)\right)\end{aligned}$$

for all $s, t \in [0,1]$. That is, while X_t and $\sqrt{(1-1/m)}U_t$ have different covariance structures, the L^2-distance between X_t and X_s is the same as that between $\sqrt{(1-1/m)}U_t$ and $\sqrt{(1-1/m)}U_s$. We then conclude again that $\mathbb{E}\tilde{H}_m = \mathbb{E}D_m$ in a manner independent of Theorem 2.3.1, which implies that $\mathbb{E}\tilde{H}_m = \mathbb{E}H_m = \mathbb{E}(D_m + Z_m) = \mathbb{E}D_m$, since $Z_m \sim N(0, 1/m)$.

Remark 2.3.4 *Let us briefly summarize the connections between random matrix theory and the Brownian functionals encountered in this thesis. Writing, for any $m \geq 1$, $x^{(m)} = (x_1, x_2, \ldots, x_m)$, letting $\Delta(x^{(m)}) = \Pi_{1\leq i<j\leq m}(x_i - x_j)$ be the Vandermonde determinant, and if $dx^{(m)} = dx_1 dx_2 \ldots dx_m$ is the Lebesgue measure on $\mathbb{R}^m$, we have the following facts.*

(i) First, $\lambda_1^{(m)} \stackrel{\mathcal{L}}{=} D_m$, where $\lambda_1^{(m)}$ is the largest eigenvalue of the $m \times m$ GUE, with the scaling taken so that the diagonal elements $X_{i,i}$ satisfy $\mathbb{E}X_{i,i}^2 = 1$, and the off-diagonal elements $X_{i,j}$, for $i \neq j$, satisfy $\mathbb{E}|X_{i,j}|^2 = 1$. Using standard random matrix results (see, e.g., [35]), the distribution of D_m, for all $m \geq 1$ and all $s \in \mathbb{R}$, is given by

$$\mathbb{P}(D_m \leq s) = c_m \int_{A_s} e^{-\frac{1}{2}\sum_{i=1}^{m} x_i^2} \Delta(x^{(m)})^2 dx^{(m)},$$

where

$$A_s = \{x \in \mathbb{R}^m : \max_{1\leq i\leq m} x_i \leq s\},$$

where

$$c_m^{-1} = \int_{\mathbb{R}^m} e^{-\frac{1}{2}\sum_{i=1}^m x_i^2} \Delta(x^{(m)})^2 dx^{(m)}.$$

(ii) Second, from [44] and our results, $\lambda_1^{(m,0)} \stackrel{\mathcal{L}}{=} \tilde{H}_m \stackrel{\mathcal{L}}{=} H_m$*, where* $\lambda_1^{(m,0)}$ *is the largest eigenvalue of the* $m \times m$ *traceless GUE, with the scaling as in (i). Using the joint density of the eigenvalues of the traceless* $m \times m$ *GUE [35, 44], the distribution function of* $\tilde{H}_m$ *can also be computed directly, for all* $m \geq 2$ *and all* $s \geq 0$*, as*

$$\mathbb{P}(\tilde{H}_m \leq s) = c_m^0 \int_{A_s^0} e^{-\frac{m}{2}\sum_{i=1}^m x_i^2} \Delta(x^{(m)})^2 dx^{(m,0)},$$

where $dx^{(m,0)}$ *is the Lebesgue measure over the set* $\{x \in \mathbb{R}^m : \sum_{i=1}^m x_i = 0\}$*, and where*

$$A_s^0 = \{x \in \mathbb{R}^m : \max_{1 \leq i \leq m} x_i \leq s\} \cap \left\{\sum_{i=1}^m x_i = 0\right\},$$

where

$$(c_m^0)^{-1} = \int_{\{\sum_{i=1}^m x_i = 0\}} e^{-\frac{m}{2}\sum_{i=1}^m x_i^2} \Delta(x^{(m)})^2 dx^{(m,0)}.$$

Note that $\tilde{H}_m$ *is a.s. non-negative, and so* $\mathbb{P}(\tilde{H}_m \leq s) = 0$*, for all* $s < 0$*.*

(iii) Third, let $J_k := \sqrt{p_{max}}\{\sqrt{1 - kp_{max}} Z_k + H_k\}$ *be the limiting functional of Proposition 2.3.1 for the m-letter non-uniform case, having its most probable letters of multiplicity* k *occurring with probability* p_{max}*. Using (ii), this functional is equal in law to the sum of a normal random variable and a variable whose distribution, up to the scaling factor* $\sqrt{p_{max}}$*, is that of the largest eigenvalue of the* $k \times k$ *traceless GUE, with the scaling as in (i). Further, from (i) and* $D_k = Z_k + H_k$ *a.s.,* J_k *is also equal in law to the sum of a normal random variable and a variable whose distribution, up to the scaling factor* $\sqrt{p_{max}}$*, is that of the largest eigenvalue of the* $k \times k$ *GUE. Its, Tracy, and Widom [29] show that, for all* $k \geq 1$ *and all* $s \in \mathbb{R}$*,* J_k *has distribution given by*

$$\mathbb{P}(J_k \le s) = c_{k,p_{max}} \int_{A_s} e^{-\frac{1}{2p_{max}}\left[\sum_{i=1}^k x_i^2 + \frac{p_{max}}{1-kp_{max}}(\sum_{i=1}^k x_i)^2\right]} \Delta(x^{(k)})^2 dx^{(k)},$$

where

$$A_s = \{x \in \mathbb{R}^k : \max_{1\le i\le k} x_i \le s\},$$

and where

$$c_{k,p_{max}}^{-1} = \int_{\mathbb{R}^k} e^{-\frac{1}{2p_{max}}\left[\sum_{i=1}^k x_i^2 + \frac{p_{max}}{1-kp_{max}}(\sum_{i=1}^k x_i)^2\right]} \Delta(x^{(k)})^2 dx^{(k)}.$$

Moreover, in the discussion prior to Theorem 2.2.1, we noted that the k-fold integral representation of the limiting distribution of J_k came from simplifying a more complex expression. This expression described the distribution of J_k as that of the largest eigenvalue associated with the $k_1 \times k_1$ submatrix of the matrix consisting of a direct sum of d mutually independent GUEs, each of size $k_j \times k_j$, $1 \le j \le d$, subject to the eigenvalue constraint $\sum_{i=1}^m \sqrt{p_i}\lambda_i = 0$. The k_j were the multiplicities of the probabilities having common values, the p_i were ordered in decreasing order, and the eigenvalues were ordered in terms of the GUEs corresponding to the appropriate values of p_i. So in our notation, k is the multiplicity of p_{max}.

Note that when $k = 1$, the limiting distribution becomes

$$\mathbb{P}(J_k \le s) = \frac{1}{\sqrt{2\pi p_{max}(1-p_{max})}} \int_{-\infty}^{s} e^{-\frac{x^2}{2p_{max}(1-p_{max})}} dx,$$

which is simply a $N(0, p_{max}(1-p_{max}))$ distribution.

(iv) The Tracy-Widom distribution function F_2, which also describes the limiting distribution of $(L\sigma_n - 2\sqrt{n})/n^{1/6}$, (see [5]), is given, for all $t \in \mathbb{R}$, by

$$F_2(t) = \exp\left(-\int_t^{\infty} (x-t)u^2(x)dx\right),$$

where $u(x)$ is the solution to the Painlevé II equation $u_{xx} = 2u^3 + xu$ with $u \sim -Ai(x)$, as $x \to \infty$, where $Ai(x)$ is the Airy function.

CHAPTER III

COUNTABLY INFINITE IID ALPHABETS

Let us now study the problem of describing LI_n for an ordered, countably infinite alphabet $\mathcal{A} = \{\alpha_n\}_{n \geq 1}$, where $\alpha_1 < \alpha_2 < \cdots < \alpha_m < \cdots$. Let $(X_i)_{1 \leq i \leq n}$, $X_i \in \mathcal{A}$, be an iid sequence, with $\mathbb{P}(X_1 = \alpha_r) = p_r > 0$, for $r \geq 1$.

The central idea in the first part of our approach is to introduce two new sequences derived from $(X_i)_{1 \leq i \leq n}$. Fix $m \geq 1$. The first sequence, which we shall term the *capped sequence*, is defined by taking $T_i^m = X_i \wedge \alpha_m$, for $i \geq 1$. The second one, $(Y_i^m)_{1 \leq i \leq N_{n,m}}$, the *reduced sequence*, consists of the subsequence of $(X_i)_{1 \leq i \leq n}$ of length $N_{n,m}$, for which $X_i \leq \alpha_m$, for $i \geq 1$. Thus, the capped sequence $(T_i^m)_{1 \leq i \leq n}$ is obtained by setting to α_m all letter values greater than α_m, while the reduced sequence $(Y_i^m)_{1 \leq i \leq N_{n,m}}$ is obtained by eliminating letter values greater than α_m altogether.

Let $LI_{n,m}^{cap}$ and $LI_{n,m}^{red}$ to be the lengths of the longest increasing subsequence of $(T_i^m)_{1 \leq i \leq n}$ and $(Y_i^m)_{1 \leq i \leq N_{n,m}}$, respectively. Now on the one hand, any subsequence of the reduced sequence is again a subsequence of the original sequence $(X_i)_{1 \leq i \leq n}$. On the other hand, any increasing subsequence of $(X_i)_{1 \leq i \leq n}$ corresponds to an increasing subsequence of the capped one. These two observations lead to the pathwise bounds

$$LI_{n,m}^{red} \leq LI_n \leq LI_{n,m}^{cap}, \tag{3.0.10}$$

for all $m \geq 1$ and $n \geq 1$.

These bounds suggest that the behavior of the iid infinite case perhaps mirrors that of the iid finite-alphabet case. Indeed, we do have the following result, which amounts to an extension of Theorem 2.2.1 (or, more precisely, of Proposition 2.3.1) to the iid infinite-alphabet case.

Theorem 3.0.3 *Let $(X_i)_{1\le i\le n}$ be a sequence of iid random variables taking values in the ordered alphabet $\mathcal{A} = \{\alpha_n\}_{n\ge 1}$. Let $\mathbb{P}(X_1 = \alpha_j) = p_j$, for $j \ge 1$. Let $p_{max} = p_{j_1} = p_{j_2} = \cdots = p_{j_k}$, $1 \le j_1 < j_2 < \cdots < j_k$, $k \ge 1$, and let $p_i < p_{max}$, otherwise. Then*

$$\frac{LI_n - p_{max}n}{\sqrt{n}} \Rightarrow \sqrt{p_{max}}\{\sqrt{1-p_{max}k}Z_k + H_k\} := R(p_{max}, k).$$

The proof of the theorem relies on an understanding of the limiting distributions of $LI^{red}_{n,m}$ and $LI^{cap}_{n,m}$. To this end, let us introduce some more notation. For a finite m-alphabet, and for $W_1, \dots, W_n$ iid with $\mathbb{P}(W_1 = \alpha_r) = q_r > 0$, let $LI_n(q) := LI_n(q_1, \dots, q_m)$ denote the length of the longest increasing subsequence of $(W_i)_{i=1}^n$. For each $m \ge 1$, let also $\pi_m = \sum_{r=1}^m p_r$.

First, let us choose m large enough so that $1 - \pi_{m-1} < p_{max}$. Next, observe that, from the capping at α_m, $LI^{cap}_{n,m}$ is distributed as $LI_n(\tilde{p})$, where $\tilde{p} = (p_1, \dots, p_{m-1}, 1 - \pi_{m-1})$. But since m is chosen large enough, the maximal probability among the entries of $\tilde{p}$ is then p_{max}, of multiplicity k, as for the original infinite alphabet. By Theorem 2.2.1, we thus have

$$\frac{LI_n(\tilde{p}) - p_{max}n}{\sqrt{n}} \Rightarrow R(p_{max}, k), \tag{3.0.11}$$

as $n \to \infty$.

Turning to $LI^{red}_{n,m}$, suppose that the number of elements $N_{n,m}$ of the reduced subsequence $(Y_i^m)_{1\le i\le N_{n,m}}$ is equal to j. Since only the elements of $(X_i)_{1\le i\le n}$ which are at most α_m are left, $LI^{red}_{n,m}$ must be distributed as $LI_j(\hat{p})$, where $\hat{p} = (p_1/\pi_m, \dots, p_m/\pi_m)$. From the way m is chosen, the maximal probability among the entries of $\hat{p}$ is then p_{max}/π_m, of multiplicity k. Invoking again the finite-alphabet result of Theorem 2.2.1, we find that

$$\frac{LI_n(\hat{p}) - (p_{max}/\pi_m)n}{\sqrt{n}} \Rightarrow R\left(\frac{p_{max}}{\pi_m}, k\right), \tag{3.0.12}$$

as $n \to \infty$.

We now relate the two limiting expressions in (3.0.11) and (3.0.12) by the following elementary lemma.

Lemma 3.0.1 *Let $k \geq 1$ be an integer, and let $(q_m)_{m=1}^{\infty}$ be a sequence of reals in $(0, 1/k]$ converging to $q \geq 0$. Then $R(q_m, k) \Rightarrow R(q, k)$, as $m \to \infty$.*

Proof. Assume $q > 0$. Then

$$\begin{aligned} R(q_m, k) &= \sqrt{q_m}\left\{\sqrt{1 - q_m k} Z_k + H_k\right\} \\ &= \sqrt{q_m}\{\sqrt{1 - qk} Z_k + H_k\} \\ &\quad + \sqrt{q_m}\{\sqrt{1 - q_m k} - \sqrt{1 - qk}\} Z_k \\ &= \sqrt{\frac{q_m}{q}} R(q, k) + c_m Z_k, \end{aligned} \tag{3.0.13}$$

where $c_m = \sqrt{1 - q_m k} - \sqrt{1 - qk}$. Since $q_m \to q$ as $m \to \infty$, $c_m \to 0$, and so $c_m Z_k \Rightarrow 0$, as $m \to \infty$. This gives the result. The degenerate case, $q = 0$, is clear. ■

The main idea developed in the proof of Theorem 3.0.3 is now to use the basic inequality (3.0.10) in conjunction with a conditioning argument for $LI_{n,m}^{red}$, in order to apply Lemma 3.0.1, *i.e.,* to use $R(p_{max}/\pi_m, k) \Rightarrow R(p_{max}, k)$, as $m \to \infty$, since $\pi_m \to 1$.

Proof. (Theorem 3.0.3) First, fix an arbitrary $s > 0$. As previously noted in Remark 2.2.2, $R(p_{max}, k)$ is absolutely continuous, with density supported on $\mathbb{R}$ ($\mathbb{R}^+$ in the uniform case), and so s is a continuity point of its distribution function. Next, choose $0 < \epsilon_1 < 1$, and $0 < \delta < 1$, and again note that $(1 + \delta)s$ is also necessarily a continuity point for $R(p_{max}, k)$.

With this choice of ϵ_1, pick $\beta > 0$ such that $\mathbb{P}(Z \geq \beta) < \epsilon_1/2$, where Z is a standard normal random variable. Finally, pick ϵ_2 such that $0 < \epsilon_2 < \epsilon_1 \mathbb{P}(R(p_{max}, k) <$

$(1+\delta)s)$. Such a choice of ϵ_2 can always be made since the support of $R(p_{max}, k)$ includes $\mathbb{R}^+$.

We have seen that, for m large enough, we can bring some finite-alphabet results to bear on the infinite case. In fact, we need a few more technical requirements to complete our proof. Setting $\sigma_m^2 = \pi_m(1-\pi_m)$, we choose large enough m so that:

(i) $1 - \pi_{m-1} < p_{max}$,

(ii) $(s + p_{max}\beta\sigma_m/\pi_m)/\sqrt{\pi_m - \beta\sigma_m} < (1+\delta)s$, and

(iii) $|\mathbb{P}(R(p_{max}, k) < (1+\delta)s) - \mathbb{P}(R(p_{max}/\pi_m, k) < (1+\delta)s)| < \epsilon_2/2$.

The conditions (i) and (ii) are clearly satisfied, since $\pi_m \to 1$ and $\sigma_m \to 0$, as $m \to \infty$. The condition (iii) is also satisfied, as seen by applying Lemma 3.0.1 to $R(p_{max}/\pi_m, k)$, with $\pi_m \to 1$, and since $(1+\delta)s$ is also a continuity point for $R(p_{max}, k)$.

Now recall that $LI_{n,m}^{cap}$ is distributed as $LI_n(\tilde{p})$, where $\tilde{p} = (p_1, \ldots, p_{m-1}, 1-\pi_{m-1})$. Hence, we have from (3.0.10) and (3.0.11) that

$$\frac{LI_n - p_{max}n}{\sqrt{n}} \leq \frac{LI_{n,m}^{cap} - p_{max}n}{\sqrt{n}} \Rightarrow R(p_{max}, k), \tag{3.0.14}$$

as $n \to \infty$, and so

$$\begin{aligned} \mathbb{P}\left(\frac{LI_n - p_{max}n}{\sqrt{n}} \leq s\right) &\geq \mathbb{P}\left(\frac{LI_{n,m}^{cap} - p_{max}n}{\sqrt{n}} \leq s\right) \\ &\to \mathbb{P}(R(p_{max}, k) \leq s), \end{aligned} \tag{3.0.15}$$

as $n \to \infty$.

More work is required to make use of the left-hand minorization in (3.0.10) (*i.e.*, $LI_{n,m}^{red} \leq LI_n$.) Recall that if the length $N_{n,m}$ of the reduced sequence is equal to j, then $LI_{n,m}^{red}$ must be distributed as $LI_j(\hat{p})$, where $\hat{p} = (p_1/\pi_m, \ldots, p_m/\pi_m)$. Now the essential observation is that $N_{n,m}$ is distributed as a binomial random variable with parameters π_m and n. It is thus natural to focus on the values of j close to $\mathbb{E}N_{n,m} = n\pi_m$. Writing the variance of $N_{n,m}$ as $n\sigma_m^2$, where, as above, $\sigma_m^2 = \pi_m(1-\pi_m)$, and setting

$$\gamma_{n,m,j} := \mathbb{P}(N_{n,m} = j) = \binom{n}{j}\pi_m^j(1-\pi_m)^{n-j},$$

we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{LI_{n,m}^{red} - p_{max}n}{\sqrt{n}} \le s\right) \\
&= \sum_{j=0}^{n} \mathbb{P}\left(\frac{LI_{n,m}^{red} - p_{max}n}{\sqrt{n}} \le s \,\middle|\, N_{n,m} = j\right)\gamma_{n,m,j} \\
&= \sum_{j=0}^{n} \mathbb{P}\left(\frac{LI_j(\hat{p}) - p_{max}n}{\sqrt{n}} \le s\right)\gamma_{n,m,j} \\
&= \sum_{j=0}^{n} \mathbb{P}\left(\frac{LI_j(\hat{p}) - \frac{p_{max}}{\pi_m}j}{\sqrt{j}} \le \sqrt{\frac{n}{j}}\left(s + \frac{p_{max}}{\sqrt{n}}\left(n - \frac{j}{\pi_m}\right)\right)\right)\gamma_{n,m,j} \\
&\le \sum_{j=\lceil n\pi_m - \beta\sigma_m\sqrt{n}\rceil}^{n} \mathbb{P}\left(\frac{LI_j(\hat{p}) - \frac{p_{max}}{\pi_m}j}{\sqrt{j}} \le \sqrt{\frac{n}{j}}\left(s + \frac{p_{max}}{\sqrt{n}}\left(n - \frac{j}{\pi_m}\right)\right)\right)\gamma_{n,m,j} \\
&\qquad + \sum_{j=0}^{\lceil n\pi_m - \beta\sigma_m\sqrt{n}\rceil - 1} \gamma_{n,m,j} \\
&< \sum_{j=\lceil n\pi_m - \beta\sigma_m\sqrt{n}\rceil}^{n} \mathbb{P}\left(\frac{LI_j(\hat{p}) - \frac{p_{max}}{\pi_m}j}{\sqrt{j}} \le \sqrt{\frac{n}{j}}\left(s + \frac{p_{max}}{\sqrt{n}}\left(n - \frac{j}{\pi_m}\right)\right)\right)\gamma_{n,m,j} \\
&\qquad + \epsilon_1, \qquad (3.0.16)
\end{aligned}$$

for sufficiently large n, where (3.0.16) follows from the Central Limit Theorem and our choice of β, and where, as usual, $\lceil\cdot\rceil$ is the ceiling function.

Next, note that for $\lceil n\pi_m - \beta\sigma_m\sqrt{n}\rceil \le j \le n$, and making use of the condition (ii),

$$
\begin{aligned}
&\sqrt{\frac{n}{j}}\left(s+\frac{p_{max}}{\sqrt{n}}\left(n-\frac{j}{\pi_m}\right)\right)\\
&\qquad<\sqrt{\frac{n}{n\pi_m-\beta\sigma_m\sqrt{n}}}\left(s+\frac{p_{max}}{\sqrt{n}}\left(n-\frac{n\pi_m-\beta\sigma_m\sqrt{n}}{\pi_m}\right)\right)\\
&\qquad=\frac{1}{\sqrt{\pi_m-\beta\sigma_m/\sqrt{n}}}\left(s+\frac{p_{max}\beta\sigma_m}{\pi_m}\right)\\
&\qquad\leq\frac{1}{\sqrt{\pi_m-\beta\sigma_m}}\left(s+\frac{p_{max}\beta\sigma_m}{\pi_m}\right)\\
&\qquad<s(1+\delta).
\end{aligned}
\tag{3.0.17}
$$

Hence, for sufficiently large n, we have

$$
\begin{aligned}
&\sum_{j=\lceil n\pi_m-\beta\sigma_m\sqrt{n}\rceil}^{n}\mathbb{P}\left(\frac{LI_j(\hat{p})-\frac{p_{max}}{\pi_m}j}{\sqrt{j}}\leq\sqrt{\frac{n}{j}}\left(s+\frac{p_{max}}{\sqrt{n}}\left(n-\frac{j}{\pi_m}\right)\right)\right)\gamma_{n,m,j}\\
&\qquad+\epsilon_1\\
&\leq\sum_{j=\lceil n\pi_m-\beta\sigma_m\sqrt{n}\rceil}^{n}\mathbb{P}\left(\frac{LI_j(\hat{p})-\frac{p_{max}}{\pi_m}j}{\sqrt{j}}\leq s(1+\delta)\right)\gamma_{n,m,j}+\epsilon_1.
\end{aligned}
\tag{3.0.18}
$$

Now from the condition (iii), and from the weak convergence, as $j\to\infty$, of $(LI_j(\hat{p})-(p_{max}/\pi_m)j)/\sqrt{j}$ to $R(p_{max}/\pi_m,k)$, we find that, for j large enough,

$$
\begin{aligned}
&\left|\mathbb{P}\left(\frac{LI_j(\hat{p})-\frac{p_{max}}{\pi_m}j}{\sqrt{j}}\leq(1+\delta)s\right)-\mathbb{P}(R(p_{max},k)\leq(1+\delta)s)\right|\\
&\leq\left|\mathbb{P}\left(\frac{LI_j(\hat{p})-\frac{p_{max}}{\pi_m}j}{\sqrt{j}}\leq(1+\delta)s\right)-\mathbb{P}\left(R\left(\frac{p_{max}}{\pi_m},k\right)\leq(1+\delta)s\right)\right|\\
&\quad+\left|\mathbb{P}(R(p_{max},k)\leq(1+\delta)s)-\mathbb{P}\left(R\left(\frac{p_{max}}{\pi_m},k\right)\leq(1+\delta)s\right)\right|\\
&<\frac{\epsilon_2}{2}+\frac{\epsilon_2}{2}\\
&<\epsilon_1\mathbb{P}(R(p_{max},k)\leq(1+\delta)s),
\end{aligned}
\tag{3.0.19}
$$

and so,

$$\mathbb{P}\left(\frac{LI_j(\hat{p}) - \frac{p_{max}}{\pi_m} j}{\sqrt{j}} \leq (1+\delta)s\right) \leq (1+\epsilon_1)\mathbb{P}\left(R(p_{max}, k) \leq (1+\delta)s\right). \qquad (3.0.20)$$

Now since $\lceil n\pi_m - \beta\sigma_m\sqrt{n}\rceil \to \infty$, as $n \to \infty$, with the help of (3.0.18) and (3.0.20), (3.0.16) becomes

$$\begin{aligned}
&\mathbb{P}\left(\frac{LI^{red}_{n,m} - p_{max}n}{\sqrt{n}} \leq s\right) \\
&\qquad \leq \sum_{j=\lceil n\pi_m - \beta\sigma_m\sqrt{n}\rceil}^{n} (1+\epsilon_1)\mathbb{P}\left(R(p_{max}, k) \leq (1+\delta)s\right)\gamma_{n,m,j} + \epsilon_1 \\
&\qquad \leq (1+\epsilon_1)\mathbb{P}\left(R(p_{max}, k) \leq (1+\delta)s\right) + \epsilon_1. \qquad (3.0.21)
\end{aligned}$$

From (3.0.10) we know that $LI^{red}_{n,m} \leq LI_n$ a.s., and so

$$\begin{aligned}
\mathbb{P}\left(\frac{LI_n - p_{max}n}{\sqrt{n}} \leq s\right) &\leq \mathbb{P}\left(\frac{LI^{red}_{n,m} - p_{max}n}{\sqrt{n}} \leq s\right) \\
&\leq (1+\epsilon_1)\mathbb{P}\left(R(p_{max}, k) \leq (1+\delta)s\right) + \epsilon_1, \qquad (3.0.22)
\end{aligned}$$

for large enough n. But since ϵ_1 and δ are arbitrary, (3.0.22) and (3.0.15) together show that

$$\mathbb{P}\left(\frac{LI_n - p_{max}n}{\sqrt{n}} \leq s\right) \to \mathbb{P}(R(p_{max}, k) \leq s), \qquad (3.0.23)$$

for all $s > 0$.

The proof for $s < 0$ is similar. Indeed, since necessarily $p_{max} < 1/k$, $R(p_{max}, k)$ describes the limiting distribution of the longest increasing subsequence for a non-uniform alphabet, and so is supported on $\mathbb{R}$. Then, one needs only to examine quantities of the form $\mathbb{P}(R(p_{max}, k) \leq (1-\delta)s)$, instead of $\mathbb{P}(R(p_{max}, k) \leq (1+\delta)s)$, as we have done throughout the proof for $s > 0$. These changes lead to the resulting statement. ■

Remark 3.0.5 *As an alternative to the above proof, one could certainly adopt the finite-alphabet development of the previous sections so as to express LI_n, for countable infinite alphabets, in terms of approximations to functionals of Brownian motion. More precisely,*

$$
\begin{aligned}
LI_n &= \sup_{m\geq 2} \max_{\substack{0\leq k_1\leq\cdots \\ \leq k_{m-1}\leq n}} \left\{ S^1_{k_1} + S^2_{k_2} + \cdots + S^{m-1}_{k_{m-1}} + a^m_n \right\} \\
&= \sup_{m\geq 2} \left\{ \frac{n}{m} - \frac{1}{m}\sum_{r=1}^{m-1} r S^r_n + \max_{\substack{0\leq k_1\leq\cdots \\ \leq k_{m-1}\leq n}} \sum_{r=1}^{m-1} S^r_{k_r} \right\},
\end{aligned}
$$

where a^m_n counts the number of occurrences of the letter α_m among $(X_i)_{1\leq i\leq n}$, and $S^r_k = \sum_{i=1}^{k} Z^r_i$ is the sum of independent random variables defined as in (2.1.3). *After centering and normalizing the S^r_k, as was done to obtain* (2.2.11) *in the non-uniform finite alphabet development, one could then try to apply Donsker's Theorem to obtain a Brownian functional, which we now know to be distributed as $R(p_{max}, k)$. However, due to the countably infinite number of Brownian motions that would result, great care would need to be taken to make such an approach rigorous.*

CHAPTER IV

GROWING IID ALPHABETS

4.1 Introduction

We have thus far focused on the limiting behavior of LI_n for iid alphabets, noting, in particular, that $LI_n \asymp n$. On the other hand, $L\sigma_n \sim 2\sqrt{n}$, as was discussed in the Introduction.

With an eye to linking both types of asymptotics, we introduce the notion of *growing alphabets*. Specifically, we assume that we have an infinite, ordered alphabet $\mathcal{A} = \{\alpha_1 < \alpha_2 < \cdots < \alpha_n \cdots\}$, and that for each $n \geq 1$, we have a finite alphabet $\mathcal{A}_n = \{\alpha_1 < \alpha_2 < \cdots < \alpha_{m_n}\} \subset \mathcal{A}$. Then $(X_n)_{n \geq 1}$ is chosen to be a sequence of independent random variables such that each X_n is uniformly distributed over $\mathcal{A}_n$, where $m_n \to \infty$, as $n \to \infty$. In this setup, each finite alphabet consists of the first m_n letters of $\mathcal{A}$, so that $\mathcal{A}_n \subset \mathcal{A}_{n+1} \subset \mathcal{A}$, for all $n \geq 1$.

4.2 A First-order Lower Bound

Now if $m_n = m$, for all n, then we have again the finite-alphabet case, so that $LI_n \asymp n$. Here, the linear asymptotic behavior essentially results from long stretches of identical values in the subsequence. On the other hand, if we allow m_n to grow so quickly, as $n \to \infty$, that $X_n \leq X_{n+1}$ with some positive probability uniformly bounded below in n, we find that, again, $LI_n \asymp n$. Such results can be made rigorous, for instance, in the exponentially growing case $m_n \sim c^n$, where $c > 1$. At more moderate growth rates of m_n, however, sub-linear behavior is more typical. In particular, for growing alphabets with a polynomial growth rate, we obtain the following theorem, which serves as something of an interpolation result between the finite-alphabet case and

Table 1: Lower bound constants in the $\sqrt{n}$ regime of Theorem 4.2.1

p	d_p	$2\sqrt{c_p}$
0.5	0.90251	1.41421
0.6	0.98865	1.36346
0.7	1.06787	1.31920
0.8	1.14160	1.28000
0.9	1.21085	1.24486
1.0	1.27635	1.21306
1.5	1.56320	1.08866
2.0	1.80503	1.00000
3.0	2.21070	0.87738
4.0	2.55269	0.79370
5.0	2.85400	0.73143
6.0	3.12640	0.68256
7.0	3.37689	0.64277
8.0	3.61005	0.60951
9.0	3.82904	0.58112
10.0	4.03616	0.55651

the random permutation case. In particular, it shows that, in expectation, LI_n must always be *at least* asymptotically $\sqrt{n}$, in this growth regime.

Theorem 4.2.1 *Let $m_n = \lceil n^p \rceil$, with $p > 0$, and let $(X_n)_{n \geq 1}$ be a sequence of independent random variables, with X_n uniformly distributed over $\mathcal{A}_n$. Then,*

$$\liminf_{n\to\infty} \frac{\mathbb{E}LI_n}{n^{1-p}} \geq \frac{1}{1-p}, \qquad 0 < p \leq \frac{1}{2}, \tag{4.2.1}$$

$$\liminf_{n\to\infty} \frac{\mathbb{E}LI_n}{\sqrt{n}} \geq (2\sqrt{c_p}) \vee d_p, \qquad p > \frac{1}{2}, \tag{4.2.2}$$

where c_p and d_p satisfy $c_1 = e^{-1}$ and $c_p = (p^{p/(1-p)} - p^{1/(1-p)})/(1-p)$, for $p \neq 1$, and $d_p = \sup_{\alpha>0}\{\sqrt{2/\alpha}(1 - e^{-2\alpha p})\}$, for all $p > 0$, respectively.

Remark 4.2.1 *The lower bound in (4.2.1) can actually be extended to $0 < p < 1$, and the lower bound in (4.2.2) can likewise be extended to all $p > 0$. However, for*

$0 < p < 1$, *the theorem as stated shows the best asymptotic rates of growth in* n, *and, moreover, the constant* $1/(1-p)$ *can also be shown to be exact for* $0 < p < 1/3$. *Finally, note that the constant* c_p *is continuous at* $p = 1$, *and that* $d_p > 2\sqrt{c_p}$, *if and only if,* p *is greater than approximately* 0.94, *as suggested in Table* 1.

Proof. To prove (4.2.1), observe that, for any $n \geq 1$,

$$LI_n \geq \sum_{k=1}^{n} \mathbf{1}_{\{X_k=\alpha_1\}}, \tag{4.2.3}$$

almost surely.

Let $0 < p < 1$. Note that, for any k, we have $m_j = k$ over the block $I_k := \{\lfloor (k-1)^{1/p} \rfloor + 1 \leq j \leq \lfloor k^{1/p} \rfloor\}$, where $\lfloor a \rfloor$ (the floor of a) is the greatest integer less than or equal to a. Then (4.2.3) clearly implies that

$$\begin{aligned}
\mathbb{E}LI_{\lfloor n^{1/p} \rfloor} &\geq \sum_{j=1}^{\lfloor n^{1/p} \rfloor} \mathbb{E}\mathbf{1}_{\{X_j=\alpha_1\}} \\
&= \sum_{j=1}^{\lfloor n^{1/p} \rfloor} \mathbb{P}(X_j = \alpha_1) \\
&= \sum_{k=1}^{n} \left(\lfloor k^{1/p} \rfloor - \lfloor (k-1)^{1/p} \rfloor\right) \frac{1}{k} \\
&\geq \sum_{k=1}^{n} \left(k^{1/p} - (k-1)^{1/p} - 1\right) \frac{1}{k} \\
&\geq \sum_{k=1}^{n} \left(k^{1/p}\left(1 - e^{-1/kp}\right) - 1\right) \frac{1}{k} \\
&\geq \sum_{k=1}^{n} \left(k^{1/p}\left(\frac{1}{kp}\right) - 1\right) \frac{1}{k} \\
&= \sum_{k=1}^{n} \frac{1}{p} k^{1/p-2} - \frac{1}{k} \\
&= \frac{1}{1-p} n^{1/p-1} + o\left(n^{1/p-1}\right).
\end{aligned} \tag{4.2.4}$$

Hence, $\liminf_{n\to\infty} \mathbb{E}LI_{\lfloor n^{1/p}\rfloor}/n^{1/p-1} \geq 1/(1-p)$, and so, by rescaling,

$$\liminf_{n\to\infty} \frac{\mathbb{E}LI_n}{n^{1-p}} \geq \frac{1}{1-p}$$

,

for $0 < p < 1$, and (4.2.1) is proved.

Turning next to (4.2.2), we first establish the bounds associated with c_p. For k and n fixed, $k \leq n$, let

$$Y_j^k = \mathbf{1}_{\{X_j \leq \alpha_{m_k}\}}, \qquad k \leq j \leq n, \tag{4.2.5}$$

and thus the number $N_{k,n}$ of X_js among $X_k, X_{k+1}, \ldots, X_n$ which do not exceed α_{m_k} is

$$N_{k,n} = \sum_{j=k}^{n} Y_j^k. \tag{4.2.6}$$

In seeking a lower bound for $\mathbb{E}N_{k,n}$, we first note that, trivially, for $k = n-1$ or n, $\mathbb{E}N_{k,n}$ does not exceed 2. For the remaining values of $1 \leq k \leq n-2$, we have, for $p \neq 1$:

$$\begin{aligned}
\mathbb{E}N_{k,n} = \sum_{j=k}^{n} \mathbb{E}Y_j^k &= \sum_{j=k}^{n} \frac{\lceil k^p \rceil}{\lceil j^p \rceil} \\
&\geq k^p \sum_{j=k}^{n} \frac{1}{(j+1)^p} \\
&\geq k^p \int_k^n \frac{1}{(y+2)^p} dy \\
&> k^p \int_k^{n-2} \frac{1}{(y+2)^p} dy \\
&= k^p \frac{n^{1-p} - (k+2)^{1-p}}{1-p} \\
&= \frac{1}{1-p}\left[n\left(\frac{k}{n}\right)^p - \left(\frac{k}{k+2}\right)^p (k+2)\right]
\end{aligned}$$

$$> \frac{1}{1-p}\left[n\left(\frac{k}{n}\right)^p - k - 2\right]. \tag{4.2.7}$$

Let us denote the expression in (4.2.7) by $f(k)$, and optimize over $1 \leq k \leq n-2$. One can easily check that f attains its maximum at $x_n := np^{1/(1-p)}$, and so

$$\begin{aligned} f(x_n) &= \frac{1}{1-p}(np^{p/(1-p)} - np^{1/(1-p)} - 2) \\ &= nc_p - \frac{2}{n}. \end{aligned} \tag{4.2.8}$$

Now among the integers, $k_n = \lceil x_n \rceil$ or $\lfloor x_n \rfloor$ one finds the maximum of f. In either case, it is clear that

$$\liminf_{n\to\infty} \frac{\mathbb{E}N_{k,n}}{n} \geq c_p. \tag{4.2.9}$$

To see that (4.2.9) holds when $p = 1$, we make use of the inequality

$$\frac{1}{2(n+1)} < \sum_{j=1}^{n} \frac{1}{j} - \log n - \gamma < \frac{1}{2n}, \tag{4.2.10}$$

where $\gamma \approx 1.57721\ldots$ is the Euler-Mascheroni constant. From this it follows that, for $2 \leq k \leq n$,

$$\log\left(\frac{n}{k-1}\right) - \frac{n-k+2}{2(k-1)(n+1)} < \sum_{j=k}^{n} \frac{1}{j} < \log\left(\frac{n}{k-1}\right) - \frac{n-k}{2kn}. \tag{4.2.11}$$

Now trivially $\mathbb{E}N_{n,n}$ does not exceed 1, and for $2 \leq k \leq n-1$ we have that

$$\begin{aligned} \mathbb{E}N_{k,n} &= \sum_{j=k}^{n} \mathbb{E}Y_j^k = \sum_{j=k}^{n} \frac{k}{j} \\ &> k\left(\log\left(\frac{n}{k-1}\right) - \frac{n-k+2}{2(k-1)(n+1)}\right) \end{aligned}$$

$$> k \log \left(\frac{n}{k-1} \right) - 1. \tag{4.2.12}$$

Denoting the final expression in (4.2.12) by $g(k)$, we find that its maximum is attained at some x_n, where $x_n/n \to e^{-1}$, as $n \to \infty$. Again denoting the maximum of $g(k)$ over the integers by k_n (either $\lceil x_n \rceil$ or $\lfloor x_n \rfloor$), we again see that

$$\begin{aligned} \liminf_{n\to\infty} \frac{\mathbb{E}N_{k,n}}{n} &\geq \liminf_{n\to\infty} \frac{1}{n} \left(k_n \log \left(\frac{n}{k_n - 1} \right) - 1 \right) \\ &= e^{-1} \log e = c_1. \end{aligned} \tag{4.2.13}$$

We thus have an estimate of how many X_js satisfy $X_j \leq \alpha_{m_{k_n}}$, for $k_n \leq j \leq n$, an estimate whose expectation is of order n. To complete the proof, we make three key observations.

Firstly, since $N_{k,n}$ consists of a sum of (independent) indicator random variables, we have

$$\mathbb{P}(N_{k,n} \geq a\mathbb{E}N_{k,n}) \geq 1 - e^{-(1-a)^2(\mathbb{E}N_{k,n})^2/2\operatorname{Var} N_{k,n}} \geq 1 - e^{-(1-a)^2\mathbb{E}N_{k,n}/2}, \tag{4.2.14}$$

for $0 < a < 1$.

Secondly, denoting by $A_{k_n,n}$ the set of indices among $k_n, k_n + 1, \ldots, n$ for which $X_j \leq \alpha_{m_{k_n}}$, we see that each $X_j \in A_{k_n,n}$ is distributed uniformly over $\alpha_1, \alpha_2, \ldots, \alpha_{m_{k_n}}$. Hence, the length of the longest increasing subsequence among such X_j, which we denote by $LI(A_{k_n,n})$, is distributed as in the m_{k_n}-letter uniform alphabet case, with the sequence length given by $N_{k_n,n}$.

Thirdly, given $(Y_i)_{1\leq i\leq n}$ iid, chosen uniformly over the m-letter alphabet $\{\alpha_k\}_{1\leq k\leq m}$, and $(U_i)_{1\leq i\leq n}$ iid, and independent of $(Y_i)_{1\leq i\leq n}$, with $U_1 \sim U(0,1)$, then the sequence $(Z_i)_{1\leq i\leq n}$ defined by $Z_i = Y_i - U_i$ has two useful properties. The first of these is that clearly each Z_i is uniform over the interval $(0, m)$, and hence the length of the longest increasing subsequence of $(Z_i)_{1\le i\le n}$ is distributed as $L\sigma_n$. The second property is that if $Z_{i_1} \le Z_{i_2} \le \cdots \le Z_{i_k}$ is an increasing subsequence of $(Z_i)_{1\le i\le n}$, then $Y_{i_1} \le Y_{i_2} \le \cdots \le Y_{i_k}$ must also be an increasing subsequence of $(Y_i)_{1\le i\le n}$, and so $LI(Y_1,\ldots,Y_n) \ge LI(Z_1,\ldots,Z_n)$.

Applying these insights, along with our understanding of the asymptotics of $N_{k_n,n}$, we conclude that, for any $0 \le a < 1$,

$$\begin{aligned}
\mathbb{E}LI_n &\ge \mathbb{E}LI(A_{k_n,n}) \\
&\ge \sum_{j=\lfloor a\mathbb{E}N_{k_n,n}\rfloor}^{n} \mathbb{E}(LI(A_{k_n,n})|N_{k_n,n}=j)\mathbb{P}(N_{k_n,n}=j) \\
&\ge \sum_{j=\lfloor a\mathbb{E}N_{k_n,n}\rfloor}^{n} \mathbb{E}L\sigma_j\mathbb{P}(N_{k_n,n}=j) \\
&> \mathbb{E}L\sigma_{\lfloor a\mathbb{E}N_{k_n,n}\rfloor} \sum_{j=\lfloor a\mathbb{E}N_{k_n,n}\rfloor}^{n} \mathbb{P}(N_{k_n,n}=j) \\
&\ge \mathbb{E}L\sigma_{\lfloor a\mathbb{E}N_{k_n,n}\rfloor}\mathbb{P}(N_{k,n} \ge a\mathbb{E}N_{k_n,n}) \\
&> \mathbb{E}L\sigma_{\lfloor a\mathbb{E}N_{k_n,n}\rfloor}\left(1-e^{-(1-a)^2\mathbb{E}N_{k_n,n}/2}\right),
\end{aligned} \tag{4.2.15}$$

using (4.2.14). Hence, since $\liminf_{n\to\infty}\mathbb{E}N_{k_n,n}/n = c_p$,

$$\begin{aligned}
\liminf_{n\to\infty}\frac{LI_n}{\sqrt{n}} &\ge \liminf_{n\to\infty}\frac{\mathbb{E}L\sigma_{\lfloor a\mathbb{E}N_{k_n,n}\rfloor}}{\sqrt{n}}\left(1-e^{-(1-a)^2\mathbb{E}N_{k_n,n}/2}\right) \\
&\ge 2\sqrt{ac_p}(1),
\end{aligned} \tag{4.2.16}$$

which is optimized as $a \nearrow 1$. Thus,

$$\liminf_{n\to\infty}\frac{LI_n}{\sqrt{n}} \ge 2\sqrt{c_p}, \tag{4.2.17}$$

and we have the bound associated with c_p.

Turning to the lower bound associated with d_p, let $\alpha > 0$ be given, and define the k^{th} block of indices to be $I_k = \{\beta_{k-1}+1, \ldots, \beta_k\}$, where $\beta_k = \lfloor \alpha k(k+1)/2 \rfloor$. We will be interested in examining the events $E_k = \cup_{j \in I_k}\{X_j \in R_k\}$, where $R_k := \{\lceil \beta_{k-2}^p + 1 \rceil, \ldots, \lceil \beta_{k-1}^p \rceil\}$, a set whose cardinality $\Delta_k = \lceil \beta_{k-1}^p \rceil - \lceil \beta_{k-2}^p \rceil$ we will estimate in terms of α and p. Note that $R_k \subset \mathcal{A}_j$, for all $j \in I_k$.

It is not hard to see that, in analogy with (4.2.3),

$$LI_{\beta_k} \geq \sum_{j=1}^{k} \mathbf{1}_{E_k}. \tag{4.2.18}$$

Now by the independence of the X_j, we have

$$\begin{aligned}
\mathbb{P}(E_k^c) &= \prod_{j \in I_k} \mathbb{P}(X^j \notin R_k) \\
&= \prod_{j=1}^{\beta_k - \beta_{k-1}} \left(1 - \frac{\Delta_k}{\lceil \beta_{k-1} + j \rceil^p}\right) \\
&\leq \exp\left(-\Delta_k \sum_{j=1}^{\beta_k - \beta_{k-1}} \frac{1}{\lceil \beta_{k-1} + j \rceil^p}\right).
\end{aligned} \tag{4.2.19}$$

We first estimate Δ_k in (4.2.19) as follows:

$$\begin{aligned}
\Delta_k &= \lceil \beta_{k-1}^p \rceil - \lceil \beta_{k-2}^p \rceil \\
&\geq \beta_{k-1}^p - \beta_{k-2}^p - 1 \\
&= \lfloor \frac{\alpha k(k-1)}{2} \rfloor^p - \lfloor \frac{\alpha(k-1)(k-2)}{2} \rfloor^p - 1 \\
&\geq \left(\frac{\alpha k(k-1)}{2}\right)^p - \left(\frac{\alpha(k-1)(k-2)}{2}\right)^p - 2 \\
&= \left(\frac{\alpha k(k-1)}{2}\right)^p \left(1 - \left(1 - \frac{2}{k}\right)^p\right) - 2.
\end{aligned} \tag{4.2.20}$$

Next, let $0 < \varepsilon < 1$. For $p \geq 1$, we note that $1 - (1 - 2/k)^p \geq p(2/k)(1 - 2/k)^{p-1}$. Then, for k large enough, our estimate in (4.2.20) becomes

$$
\begin{aligned}
\Delta_k &\geq \left(\frac{\alpha k(k-1)}{2}\right)^p \left(p\left(\frac{2}{k}\right)\left(1-\frac{2}{k}\right)^{p-1}\right) - 2 \\
&\geq (1-\varepsilon)\frac{\alpha^p k^{p-1}(k-1)^p}{2^{p-1}} p. \qquad (4.2.21)
\end{aligned}
$$

Similarly, for $0 < p < 1$, we also note that $1-(1-2/k)^p \geq p(2/k)$. Then, for k large enough, our estimate in (4.2.20) again becomes

$$
\begin{aligned}
\Delta_k &\geq \left(\frac{\alpha k(k-1)}{2}\right)^p \left(p\left(\frac{2}{k}\right)\right) - 2 \\
&\geq (1-\varepsilon)\frac{\alpha^p k^{p-1}(k-1)^p}{2^{p-1}} p. \qquad (4.2.22)
\end{aligned}
$$

To estimate the sum in (4.2.19), we also have, for sufficiently large k,

$$
\begin{aligned}
\sum_{j=1}^{\beta_k - \beta_{k-1}} \frac{1}{\lceil \beta_{k-1} + j \rceil^p} &\geq \sum_{j=2}^{\beta_k - \beta_{k-1}+1} \frac{1}{(\beta_{k-1}+j)^p} \\
&\geq \int_2^{\beta_k - \beta_{k-1}+2} \frac{1}{(\beta_{k-1}+x)^p} dx. \qquad (4.2.23)
\end{aligned}
$$

To continue estimating (4.2.23), we first consider the case $p > 1$. Then

$$
\begin{aligned}
&\int_2^{\beta_k - \beta_{k-1}+2} \frac{1}{(\beta_{k-1}+x)^p} dx \\
&\quad = \frac{1}{p-1}\left(\frac{1}{(\beta_{k-1}+2)^{p-1}} - \frac{1}{(\beta_k+2)^{p-1}}\right) \\
&\quad = \frac{1}{(p-1)(\beta_k+2)^{p-1}}\left(\frac{(\beta_k+2)^{p-1}}{(\beta_{k-1}+2)^{p-1}} - 1\right) \\
&\quad \geq \frac{1}{(p-1)(\alpha k(k+1)/2+2)^{p-1}}\left(\left(\frac{\alpha k(k+1)/2+1}{\alpha k(k-1)/2+2}\right)^{p-1} - 1\right) \\
&\quad = \frac{2^{p-1}}{(p-1)(\alpha k(k+1)+4)^{p-1}}\left(\left(1+\frac{2(\alpha k-1)}{\alpha k(k-1)+4}\right)^{p-1} - 1\right)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \frac{2^{p-1}}{(p-1)(\alpha k(k+1)+4)^{p-1}} \left(\frac{2(p-1)(\alpha k - 1)}{\alpha k(k-1)+4} \right) \\
&\geq (1-\varepsilon) \frac{2^p(\alpha k - 1)}{(\alpha k(k+1))^p}. \qquad (4.2.24)
\end{aligned}
$$

Similarly, for $0 < p < 1$, we again find that

$$
\begin{aligned}
&\int_2^{\beta_k - \beta_{k-1}+2} \frac{1}{(\beta_{k-1}+x)^p} dx \\
&\quad = \frac{1}{1-p} \left((\beta_k + 2)^{1-p} - (\beta_{k-1}+2)^{1-p} \right) \\
&\quad = \frac{1}{1-p} (\beta_k + 2)^{1-p} \left(1 - \left(\frac{(\beta_{k-1}+2)}{\beta_k + 2} \right)^{1-p} \right) \\
&\quad \geq \frac{1}{1-p} \left(\frac{\alpha k(k+1)}{2} + 1 \right)^{1-p} \left(1 - \left(\frac{\alpha k(k-1)/2+1}{\alpha k(k+1)/2+2} \right)^{1-p} \right) \\
&\quad = \frac{(\alpha k(k+1)+2)^{1-p}}{2^{1-p}(1-p)} \left(1 - \left(1 - \frac{2(\alpha k+1)}{\alpha k(k+1)+4} \right)^{1-p} \right) \\
&\quad \geq \frac{(\alpha k(k+1)+2)^{1-p}}{2^{1-p}(1-p)} \left(1 - \frac{2(1-p)(\alpha k+1)}{\alpha k(k+1)+4} \right) \\
&\quad = \frac{2^p(\alpha k+1)(\alpha k(k+1)+2)^{1-p}}{\alpha k(k+1)+4} \\
&\quad \geq (1-\varepsilon) \frac{2^p(\alpha k+1)}{(\alpha k(k+1))^p} \\
&\quad > (1-\varepsilon) \frac{2^p(\alpha k-1)}{(\alpha k(k+1))^p}. \qquad (4.2.25)
\end{aligned}
$$

Finally, for $p = 1$,

$$
\begin{aligned}
&\int_2^{\beta_k - \beta_{k-1}+2} \frac{1}{(\beta_{k-1}+x)^p} dx \\
&\quad = \log(\beta_k + 2) - \log(\beta_{k-1}+2) \\
&\quad = \log \frac{\beta_k + 2}{\beta_{k-1}+2} \\
&\quad \geq \log \frac{\alpha k(k+1)/2+1}{\alpha k(k-1)/2+2}
\end{aligned}
$$

$$= \log \frac{\alpha k(k+1)+2}{\alpha k(k-1)+4}$$
$$= \log\left((1+\frac{2(\alpha k-1)}{\alpha k(k-1)+4}\right)$$
$$\geq (1-\varepsilon)\frac{2(\alpha k-1)}{\alpha k(k+1)}. \tag{4.2.26}$$

Thus, for any $p > 0$, and for k large enough, say $k > K$, we have from our estimates in (4.2.21) - (4.2.26) that

$$\begin{aligned}
\mathbb{P}(E_k^c) &\leq \exp\left(-\Delta_k \sum_{j=1}^{\beta_k-\beta_{k-1}} \frac{1}{\lceil \beta_{k-1}+j\rceil^p}\right) \\
&\leq \exp\left(-\left[(1-\varepsilon)\frac{\alpha^p k^{p-1}(k-1)^p}{2^{p-1}}p\right]\left[(1-\varepsilon)\frac{2^p(\alpha k-1)}{(\alpha k(k+1))^p}\right]\right) \\
&= \exp\left(-(1-\varepsilon)^2\frac{2p(\alpha k-1)}{k}\right) \\
&\leq \exp(-(1-\varepsilon)^3(2\alpha p)).
\end{aligned} \tag{4.2.27}$$

Applying the estimate in (4.2.27) to (4.2.18), we see that for $n > K$,

$$\begin{aligned}
\mathbb{E}LI_{\beta_n} &\geq \sum_{k=1}^{n} \mathbb{E}\mathbf{1}_{E_k} \\
&\geq \sum_{k=K+1}^{n} \mathbb{E}\mathbf{1}_{E_k} \\
&\geq \sum_{k=K+1}^{n} \left(1-e^{-(1-\varepsilon)^3(2\alpha p))}\right) \\
&= (n-K)\left(1-e^{-(1-\varepsilon)^3(2\alpha p)}\right).
\end{aligned} \tag{4.2.28}$$

Since $n/\sqrt{\beta_n} \to \sqrt{2/\alpha}$, as $n \to \infty$, we have

$$\liminf_{n\to\infty} \frac{LI_{\beta_n}}{\sqrt{\beta_n}} = \liminf_{n\to\infty} \frac{LI_n}{\sqrt{n}} \geq \sqrt{\frac{2}{\alpha}}\left(1-e^{-(1-\varepsilon)^3(2\alpha p)}\right), \tag{4.2.29}$$

and since ε was arbitrary,

$$\liminf_{n\to\infty} \frac{LI_n}{\sqrt{n}} \geq \sqrt{\frac{2}{\alpha}}\left(1 - e^{-2\alpha p}\right), \tag{4.2.30}$$

and we may optimize over $\alpha > 0$, and so complete the proof. ∎

CHAPTER V

MARKOVIAN ALPHABETS

Recall that in the combinatorial development above, the expression for LI_n in (2.1.5), namely,

$$LI_n = \frac{n}{m} - \frac{1}{m}\sum_{r=1}^{m-1} rS_n^r + \max_{\substack{0\le k_1\le\cdots\\ \le k_{m-1}\le n}} \{S_{k_1}^1 + S_{k_2}^2 + \cdots + S_{k_{m-1}}^{m-1}\}, \tag{5.0.31}$$

is of a *purely combinatorial nature or, in more probabilistic terms, is of a pathwise nature.* We wish to extend our analysis of this expression to Markovian sequences.

Moreover, at the same time we wish to generalize from LI_n to the shape of the entire associated Young tableau, which we now define and relate to the sequence $(X_n)_{n\ge1}$.

A *Young tableau of size* n is a diagram consisting of a collection of n boxes arranged in rows and aligned at the left, such that:

- The number of boxes in each row is no greater than the number of boxes in the row above, and
- Each box contains entries which are weakly or strictly increasing in each row and strictly increasing down each column. If the entries are row-wise *weakly* increasing, we say that the Young tableau is *semi-standard*, while if the entries are row-wise *strictly* increasing, we say that it is *standard*.

The *shape* of a Young tableau will refer to the lengths of the rows, irrespective of the entries, and it is the shape that will be of primary concern to us.

Young tableaux are connected to sequence analysis via the well-known Robinson-Schensted-Knuth (RSK) correspondence, which states that for any sequence $(X_k)_{1\le k\le n}$

1	1	1	2	3	4	4	4	4	5	5	5	6
2	2	2	4	5	5	5	5	5	6	7		
3	4	4	6	6	6	6	6	7	7	8		
5	6	6	7	7	8							

Figure 1: Example of a 4-Row, Semi-Standard Young Tableau

drawn from an ordered alphabet there is a unique *pair* of Young tableaux (P, Q), both of the same shape, with P semi-standard, and Q standard. The pair (P, Q) is constructed as follows.

Beginning with a Young tableau P consisting of a single box containing X_1, and a corresponding Young tableau Q, called the *recording tableau*, also consisting of a single box but containing the integer 1, we successively augment P and Q according to the values of $X_2, \ldots, X_n$ using the following algorithm. For each $k \geq 2$:

- If X_k is *greater than or equal to* the final entry of the first row of P, then we simply add another box containing X_k to the end of the first row, completing the augmentation of P by X_k.

- If X_k is *strictly less than* the final entry of the first row of P, then we locate the left-most box of the first row whose entry exceeds X_k, replace that entry with X_k, and "bump" the original entry to the next row.

- For any "bumped" entry, we proceed, in each successive row, as with the first row until an entry is added to the end of a (possibly empty) row, at which point the augmentation of P by X_k is complete.

- Once P has been augmented, we augment Q with a box containing the integer k, where the location of the box corresponds the location of the box in P that was added to the end of a row. (This explains the name *recording tableau*).

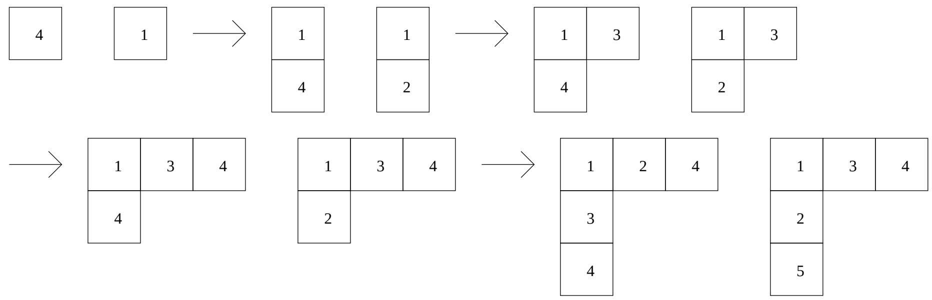

Figure 2: RSK Algorithm Applied to the Sequence $(4, 1, 3, 4, 2)$.

(See Figure 2 for a short example of the RSK algorithm applied to a sequence of length 5.)

Moreover, one can always recover the sequence $(X_k)_{1 \leq k \leq n}$ from (P, Q). Indeed, the RSK correspondence states that there is actually a one-to-one correspondence between all possible sequences $(X_k)_{1 \leq k \leq n}$ of letters from an alphabet of size m, and all possible pairs (P, Q) of Young tableaux, with P a semi-standard Young tableau with entries in the alphabet of size m, and Q a standard Young tableau with entries consisting of the first n positive integers.

5.1 2-*Letter Case*

We begin our study of Markovian alphabets by first concentrating on the 2-letter case. Now $R_n^1 = LI_n$, and with $m = 2$, $R_n^2 = n - LI_n$, it suffices to describe LI_n. Here $(X_n)_{n \geq 0}$ is described by the following transition probabilities between the two states (which we identify with the two letters α_1 and α_2): $\mathbb{P}(X_{n+1} = \alpha_2 | X_n = \alpha_1) = a$ and $\mathbb{P}(X_{n+1} = \alpha_1 | X_n = \alpha_2) = b$, where $0 < a + b < 2$. We later examine the degenerate cases $a = b = 0$ and $a = b = 1$. In keeping with the common usage within the Markov chain literature, we begin our sequence at $n = 0$, although our focus will be on $n \geq 1$. Denoting by (p_n^1, p_n^2) the vector describing the probability distribution on $\{\alpha_1, \alpha_2\}$ at time n, we have

$$\left(p_{n+1}^1, p_{n+1}^2\right) = \left(p_n^1, p_n^2\right) \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}. \tag{5.1.1}$$

The eigenvalues of the matrix in (5.1.1) are $\lambda_1 = 1$ and $-1 < \lambda_2 = 1 - a - b < 1$, with respective left eigenvectors $(\pi_1, \pi_2) = (b/(a+b), a/(a+b))$ and $(1, -1)$. Moreover, (π_1, π_2) is also the stationary distribution. Given any initial distribution (p_0^1, p_0^2), we find that

$$\left(p_n^1, p_n^2\right) = \left(\pi_1, \pi_2\right) + \lambda_2^n \frac{ap_0^1 - bp_0^2}{a+b}\left(1, -1\right) \to \left(\pi_1, \pi_2\right), \tag{5.1.2}$$

as $n \to \infty$, since $\lambda_2 < 1$.

Our goal is now to use these probabilistic expressions to describe the random variables Z_k^1 and S_k^1 defined in Section 2.2. (We retain the redundant superscript "1" in Z_k^1 and S_k^1 in the interest of uniformity.)

Setting $\beta = ap_0^1 - bp_0^2$, we easily find that

$$\begin{aligned} \mathbb{E}Z_k^1 &= (+1)\left(\pi_1 + \frac{\beta}{a+b}\lambda_2^k\right) + (-1)\left(\pi_2 - \frac{\beta}{a+b}\lambda_2^k\right) \\ &= \frac{b-a}{a+b} + 2\frac{\beta}{a+b}\lambda_2^k, \end{aligned} \tag{5.1.3}$$

for each $1 \le k \le n$. Thus,

$$\mathbb{E}S_k^1 = \frac{b-a}{a+b}k + 2\left(\frac{\beta\lambda_2}{a+b}\right)\left(\frac{1-\lambda_2^k}{1-\lambda_2}\right), \tag{5.1.4}$$

and so $\mathbb{E}S_k^1/k \to (b-a)/(a+b)$, as $k \to \infty$.

Turning to the second moments of Z_k^1 and S_k^1, first note that $\mathbb{E}(Z_k^1)^2 = 1$, since $(Z_k^1)^2 = 1$ a.s. Next, we consider $\mathbb{E}Z_k^1 Z_\ell^1$, for $k < \ell$. Using the Markovian structure of $(X_n)_{n\ge0}$, it quickly follows that

$$\mathbb{P}((X_k, X_\ell) = (x_k, x_\ell))$$
$$= \begin{cases} \left(\pi_1 + \lambda_2^{\ell-k}\frac{a}{a+b}\right)\left(\pi_1 + \lambda_2^k \frac{\beta}{a+b}\right), & \text{if } (x_k, x_\ell) = (\alpha_1, \alpha_1), \\ \left(\pi_1 - \lambda_2^{\ell-k}\frac{b}{a+b}\right)\left(\pi_2 - \lambda_2^k \frac{\beta}{a+b}\right), & \text{if } (x_k, x_\ell) = (\alpha_1, \alpha_2), \\ \left(\pi_2 - \lambda_2^{\ell-k}\frac{a}{a+b}\right)\left(\pi_1 + \lambda_2^k \frac{\beta}{a+b}\right), & \text{if } (x_k, x_\ell) = (\alpha_2, \alpha_1), \\ \left(\pi_2 + \lambda_2^{\ell-k}\frac{b}{a+b}\right)\left(\pi_2 - \lambda_2^k \frac{\beta}{a+b}\right), & \text{if } (x_k, x_\ell) = (\alpha_2, \alpha_2). \end{cases} \tag{5.1.5}$$

For simplicity, we will henceforth assume that our initial distribution is the stationary one, *i.e.*, $(p_0^1, p_0^2) = (\pi_1, \pi_2)$. Later, (see Chapter VI) we note that we may drop this assumption and deal with initial distributions concentrated on an arbitrary state. Under this assumption, $\beta = 0$, $\mathbb{E}S_k^1 = k\mu$, where $\mu = \mathbb{E}Z_k^1 = (b-a)/(a+b)$, and (5.1.5) simplifies to

$$\mathbb{P}((X_k, X_\ell) = (x_k, x_\ell))$$
$$= \begin{cases} \left(\pi_1 + \lambda_2^{\ell-k}\frac{a}{a+b}\right)\pi_1, & \text{if } (x_k, x_\ell) = (\alpha_1, \alpha_1), \\ \left(\pi_1 - \lambda_2^{\ell-k}\frac{b}{a+b}\right)\pi_2, & \text{if } (x_k, x_\ell) = (\alpha_1, \alpha_2), \\ \left(\pi_2 - \lambda_2^{\ell-k}\frac{a}{a+b}\right)\pi_1, & \text{if } (x_k, x_\ell) = (\alpha_2, \alpha_1), \\ \left(\pi_2 + \lambda_2^{\ell-k}\frac{b}{a+b}\right)\pi_2, & \text{if } (x_k, x_\ell) = (\alpha_2, \alpha_2). \end{cases} \tag{5.1.6}$$

We can now compute $\mathbb{E}Z_k^1 Z_\ell^1$:

$$\begin{aligned} \mathbb{E}Z_k^1 Z_\ell^1 &= \mathbb{P}(Z_k^1 Z_\ell^1 = +1) - \mathbb{P}(Z_k^1 Z_\ell^1 = -1) \\ &= \mathbb{P}((X_k, X_\ell) \in \{(\alpha_1, \alpha_1), (\alpha_2, \alpha_2)\}) \\ &\qquad - \mathbb{P}((X_k, X_\ell) \in \{(\alpha_1, \alpha_2), (\alpha_2, \alpha_1)\}) \\ &= \left(\pi_1^2 + \lambda_2^{\ell-k}\frac{a}{a+b}\pi_1 + \pi_2^2 + \lambda_2^{\ell-k}\frac{b}{a+b}\pi_2\right) \end{aligned}$$

$$
\begin{aligned}
&- \left(\pi_1\pi_2 - \lambda_2^{\ell-k}\frac{b}{a+b}\pi_2 + \pi_1\pi_2 - \lambda_2^{\ell-k}\frac{a}{a+b}\pi_1 \right) \\
&= \left(\pi_1^2 + \pi_2^2 + \frac{2ab}{(a+b)^2}\lambda_2^{\ell-k} \right) - \left(2\pi_1\pi_2 - \frac{2ab}{(a+b)^2}\lambda_2^{\ell-k} \right) \\
&= \frac{(b-a)^2}{(a+b)^2} + \frac{4ab}{(a+b)^2}\lambda_2^{\ell-k}. \qquad (5.1.7)
\end{aligned}
$$

Hence, recalling that $\beta = 0$,

$$
\begin{aligned}
\sigma^2 := \operatorname{Var} Z_k^1 &= 1 - \left(\frac{b-a}{a+b} \right)^2 \\
&= \frac{4ab}{(a+b)^2}, \qquad (5.1.8)
\end{aligned}
$$

for all $k \geq 1$, and, for $k < \ell$, the covariance of Z_k^1 and Z_ℓ^1 is

$$
\operatorname{Cov}(Z_k^1, Z_\ell^1) = \frac{(b-a)^2}{(a+b)^2} + \sigma^2\lambda_2^{\ell-k} - \left(\frac{b-a}{a+b} \right)^2 = \sigma^2\lambda_2^{\ell-k}. \qquad (5.1.9)
$$

Proceeding to the covariance structure of S_k^1, we first find that

$$
\begin{aligned}
\operatorname{Var} S_k^1 &= \sum_{j=1}^{k} \operatorname{Var} Z_j^1 + 2\sum_{j<\ell} \operatorname{Cov}(Z_j^1, Z_l^1) \\
&= \sigma^2 k + 2\sigma^2 \sum_{j<\ell} \lambda_2^{\ell-j} \\
&= \sigma^2 k + 2\sigma^2 \left(\frac{\lambda_2^{k+1} - k\lambda_2^2 + (k-1)\lambda_2}{(1-\lambda_2)^2} \right) \\
&= \sigma^2 \left(\frac{1+\lambda_2}{1-\lambda_2} \right) k + 2\sigma^2 \left(\frac{\lambda_2(\lambda_2^k - 1)}{(1-\lambda_2)^2} \right). \qquad (5.1.10)
\end{aligned}
$$

Next, for $k < \ell$, and using (5.1.9) and (5.1.10), the covariance of S_k^1 and S_ℓ^1 is given by

$$
\operatorname{Cov}(S_k^1, S_\ell^1) = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \operatorname{Cov}(Z_i^1, Z_j^1)
$$

$$
\begin{aligned}
&= \sum_{i=1}^{k} \operatorname{Var} Z_i^1 + 2 \sum_{i<j<k} \operatorname{Cov}(Z_i^1, Z_j^1) + \sum_{i=1}^{k} \sum_{j=k+1}^{\ell} \operatorname{Cov}(Z_i^1, Z_j^1) \\
&= \operatorname{Var} S_k^1 + \sum_{i=1}^{k} \sum_{j=k+1}^{\ell} \operatorname{Cov}(Z_i^1, Z_j^1) \\
&= \operatorname{Var} S_k^1 + \sigma^2 \left(\frac{\lambda_2 (1-\lambda_2^k)(1-\lambda_2^{\ell-k})}{(1-\lambda_2)^2} \right) \\
&= \sigma^2 \left(\left(\frac{1+\lambda_2}{1-\lambda_2} \right) k - \frac{\lambda_2(1-\lambda_2^k)(1+\lambda_2^{\ell-k})}{(1-\lambda_2)^2} \right). \qquad (5.1.11)
\end{aligned}
$$

From (5.1.10) and (5.1.11) we see that, as $k \to \infty$,

$$
\frac{\operatorname{Var} S_k^1}{k} \to \sigma^2 \left(\frac{1+\lambda_2}{1-\lambda_2} \right), \qquad (5.1.12)
$$

and, moreover, as $k \wedge \ell \to \infty$,

$$
\frac{\operatorname{Cov}(S_k^1, S_\ell^1)}{(k \wedge \ell)} \to \sigma^2 \left(\frac{1+\lambda_2}{1-\lambda_2} \right). \qquad (5.1.13)
$$

When $a = b$, $\mathbb{E} S_k^1 = 0$, and in (5.1.12) the asymptotic variance becomes

$$
\begin{aligned}
\frac{\operatorname{Var} S_k^1}{k} &\to \frac{4a^2}{(2a)^2} \left(\frac{1+(1-2a)}{1-(1-2a)} \right) \\
&= \frac{1}{a} - 1.
\end{aligned}
$$

For a small, we have a "lazy" Markov chain, that is, a Markov chain which tends to remain in a given state for long periods of time. In this regime, the random variable S_k^1 has long periods of increase followed by long periods of decrease. In this way, linear asymptotics of the variance with large constants occur. If, on the other hand, a is close to 1, the Markov chain rapidly shifts back and forth between α_1 and α_2, and so the constant associated with the linearly increasing variance of S_k^1 is small.

As in Chapter II, Brownian functionals play a central rôle in describing the limiting distribution of LI_n.

To move towards a Brownian functional expression for the limiting law of LI_n, define the polygonal function

$$\hat{B}_n(t) = \frac{S^1_{[nt]} - [nt]\mu}{\sigma\sqrt{n(1+\lambda_2)/(1-\lambda_2)}} + \frac{(nt-[nt])(Z^1_{[nt]+1} - \mu)}{\sigma\sqrt{n(1+\lambda_2)/(1-\lambda_2)}}, \tag{5.1.14}$$

for $0 \le t \le 1$. In our finite-state, irreducible, aperiodic, stationary Markov chain setting, we may conclude that $\hat{B}_n \Rightarrow B$, as desired. (See, for example, even more general settings, such as Gordin's martingale approach to dependent invariance principles [21], and the stationary ergodic invariance principle found in Theorem 19.1 of Billingsley [7].)

Turning now to LI_n, we see that for the present 2-letter situation, (5.0.31) simply becomes

$$LI_n = \frac{n}{2} - \frac{1}{2}S^1_n + \max_{1\le k\le n} S^1_k.$$

To find the limiting distribution of LI_n from this expression, recall that $\pi_1 = b/(a+b)$, $\pi_2 = a/(a+b)$, $\mu = \pi_1 - \pi_2 = (b-a)/(a+b)$, $\sigma^2 = 4ab/(a+b)^2$, and that $\lambda_2 = 1-a-b$. Define $\pi_{max} = \max\{\pi_1, \pi_2\}$ and $\tilde{\sigma}^2 = \sigma^2(1+\lambda_2)/(1-\lambda_2)$. Rewriting (5.1.14) as

$$\hat{B}_n(t) = \frac{S^1_{[nt]} - [nt]\mu}{\tilde{\sigma}\sqrt{n}} + \frac{(nt-[nt])(Z^1_{[nt]+1} - \mu)}{\tilde{\sigma}\sqrt{n}},$$

LI_n becomes

$$\begin{aligned} LI_n &= \frac{n}{2} - \frac{1}{2}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(1) + \mu n\right) + \max_{0\le t\le 1}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(t) + \mu n t\right) \\ &= n\pi_2 - \frac{1}{2}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(1)\right) + \max_{0\le t\le 1}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(t) + (\pi_1-\pi_2)nt\right) \\ &= n\pi_{max} - \frac{1}{2}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(1)\right) \\ &\qquad + \max_{0\le t\le 1}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(t) + (\pi_1-\pi_2)nt - (\pi_{max}-\pi_2)n\right). \end{aligned} \tag{5.1.15}$$

This immediately gives

$$\begin{aligned}\frac{LI_n - \pi_{max} n}{\tilde{\sigma}\sqrt{n}} &= -\frac{1}{2}\hat{B}_n(1) \\ &\quad + \max_{0 \le t \le 1}\left(\hat{B}_n(t) + \frac{\sqrt{n}}{\tilde{\sigma}}((\pi_1 - \pi_2)t - (\pi_{max} - \pi_2))\right). \end{aligned} \tag{5.1.16}$$

Let us examine (5.1.16) on a case-by-case basis. First, if $\pi_{max} = \pi_1 = \pi_2 = 1/2$, *i.e.*, if $a = b$, then $\sigma = 1$ and $\tilde{\sigma} = (1-a)/a$, and so (5.1.16) becomes

$$\frac{LI_n - n/2}{\sqrt{(1-a)n/a}} = -\frac{1}{2}\hat{B}_n(1) + \max_{0 \le t \le 1}\hat{B}_n(t). \tag{5.1.17}$$

Then, by the Invariance Principle and the Continuous Mapping Theorem,

$$\frac{LI_n - n/2}{\sqrt{(1-a)n/a}} \Rightarrow -\frac{1}{2}B(1) + \max_{0 \le t \le 1} B(t). \tag{5.1.18}$$

Next, if $\pi_{max} = \pi_2 > \pi_1$, (5.1.16) becomes

$$\begin{aligned}\frac{LI_n - \pi_{max} n}{\tilde{\sigma}\sqrt{n}} &= -\frac{1}{2}\hat{B}_n(1) \\ &\quad + \max_{0 \le t \le 1}\left(\hat{B}_n(t) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_1)t\right). \end{aligned} \tag{5.1.19}$$

On the other hand, if $\pi_{max} = \pi_1 > \pi_2$, (5.1.16) becomes

$$\begin{aligned}\frac{LI_n - \pi_{max} n}{\tilde{\sigma}\sqrt{n}} &= -\frac{1}{2}\hat{B}_n(1) \\ &\quad + \max_{0 \le t \le 1}\left(\hat{B}_n(t) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2)(1-t)\right) \\ &= \frac{1}{2}\hat{B}_n(1) \\ &\quad + \max_{0 \le t \le 1}\left(\hat{B}_n(t) - \hat{B}_n(1) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2)(1-t)\right). \end{aligned} \tag{5.1.20}$$

In both (5.1.19) and (5.1.20) we have a term in our maximal functional which is linear in t or $1-t$, with a negative slope. We now show, in an elementary fashion, that in both cases, as $n \to \infty$, the maximal functional goes to zero in probability.

Consider first (5.1.19). Let $c_n = \sqrt{n}(\pi_{max} - \pi_1)/\tilde{\sigma} > 0$, and for any $c > 0$, let $M_c = \max_{0\le t\le 1}(B(t) - ct)$, where $B(t)$ is a standard Brownian motion. Now for n large enough,

$$\hat{B}_n(t) - ct \geq \hat{B}_n(t) - c_n t$$

a.s., for all $0 \le t \le 1$. Then for any $z > 0$, and n large enough,

$$\mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - c_n t) > z) \leq \mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - ct) > z), \tag{5.1.21}$$

and so by the Invariance Principle and the Continuous Mapping Theorem,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - c_n t) > z) &\leq \lim_{n\to\infty} \mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - ct) > z) \\ &= \mathbb{P}(M_c > z). \end{aligned} \tag{5.1.22}$$

Now, as is well-known, $\mathbb{P}(M_c > z) \to 0$ as $c \to \infty$. One can confirm this intuitive fact with the following simple argument. For $z > 0$, $c > 0$, and $0 < \varepsilon < 1$, we have that

$$\begin{aligned}\mathbb{P}(M_c > z) &\leq \mathbb{P}(\max_{0\le t\le \varepsilon}(B(t) - ct) > z) + \mathbb{P}(\max_{\varepsilon< t\le 1}(B(t) - ct) > z) \\ &\leq \mathbb{P}(\max_{0\le t\le \varepsilon} B(t) > z) + \mathbb{P}(\max_{\varepsilon< t\le 1}(B(t) - c\varepsilon) > z) \\ &\leq \mathbb{P}(\max_{0\le t\le \varepsilon} B(t) > z) + \mathbb{P}(\max_{0< t\le 1} B(t) > c\varepsilon + z) \\ &= 2\left(1 - \Phi\left(\frac{z}{\sqrt{\varepsilon}}\right)\right) + 2\left(1 - \Phi(c\varepsilon + z)\right). \end{aligned} \tag{5.1.23}$$

But, as c and ε are arbitrary, we can first take the limsup of (5.1.23) as $c \to \infty$, and then let $\varepsilon \to 0$, proving the claim.

We have thus shown that

$$\limsup_{n\to\infty} \mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - c_n t) > z) \le 0,$$

and since the functional clearly is equal to zero when $t = 0$, we have

$$\max_{0\le t\le 1}(\hat{B}_n(t) - c_n t) \xrightarrow{\mathbb{P}} 0, \tag{5.1.24}$$

as $n \to \infty$. Thus, by the Continuous Mapping Theorem, and the Converging Together Lemma, we obtain the weak convergence result

$$\frac{LI_n - \pi_{max} n}{\tilde{\sigma}\sqrt{n}} \Rightarrow -\frac{1}{2}B(1). \tag{5.1.25}$$

Lastly, consider (5.1.20). Here we need simply note the following equality in law, which follows from the stationary and Markovian nature of the underlying sequence $(X_n)_{n\ge 0}$:

$$\begin{aligned} \hat{B}_n(t) - \hat{B}_n(1) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2))(1-t) \\ \stackrel{\mathcal{L}}{=} -\hat{B}_n(1-t) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2))(1-t), \end{aligned} \tag{5.1.26}$$

for $t = 0, 1/n, \ldots, (n-1)/n, 1$. With a change of variables ($u = 1-t$), and noting that $B(t)$ and $-B(t)$ are equal in law, our previous convergence result (5.1.24) implies that

$$\max_{0\le t\le 1}(\hat{B}_n(t) - \hat{B}_n(1) - c_n(1-t)) \stackrel{\mathcal{L}}{=} \max_{0\le u\le 1}(-\hat{B}_n(u) - c_n u) \xrightarrow{\mathbb{P}} 0, \tag{5.1.27}$$

as $n \to \infty$. Our limiting functional is thus of the form

$$\frac{LI_n - \pi_{max} n}{\tilde{\sigma}\sqrt{n}} \Rightarrow \frac{1}{2} B(1). \tag{5.1.28}$$

Since $B(1)$ is simply a standard normal random variable, the different signs in (5.1.25) and (5.1.28) are inconsequential.

Finally, consider the degenerate cases. If either $a = 0$ or $b = 0$, then the sequence $(X_n)_{n\geq 0}$ will be a.s. constant, regardless of the starting state, and so $LI_n \sim n$. On the other hand, if $a = b = 1$, then the sequence oscillates back and forth between α_1 and α_2, so that $LI_n \sim n/2$. Combining these trivial cases with the previous development, we have proved the following theorem:

Theorem 5.1.1 *Let $(X_n)_{n\geq 0}$ be a 2-state Markov chain, with transition probabilities $\mathbb{P}(X_{n+1} = \alpha_2 |\ X_n = \alpha_1) = a$ and $\mathbb{P}(X_{n+1} = \alpha_1 | X_n = \alpha_2) = b$. Let the law of X_0 be the invariant distribution $(\pi_1, \pi_2) = (b/(a+b), a/(a+b))$, for $0 < a + b \leq 2$, and $(\pi_1, \pi_2) = (1, 0)$, for $a = b = 0$. Then, for $a = b > 0$,*

$$\frac{LI_n - n/2}{\sqrt{n}} \Rightarrow \sqrt{\frac{1-a}{a}} \left(-\frac{1}{2} B(1) + \max_{0\leq t\leq 1} B(t) \right), \tag{5.1.29}$$

where $B(t)$ is a standard Brownian motion, and for $a \neq b$ or $a = b = 0$, and $\pi_{max} = \max\{\pi_1, \pi_2\}$,

$$\frac{LI_n - \pi_{max} n}{\sqrt{n}} \Rightarrow N(0, \tilde{\sigma}^2/4), \tag{5.1.30}$$

where $N(0, \tilde{\sigma}^2/4)$ is a centered normal random variable with variance $\tilde{\sigma}^2/4 = ab(2 - a - b)/(a+b)^3$, for $a \neq b$, and $\tilde{\sigma}^2 = 0$, for $a = b = 0$. (If $a = b = 1$, or $\tilde{\sigma}^2 = 0$, then the distributions in (5.1.29) and (5.1.30), respectively, are understood to be degenerate at the origin.)

To extend this result to the entire Young tableau, let us introduce the following notation. By

$$(Y_n^{(1)}, Y_n^{(2)}, \dots, Y_n^{(k)}) \Rightarrow (Y_\infty^{(1)}, Y_\infty^{(2)}, \dots, Y_\infty^{(k)}) \tag{5.1.31}$$

we shall mean the weak convergence of the *joint* law of a k-vector $(Y_n^{(1)}, Y_n^{(2)}, \dots, Y_n^{(k)})$ to that of $(Y_\infty^{(1)}, Y_\infty^{(2)}, \dots, Y_\infty^{(k)})$, as $n \to \infty$. As noted above, since LI_n is the length of the top row of the associated Young tableau, the length of the second row is simply $n - LI_n$. Denoting the length of the i^{th} row by R_n^i, (5.1.31), together with an application of the Cramér-Wold Theorem, recovers the result of Chistyakov and Götze [13] as part of the following easy corollary, which is in fact equivalent to Theorem 5.1.1:

Corollary 5.1.1 *For the sequence in Theorem 5.1.1, if $a = b > 0$, then*

$$\left(\frac{R_n^1 - n/2}{\sqrt{n}}, \frac{R_n^2 - n/2}{\sqrt{n}}\right) \Rightarrow R_\infty := (R_\infty^1, R_\infty^2), \tag{5.1.32}$$

where the law of Y_∞ is supported on the 2^{nd} main diagonal of $\mathbb{R}^2$, and with

$$R_\infty^1 \stackrel{\mathcal{L}}{=} \sqrt{\frac{1-a}{a}} \left(-\frac{1}{2}B(1) + \max_{0 \le t \le 1} B(t)\right).$$

If $a \neq b$ or $a = b = 0$, then setting $\pi_{min} = \min\{\pi_1, \pi_2\}$, we have

$$\left(\frac{R_n^1 - \pi_{max} n}{\sqrt{n}}, \frac{R_n^2 - \pi_{min} n}{\sqrt{n}}\right) \Rightarrow N((0,0), \tilde{\Sigma}), \tag{5.1.33}$$

where $\tilde{\Sigma}$ is the covariance matrix

$$(\tilde{\sigma}^2/4) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

where $\tilde{\sigma}^2 = 4ab(2 - a - b)/(a + b)^3$, for $a \neq b$, and $\tilde{\sigma}^2 = 0$, for $a = b = 0$.

Remark 5.1.1 *The joint distributions in* (5.1.32) *and* (5.1.33) *are of course degenerate, in that the sum of the two components is a.s. identically zero in each case. In*

(5.1.32), *the density of the first component of* R_∞ *is easy to find, and is given by (e.g., see [27])*

$$f(y) = \frac{16}{\sqrt{2\pi}} \left(\frac{a}{1-a} \right)^{3/2} y^2 e^{-2ay^2/(1-a)}, \qquad y \geq 0. \tag{5.1.34}$$

As in Chistyakov and Götze [13], (5.1.32) *can then be stated as: For any bounded, continuous function* $g : \mathbb{R}^2 \to \mathbb{R}$,

$$\lim_{n\to\infty} \left(g \left(\frac{R_n^1 - n/2}{\sqrt{(1-a)n/a}}, \frac{R_n^2 - n/2}{\sqrt{(1-a)n/a}} \right) \right)$$
$$= 2\sqrt{2\pi} \int_0^\infty g(x, -x)\phi_{GUE,2}(x, -x)dx,$$

where $\phi_{GUE,2}$ *is the density of the eigenvalues of the* 2×2 *GUE, and is given by*

$$\phi_{GUE,2}(x_1, x_2) = \frac{1}{\pi}(x_1 - x_2)^2 e^{-(x_1^2 + x_2^2)}.$$

To see the GUE connection more explicitly, consider the 2×2 *traceless GUE matrix*

$$M_0 = \begin{pmatrix} X_1 & Y + iZ \\ Y - iZ & X_2 \end{pmatrix},$$

where X_1, X_2, Y, *and* Z *are centered, normal random variables. Since Corr* $(X_1, X_2) = -1$, *the largest eigenvalue of* M_0 *is*

$$\lambda_{1,0} = \sqrt{X_1^2 + Y^2 + Z^2},$$

almost surely, so that $\lambda_{1,0}^2 \sim \chi_3^2$ *if* $\operatorname{Var} X_1 = \operatorname{Var} Y = \operatorname{Var} Z = 1$. *Hence, up to a scaling factor, the density of* $\lambda_{1,0}$ *is given by* (5.1.34). *Next, let us perturb* M_0 *to*

$$M = \alpha G I + \beta M_0,$$

where α and β are constants, G is a standard normal random variable independent of M_0, and I is the identity matrix. The covariance of the diagonal elements of M is then computed to be $\rho := \alpha^2 - \beta^2$. Hence, to obtain a desired value of ρ, we may take $\alpha = \sqrt{(1+\rho)/2}$ and $\beta = \sqrt{(1-\rho)/2}$. Clearly, the largest eigenvalue of M can then be expressed as

$$\lambda_1 = \sqrt{\frac{1+\rho}{2}}G + \sqrt{\frac{1-\rho}{2}}\lambda_{1,0}. \tag{5.1.35}$$

At one extreme, $\rho = -1$, we recover $\lambda_1 = \lambda_{1,0}$. At the other extreme, $\rho = 1$, we obtain $\lambda_1 = Z$. Midway between these two extremes, at $\rho = 0$, we have a standard GUE matrix, so that

$$\lambda_1 = \sqrt{\frac{1}{2}}\left(G + \lambda_{1,0}\right).$$

5.2 Combinatorics Revisited

The original combinatorial development for the m-letter alphabet resulted in $m-1$ quantities S_n^r, $1 \le r \le m-1$. In the 2-letter case we were then able to proceed with a probabilistic development which involved a *single* Brownian motion. Using an even more straightforward development which involves m quantities instead, we can obtain more symmetric expressions for LI_n. This is done next, and will prove useful when studying the shape of the whole Young tableau.

Recall that a_k^r counts the number of occurrences of α_r among $(X_i)_{1\le i\le k}$. Moving beyond the purely combinatorial setting, assume that $(X_k)_{k\ge 0}$ is an infinite sequence generated by an irreducible homogeneous Markov chain having a stationary distribution $(\pi_1, \pi_2, \ldots, \pi_m)$. (For no $k \ge 0$ is the law of X_k necessarily assumed to be the stationary distribution.) For each $1 \le r \le m$, set $T_k^r = a_k^r - \pi_r k$, for $k \ge 1$, and $T_0^r = 0$. Beginning again with (2.1.1), we find that

$$
\begin{aligned}
LI_n &= \max_{\substack{0\le k_1\le\cdots\\ \le k_{m-1}\le n}} \Big[(a^1_{k_1}-a^1_0)+(a^2_{k_2}-a^2_{k_1})+\cdots+(a^m_n-a^m_{k_{m-1}})\Big]\\
&= \max_{\substack{0\le k_1\le\cdots\\ \le k_{m-1}\le n}} \Big[((T^1_{k_1}+\pi_1 k_1)-(T^1_{k_0}+\pi_1 k_0))+((T^2_{k_2}+\pi_2 k_2)-(T^2_{k_1}+\pi_2 k_1))\\
&\qquad +\cdots+((T^m_{k_m}+\pi_m k_m)-(T^m_{k_{m-1}}+\pi_m k_{m-1}))\Big]\\
&= \max_{\substack{0\le k_1\le\cdots\\ \le k_{m-1}\le n}} \Big[(T^1_{k_1}-T^1_{k_0})+(T^2_{k_2}-T^2_{k_1})+\cdots+(T^m_{k_m}-T^m_{k_{m-1}})\\
&\qquad +\pi_1(k_1-k_0)+\pi_2(k_2-k_1)+\cdots+\pi_m(k_m-k_{m-1})\Big]. \qquad (5.2.1)
\end{aligned}
$$

Setting $\pi_{max}=\max\{\pi_1,\pi_2,\ldots,\pi_m\}$, (5.2.1) becomes

$$
LI_n-\pi_{max}n=\max_{\substack{0=k_0\le k_1\le\cdots\\ \le k_{m-1}\le k_m=n}}\sum_{r=1}^{m}\big[(T^r_{k_r}-T^r_{k_{r-1}})+(\pi_r-\pi_{max})(k_r-k_{r-1})\big]. \qquad (5.2.2)
$$

For a uniform stationary distribution, $\pi_{max}=\pi_r=1/m$, for all r, and (5.2.2) simplifies to

$$
LI_n-\frac{n}{m}=\max_{\substack{0=k_0\le k_1\le\cdots\\ \le k_{m-1}\le k_m=n}}\sum_{r=1}^{m}(T^r_{k_r}-T^r_{k_{r-1}}). \qquad (5.2.3)
$$

To introduce a random walk formalism into the picture, we next set, for $i=1,\ldots,n$ and $r=1,2,\ldots,m$,

$$
W^r_i=\begin{cases}1, & \text{if } X_i=\alpha_r,\\ 0, & \text{otherwise.}\end{cases} \qquad (5.2.4)
$$

Clearly, $a^r_k=\sum_{i=1}^k W^r_i$, and so $T^r_k=\sum_{i=1}^k(W^r_i-\pi_r)$, for $1\le r\le m$.

To understand the limiting law of (5.2.2) or (5.2.3), we must have a more precise description of the underlying Markovian structure. To that end, let $p_{r,s}=\mathbb{P}(X_{k+1}=\alpha_s|X_k=\alpha_r)$ be the transition probability from state α_r to state α_s, and let $P=(p_{r,s})$ be the associated Markov transition matrix. In this setting,

$$(p_1^{n+1}, p_2^{n+1}, \ldots, p_m^{n+1}) = (p_1^n, p_2^n, \ldots, p_m^n)P.$$

Moreover, as usual, let $p_{r,s}^{(k)}$ denote the k-step transition probability from α_r to α_s; its associated transition matrix is simply P^k.

Assume now that the law of X_0 is the stationary distribution. Thus, by construction, $\mathbb{E}T_k^r = 0$ for all $1 \le r \le m$ and $1 \le k \le n$, and our primary task is to describe the covariance structure of these random variables T_k^r.

Since W_i^r is, simply, a Bernoulli random variable with parameter π_r, $\operatorname{Var} W_i^r = \pi_r(1-\pi_r)$. We then find that, for $k \ge 1$,

$$\begin{aligned}
\operatorname{Var} T_k^r &= \operatorname{Var}\left(\sum_{i=1}^{k}(W_i^r - \pi_r)\right) \\
&= \sum_{i=1}^{k} \operatorname{Var} W_i^r + \sum_{i_1=1}^{k-1}\sum_{i_2=i_1+1}^{k} \operatorname{Cov}(W_{i_1}^r, W_{i_2}^r) \\
&\qquad + \sum_{i_1=2}^{k}\sum_{i_2=1}^{i_1-1} \operatorname{Cov}(W_{i_1}^r, W_{i_2}^r).
\end{aligned} \tag{5.2.5}$$

By stationarity, (5.2.5) becomes

$$\begin{aligned}
\operatorname{Var} T_k^r &= \sum_{i=1}^{k} \operatorname{Var} W_i^r + \sum_{i_1=1}^{k-1}\sum_{i_2=i_1+1}^{k} \operatorname{Cov}(W_0^r, W_{i_2-i_1}^r) \\
&\qquad + \sum_{i_1=2}^{k}\sum_{i_2=1}^{i_1-1} \operatorname{Cov}(W_0^r, W_{i_1-i_2}^r) \\
&= k\pi_r(1-\pi_r) + \sum_{i_1=1}^{k-1}\sum_{i_2=i_1+1}^{k} (\pi_r p_{r,r}^{(i_2-i_1)} - \pi_r^2) \\
&\qquad + \sum_{i_1=2}^{k}\sum_{i_2=1}^{i_1-1} (\pi_r p_{r,r}^{(i_1-i_2)} - \pi_r^2) \\
&= k\pi_r - k^2\pi_r^2 + \pi_r \sum_{i_1=1}^{k-1}\sum_{i_2=i_1+1}^{k} e_r P^{i_2-i_1} e_r^T \\
&\qquad + \pi_r \sum_{i_1=2}^{k}\sum_{i_2=1}^{i_1-1} e_r P^{i_1-i_2} e_r^T,
\end{aligned} \tag{5.2.6}$$

where $e_r = (0, 0, \ldots, 0, 1, 0, \ldots 0)$ is the r^{th} standard basis vector of $\mathbb{R}^m$. Setting

$$Q_k = \sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} P^{i_2-i_1} = \sum_{i=1}^{k}(k-i)P^i, \tag{5.2.7}$$

we can rewrite (5.2.6) in the simple form

$$\operatorname{Var} T_k^r = k\pi_r - k^2\pi_r^2 + 2\pi_r e_r Q_k e_r^T. \tag{5.2.8}$$

Our description of the covariance structure can now be completed using the above results. For $r_1 \neq r_2$ and $k \geq 1$,

$$\begin{aligned}
\operatorname{Cov}(T_k^{r_1}, T_k^{r_2}) &= \sum_{i=1}^{k} \operatorname{Cov}(W_i^{r_1}, W_i^{r_2}) + \sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} \operatorname{Cov}(W_{i_1}^{r_1}, W_{i_2}^{r_2}) \\
&\quad + \sum_{i_1=2}^{k} \sum_{i_2=1}^{i_1-1} \operatorname{Cov}(W_{i_1}^{r_1}, W_{i_2}^{r_2}) \\
&= \sum_{i=1}^{k} \operatorname{Cov}(W_i^{r_1}, W_i^{r_2}) + \sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} \operatorname{Cov}(W_0^{r_1}, W_{i_2-i_1}^{r_2}) \\
&\quad + \sum_{i_1=2}^{k} \sum_{i_2=1}^{i_1-1} \operatorname{Cov}(W_0^{r_2}, W_{i_1-i_2}^{r_1}) \\
&= -k\pi_{r_1}\pi_{r_2} + \sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} (\pi_{r_1} p_{r_1,r_2}^{(i_2-i_1)} - \pi_{r_1}\pi_{r_2}) \\
&\quad + \sum_{i_1=2}^{k} \sum_{i_2=1}^{i_1-1} (\pi_{r_2} p_{r_2,r_1}^{(i_1-i_2)} - \pi_{r_1}\pi_{r_2}) \\
&= -k^2\pi_{r_1}\pi_{r_2} + \pi_{r_1} \sum_{i_1=1}^{k-1} \sum_{i_2=i_1+1}^{k} e_{r_1} P^{i_2-i_1} e_{r_2}^T \\
&\quad + \pi_{r_2} \sum_{i_1=2}^{k} \sum_{i_2=1}^{i_1-1} e_{r_2} P^{i_1-i_2} e_{r_1}^T \\
&= -k^2\pi_{r_1}\pi_{r_2} + \pi_{r_1} e_{r_1} Q_k e_{r_2}^T + \pi_{r_2} e_{r_2} Q_k e_{r_1}^T.
\end{aligned} \tag{5.2.9}$$

Remark 5.2.1 *Both* (5.2.8) *and* (5.2.9) *appear to be asymptotically quadratic in* k. *However, since* $Q_k = \sum_{i=i}^{k}(k-i)P^i$, *cancellations will show that when the Markov chain is irreducible and aperiodic, the order of the variance is, in fact, linear in* k.

In order to further analyze the asymptotics of Q_k, we first examine the diagonalization of P for a very general class of transition matrices.

Proposition 5.2.1 *Let* P *be the* $m \times m$ *transition matrix of an irreducible, aperiodic, homogeneous Markov chain with eigenvalues* $\lambda_1 = 1, \lambda_2, \ldots, \lambda_m$, *and let* $\Lambda = \mathrm{diag}(1, \lambda_2, \ldots, \lambda_m)$. *Let* $P = S^{-1}\Lambda S$ *be the diagonalization of* P, *where the rows of* S *consist of the left-eigenvectors of* P, *with, moreover, the first row of* S *being the stationary distribution* $(\pi_1, \pi_2, \ldots, \pi_m)$. *Then the first column of* S^{-1} *is* $(1, 1, \ldots, 1)^T$.

Proof. Since $P = S^{-1}\Lambda S$, then $PS^{-1} = S^{-1}\Lambda$. Denoting the first column of S^{-1} by c_1, we have $Pc_1 = c_1$. But since the rows of P sum to 1, we see that $c_1 = (1, 1, \ldots, 1)^T$ satisfies $Pc_1 = c_1$. Moreover, c_1 must be unique, up to normalization, since the irreducibility of P implies that $\lambda_1 = 1$ has multiplicity 1. Finally, since the inner product of the first row of S and the first column of S^{-1} is 1, the correct normalization is indeed $(1, 1, \ldots, 1)^T$. ■

Returning to Q_k, as given in (5.2.7), and using Proposition 5.2.1, we then obtain:

Theorem 5.2.1 *Let* $(X_n)_{n \geq 0}$ *be a sequence generated by an* m*-letter, aperiodic, irreducible, homogeneous Markov chain with state space* $\mathcal{A}_m = \{\alpha_1 < \cdots < \alpha_m\}$, *transition matrix* P, *and stationary distribution* $(\pi_1, \pi_2, \ldots, \pi_m)$. *Let also the law of* X_0 *be the stationary distribution. Moreover, for* $1 \leq r \leq m$, *let* $T_k^r = a_k^r - \pi_r k$, *for* $k \geq 1$, *and* $T_0^r = 0$, *where* a_k^r *is the number of occurrences of* α_r *among* $(X_i)_{1 \leq i \leq k}$. *Then, for* $1 \leq r \leq m$,

$$\lim_{k \to \infty} \frac{\mathrm{Var}\, T_k^r}{k} = \pi_r \left(1 + 2e_r S^{-1} D S e_r^T\right), \tag{5.2.10}$$

and for $r_1 \neq r_2$,

$$\lim_{k\to\infty} \frac{\mathrm{Cov}(T_k^{r_1}, T_k^{r_2})}{k} = \pi_{r_1} e_{r_1} S^{-1} D S e_{r_2}^T + \pi_{r_2} e_{r_2} S^{-1} D S e_{r_1}^T, \tag{5.2.11}$$

where $P = S^{-1}\Lambda S$ *is the standard diagonalization of* P *in Proposition 5.2.1, and* $D = \mathrm{diag}(-1/2, \lambda_2/(1-\lambda_2), \ldots, \lambda_m/(1-\lambda_m))$. *That is, the asymptotic covariance matrix of* $(T_k^r,)_{1\leq r\leq m}$ *is given by*

$$\Sigma = \Pi + \Pi(S^{-1}DS) + (S^{-1}DS)^T\Pi, \tag{5.2.12}$$

where $\Pi = \mathrm{diag}(\pi_1, \pi_2, \ldots, \pi_m)$.

Proof. Beginning with (5.2.7), we diagonalize P and find that

$$\begin{aligned} Q_k &= \sum_{i=1}^{k-1} (k-i)(S^{-1}\Lambda S)^i \\ &= S^{-1}\left(\sum_{i=1}^{k-1}(k-i)\Lambda^i\right) S \\ &= S^{-1}\,\mathrm{diag}(h(1), h(\lambda_2), \ldots, h(\lambda_m))S, \end{aligned} \tag{5.2.13}$$

where $h(\lambda) := \sum_{k=1}^{n-1}(n-k)\lambda^k$. Now $h(1) = k(k-1)/2$ is quadratic in k, while for $\lambda \neq 1$,

$$h(\lambda) = k\frac{\lambda}{(1-\lambda)} + \frac{\lambda(\lambda^k - 1)}{(1-\lambda)^2},$$

so that $h(\lambda)$ is linear in k. We thus can write Q_k as the sum of terms which are, respectively, quadratic and linear in k. Recalling, moreover, that the first row of S contains the stationary distribution, and that the first column of S^{-1} is $(1, 1, \ldots, 1)^T$, we have

$$
\begin{aligned}
Q_k &= S^{-1} \operatorname{diag}(h(1), h(\lambda_2), \ldots, h(\lambda_m))S, \\
&= \frac{k^2}{2} S^{-1} \operatorname{diag}(1, 0, \ldots, 0)S \\
&\quad + kS^{-1} \operatorname{diag}\left(-\frac{1}{2}, \frac{\lambda_2}{1-\lambda_2}, \ldots, \frac{\lambda_m}{1-\lambda_m}\right) S + o(k) \\
&= \frac{k^2}{2} \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_m \\ \pi_1 & \pi_2 & \cdots & \pi_m \\ \vdots & \vdots & \cdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_m \end{pmatrix} + kS^{-1}DS + o(k).
\end{aligned} \tag{5.2.14}
$$

Starting with the variance in (5.2.8), we now find that, for each $1 \leq r \leq m$,

$$
\begin{aligned}
\operatorname{Var} T_k^r &= k\pi_r - k^2\pi_r^2 + 2\pi_r e_r Q_k e_r^T \\
&= k\pi_r - k^2\pi_r^2 + 2\pi_r \left(\frac{k^2}{2}\pi_r + ke_r S^{-1} DS e_r^T\right) + o(k) \\
&= k\pi_r \left(1 + 2e_r S^{-1} DS e_r^T\right) + o(k),
\end{aligned} \tag{5.2.15}
$$

from which the asymptotic result (5.2.10) follows immediately.

An identical development shows that, for $r_1 \neq r_2$, (5.2.9) simplifies to

$$
\begin{aligned}
\operatorname{Cov}(T_k^{r_1}, T_k^{r_2}) &= -k^2\pi_{r_1}\pi_{r_2} + \pi_{r_1} e_{r_1} Q_k e_{r_2}^T + \pi_{r_2} e_{r_2} Q_k e_{r_1}^T \\
&= -k^2\pi_{r_1}\pi_{r_2} + \pi_{r_1}\left(\frac{k^2}{2}\pi_{r_2} + ke_{r_1} S^{-1} DS e_{r_2}^T\right) \\
&\quad + \pi_{r_2}\left(\frac{k^2}{2}\pi_{r_1} + ke_{r_2} S^{-1} DS e_{r_1}^T\right) + o(k) \\
&= k\left(\pi_{r_1} e_{r_1} S^{-1} DS e_{r_2}^T + \pi_{r_2} e_{r_2} S^{-1} DS e_{r_1}^T\right) + o(k),
\end{aligned} \tag{5.2.16}
$$

from which the asymptotic result (5.2.11) follows, and so does (5.2.12). ■

Remark 5.2.2 *To see that* (5.2.10) *and* (5.2.11) *both recover the covariance results for the iid case in Chapter II, let P be the transition matrix whose rows each consist of the stationary distribution $(\pi_1, \pi_2, \ldots, \pi_m)$. In this case $\lambda_2 = \cdots = \lambda_m = 0$, and so $D = \mathrm{diag}(-1/2, 0, \ldots, 0)$. Hence,*

$$\begin{aligned} e_{r_1} S^{-1} D S e_{r_2}^T &= (1, *, \ldots, *)\, D\, (\pi_{r_2}, *, \ldots, *)^T \\ &= -\frac{\pi_{r_2}}{2}, \end{aligned}$$

for all r_1 and r_2, and so, for each r,

$$\lim_{k\to\infty} \frac{\mathrm{Var}\, T_k^r}{k} = \pi_r \left(1 + 2\left(-\frac{\pi_r}{2}\right)\right) = \pi_r(1 - \pi_r),$$

while, for $r_1 \neq r_2$,

$$\lim_{k\to\infty} \frac{\mathrm{Cov}(T_k^{r_1}, T_k^{r_2})}{k} = \pi_{r_1}\left(-\frac{\pi_{r_2}}{2}\right) + \pi_{r_2}\left(-\frac{\pi_{r_1}}{2}\right) = -\pi_{r_1}\pi_{r_2}.$$

Note that, in the uniform iid case, we have $\pi_r = 1/m$, for all $1 \leq r \leq m$. Hence, for $r_1 \neq r_2$, the asymptotic correlation between $T_k^{r_1}$ and $T_k^{r_2}$ is given by $(-1/(m^2))/((1/m)(1-1/m)) = -1/(m-1)$, so that the covariance matrix is indeed the permutation-symmetric one obtained in the iid uniform case in Chapter II.

5.3 The Limiting Shape of the Young Tableau

Thus far, our results have centered on LI_n alone, essentially ignoring the larger question of the structure of the entire Young tableau. The present section extends the combinatorial development of the previous section to answer the question of the limiting shape of the Young tableau.

Our first result in this direction is a purely combinatorial expression generalizing (5.0.31). It is standard in the Young tableau literature to have entries chosen from the set $\{1, 2, \ldots, m\}$. Below, without loss of generality, we allow our entries to be chosen from the m-letter ordered alphabet $\mathcal{A}_m = \{\alpha_1 < \cdots < \alpha_m\}$.

Theorem 5.3.1 *Let $R_n^1, R_n^2, \ldots, R_n^r$ be the lengths of the first $1 \leq r \leq m$ rows of the Young tableau generated by the sequence $(X_k)_{1 \leq k \leq n}$ whose elements belong to an ordered alphabet $\mathcal{A}_m = \{\alpha_1 < \cdots < \alpha_m\}$. Then, for each $1 \leq r \leq m$, the sum of the lengths of the first r rows of the Young tableau is given by*

$$\sum_{j=1}^{r} R_n^j = \max_{k_{j,\ell} \in J_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \left(a_{k_{j,\ell}}^{\ell} - a_{k_{j,\ell-1}}^{\ell} \right), \tag{5.3.1}$$

where $J_{r,m} = \{(k_{j,\ell}, 1 \leq j \leq r, 0 \leq \ell \leq m) : k_{j,j-1} = 0, k_{j,m-r+j} = n, 1 \leq j \leq r; k_{j,\ell-1} \leq k_{j,\ell}, 1 \leq j \leq r, 1 \leq \ell \leq m; k_{j,\ell} \leq k_{j-1,\ell}, 2 \leq j \leq r, 1 \leq \ell \leq m\}$, and where a_k^{ℓ} is the number of occurrences of α_ℓ among $(X_i)_{1 \leq k \leq k}$.

Proof. Recall that the sum of the lengths of the first r rows of the Young tableau generated by a sequence $(X_k)_{1 \leq k \leq n}$, whose letters arise from an m-letter alphabet, has an interpretation in terms of the length of certain increasing sequences. Indeed, the sum $R_n^1 + R_n^2 + \cdots + R_n^r$ is equal to the maximum sum of the lengths of r disjoint, increasing subsequences of $(X_k)_{1 \leq k \leq n}$, where by *disjoint* it is meant that each element of $(X_k)_{1 \leq k \leq n}$ occurs in at most one of the r subsequences. (See Lemma 1 of Section 3.2 in [18]). More general results of this sort, involving partial orderings of the alphabet and associated antichains, are known as Greene's Theorem [23]. However, such results are not enough for our purpose. Below we need a different way of reconstructing disjoint subsequences.

We begin by examining an arbitrary collection of r disjoint, increasing subsequences of $(X_k)_{1 \leq k \leq n}$, and show that we can always map these r subsequences onto another collection of r disjoint, increasing subsequences whose properties will be amenable to our combinatorial analysis.

Specifically, with the number of rows r fixed, suppose that, for each $1 \leq j \leq r$, we have an increasing subsequence $(X^j_{k^j_\ell})_{1\leq \ell \leq n_j}$ of length $n_j \leq n$, and that the r subsequences are disjoint.

We first construct the new subsequence $(\tilde{X}^1_{\tilde{k}^1_\ell})_{1\leq \ell \leq \tilde{n}_1}$ as follows. First, place all α_1s occurring among the r original subsequences into $(\tilde{X}^1_{\tilde{k}^1_\ell})_{1\leq \ell \leq \tilde{n}_1}$, if there are any. If the last α_1 occurs at the n^{th} index, then $(\tilde{X}^1_{\tilde{k}^1_\ell})_{1\leq \ell \leq \tilde{n}_1}$ is complete. Otherwise, place all α_2s which occur after the final α_1 into $(\tilde{X}^1_{\tilde{k}^1_\ell})_{1\leq \ell \leq \tilde{n}_1}$, if there are any. If the last α_2 occurs at the n^{th} index, then $(\tilde{X}^1_{\tilde{k}^1_\ell})_{1\leq \ell \leq \tilde{n}_1}$ is complete. Otherwise, continue adding, successively, $\alpha_3, \ldots, \alpha_{m-r+1}$ in the same manner. Thus, $(\tilde{X}^1_{\tilde{k}^1_\ell})_{1\leq \ell \leq \tilde{n}_1}$ consists of a weakly increasing sequence of length $\tilde{n}_1$ having values in $\{\alpha_1, \ldots, \alpha_{m-r+1}\}$.

Next, we construct the new subsequence $(\tilde{X}^2_{\tilde{k}^2_\ell})_{1\leq \ell \leq \tilde{n}_2}$ similarly. By considering only those letters among the r original subsequences which have not already been moved to the first new subsequence, start with the smallest available letter, α_2, and continue adding, successively, $\alpha_3, \ldots, \alpha_{m+r-2}$. Note that, crucially, all α_2s added to $(\tilde{X}^2_{\tilde{k}^2_\ell})_{1\leq \ell \leq \tilde{n}_2}$ occur before the last index at which α_1 was added to the first subsequence. More generally, each α_j, $2 \leq j \leq m-r+2$, added to $(\tilde{X}^2_{\tilde{k}^2_\ell})_{1\leq \ell \leq \tilde{n}_2}$ occurs before the last α_{j-1} was added to the first subsequence. Thus, $(\tilde{X}^2_{\tilde{k}^2_\ell})_{1\leq \ell \leq \tilde{n}_2}$ consists of a weakly increasing subsequence of length $\tilde{n}_2$ having values in $\{\alpha_2, \ldots, \alpha_{m-r+2}\}$.

The construction of $(\tilde{X}^j_{\tilde{k}^j_\ell})_{1\leq \ell \leq \tilde{n}_j}$, for $3 \leq j \leq r$, continues in the same manner, with $(\tilde{X}^j_{\tilde{k}^j_\ell})_{1\leq \ell \leq \tilde{n}_j}$, constructed from among the entries of the r original subsequences which were not moved into any of the first $j-1$ new subsequences, so that $(\tilde{X}^j_{\tilde{k}^j_\ell})_{1\leq \ell \leq \tilde{n}_j}$, consists of a weakly increasing sequence of length $\tilde{n}_j$ having values in $\{\alpha_j, \ldots, \alpha_{m-r+j}\}$. It is possible that beyond some $j \geq 2$ the new subsequences may be empty.

We claim that, indeed, the construction of the r^{th} new subsequence exhausts the set of available entries. Indeed, without loss of generality, assume that after we have created the $(r-1)^{th}$ new subsequence, the set of available entries is non-empty, and designate the location of the final α_ℓ to be included in the j^{th} new subsequence by $k_{j,\ell}$, for $1 \le j \le r$ and $1 \le \ell \le m$. (If no α_ℓ was available for inclusion, set $k_{j,\ell} = k_{j,\ell-1}$, where $k_{j,0} = 0$, for all $1 \le j \le r$.) Clearly, all $\alpha_1, \alpha_2, \dots, \alpha_{r-1}$ have been included in the first $r-1$ new subsequences. If $r = m$, we are done: simply put the remaining α_rs into the r^{th} new subsequence. If $r < m$, we may still ask whether there was, for some $r+1 \le \ell \le m$, an α_ℓ from among the available entries which occurred before $k_{r,\ell-1}$. Assume that there is such an α_ℓ. Now by construction, $k_{j+1,\ell-r+j} \le k_{j,\ell-r+j-1}$, for $1 \le j \le r-1$. Hence, there exist letters $\alpha_{j_1} < \alpha_{j_2} < \cdots < \alpha_{j_r} \le \alpha_{\ell-1}$ among the original subsequences which occurred after $k_{r,\ell-1}$, and, moreover, each letter must come from a different subsequence. But since each original subsequence was increasing, none of them could have contained an α_ℓ before $k_{r,\ell-1}$, and we have a contradiction.

To better understand this construction, consider the first row of Figure 3, which shows an initial sequence of length $n = 12$, with $m = 4$ letters, broken into $r = 3$ disjoint, increasing subsequences of lengths $n_1 = 3, n_2 = 4$, and $n_3 = 3$, and so with total length 10. The final three rows of the diagram show the results of the operations described above, producing 3 new increasing subsequences of length $\tilde{n}_1 = 4, \tilde{n}_2 = 3$, and $\tilde{n}_3 = 3$.

Hence, if we wish to find r disjoint, increasing subsequences whose length sum is maximal, it suffices to consider only those disjoint, increasing subsequences for which the final occurrence of the letter α_ℓ in the subsequence i happens after the final occurrence in the subsequence j, whenever $i < j$. Because such ranges do not overlap, if we wish to count the number of α_ℓs in the j^{th} subsequence, it suffices to simply count the number of α_ss in $(X_k)_{1 \le k \le n}$ over that range.

Indeed, returning to the fundamental combinatorial objects of our development, the a_k^j, we see that since $a_\ell^j - a_k^j$ counts the number of α_js in the range $\ell+1, \dots, k$, we can describe the valid index ranges over which to search for the maximal sum as $J_{r,m} = \{(k_{j,\ell}, 1 \le j \le r, 0 \le \ell \le m) : k_{j,j-1} = 0, k_{j,m-r+j} = n, 1 \le j \le r; k_{j,\ell-1} \le$

1	4	2	1	3	4	3	4	1	4	2	4

			1					1		2	
1					4				4		4
		2		3		3					

1			1					1		2	
		2		3		3					
					4				4		4

Figure 3: Transformation of $r = 3$ subsequences.

$k_{j,\ell}, 1 \leq j \leq r, 1 \leq \ell \leq m; k_{j,\ell} \leq k_{j-1,\ell}, 2 \leq j \leq r, 1 \leq \ell \leq m\}$. The constraints on the $k_{j,\ell}$ follow simply from the fact that each subsequence is increasing and that, moreover, the intervals associated with a given letter do not overlap. Figure 4 indicates the relative positions of each range, for $r = 4$ and $m = 7$.

Since the first possible letter of each subsequence grows from α_1 to α_r, and the last possible letter grows from α_{m+r-1} to α_m, the result is proved. ■

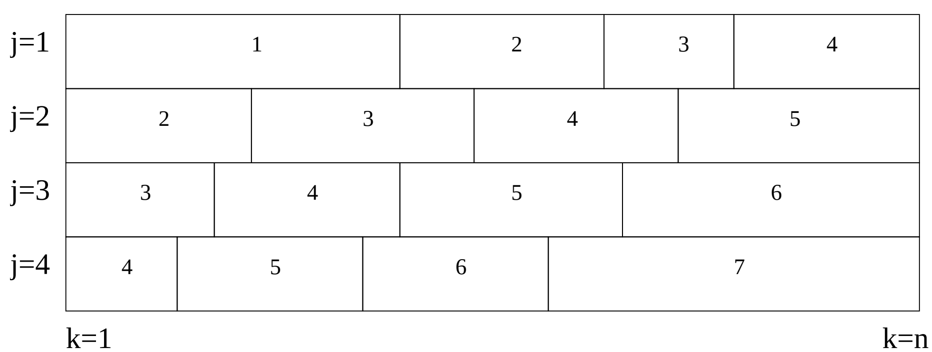

Figure 4: Schematic diagram of $J_{r,m}$, for $r = 4, m = 7$.

We are now ready to apply our asymptotic covariance results (Theorem 5.2.1), along with a Brownian sample-path approximation, to the combinatorial expression (5.3.1), and so obtain a Brownian functional expression for the limiting shape of the Young tableau for all irreducible, aperiodic, homogeneous Markov chains.

Indeed, for each $1 \leq r \leq m$, let the sum of the first r rows of the Young tableau be given by

$$V_n^r := \sum_{j=1}^{r} R_n^j = \max_{k_{j,\ell} \in J_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \left(a_{k_{j,\ell}}^{\ell} - a_{k_{j,\ell-1}}^{\ell} \right), \tag{5.3.2}$$

where the index set $J_{r,m}$ is defined as in Theorem 5.3.1. Define, as before, $T_k^r = \sum_{i=1}^{k}(W_i^r - \pi_r) = a_k^r - \pi_r k$, and so rewrite (5.3.2) as

$$\begin{aligned} V_n^r &= \max_{k_{j,\ell} \in J_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \left(\left(T_{k_{j,\ell}}^{\ell} + \pi_\ell k_{j,\ell} \right) - \left(T_{k_{j,\ell-1}}^{\ell} + \pi_\ell k_{j,\ell-1} \right) \right) \\ &= \max_{k_{j,\ell} \in J_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \left(\left(T_{k_{j,\ell}}^{\ell} - T_{k_{j,\ell-1}}^{\ell} \right) + \pi_\ell \left(k_{j,\ell} - k_{j,\ell-1} \right) \right). \end{aligned} \tag{5.3.3}$$

Next, let τ be a permutation of the indices $1, 2, \ldots, m$ such that $\pi_{\tau(1)} \geq \pi_{\tau(2)} \geq \cdots \geq \pi_{\tau(m)} > 0$. Moreover, we demand that if $\pi_{\tau(i)} = \pi_{\tau(j)}$ for $i < j$, then $\tau(i) < \tau(j)$. (The permutation so defined is thus unique.) As we are considering V_n^r, it is natural to define $\nu_r = \sum_{j=1}^{r} \pi_{\tau(j)}$, the sum of the r largest values among $\pi_1, \pi_2, \ldots, \pi_m$. We obtain, below, the limiting distribution of $(V_n^r - \nu_r n)/\sqrt{n}$ as a Brownian functional.

To introduce Brownian sample-path approximations, and for each $1 \leq r \leq m$, we first define the asymptotic variance of T_n^r as in (5.2.10), by

$$\sigma_r^2 := \lim_{n \to \infty} \frac{\operatorname{Var} T_n^r}{n} = e_r \Sigma e_r^T, \tag{5.3.4}$$

and, for $r_1 \neq r_2$, the asymptotic covariance of $T_n^{r_1}$ and $T_n^{r_2}$ by

$$\sigma_{r_1,r_2} := \lim_{n \to \infty} \frac{\operatorname{Cov}(T_n^{r_1}, T_n^{r_2})}{n} = e_{r_1} \Sigma e_{r_2}^T, \tag{5.3.5}$$

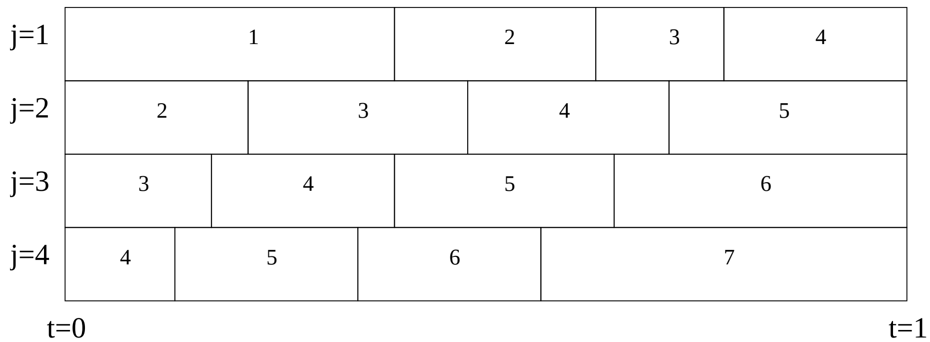

Figure 5: Schematic diagram of $I_{s,d}$, for $s = 4, d = 7$.

where Σ is the covariance matrix of Theorem 5.2.1 associated with the transition matrix P. For each $1 \leq r \leq m$, we then let

$$\hat{B}_n^r(t) = \frac{T_{[nt]}^r + (nt - [nt])(W_{[nt]+1}^r - \pi_r)}{\sigma_r \sqrt{n}}, \tag{5.3.6}$$

for $0 \leq t \leq 1$. This rescaling of $[0, n]$ to $[0, 1]$ calls for us to define a new parameter set over which we will maximize a functional arising from the expressions in (5.3.6). Indeed, for any positive integers s and d, with $s \leq d$, define the set

$$\begin{aligned} I_{s,d} = \Big\{(t_{j,\ell}, 1 \leq j \leq s, 0 \leq \ell \leq d) :& t_{j,j-1} = 0, t_{j,d-s+j} = 1, 1 \leq j \leq s; \\ & t_{j,\ell-1} \leq t_{j,\ell}, 1 \leq j \leq s, 1 \leq \ell \leq d; \\ & t_{j,\ell} \leq t_{j-1,\ell}, 2 \leq j \leq s, 1 \leq \ell \leq d\Big\}. \end{aligned}$$

Note that the constraints $t_{j,j-1} = 0$ and $t_{j,d-s+j} = 1$, for $1 \leq j \leq s$, force many of the $t_{j,\ell}$ to be zero or one. We will denote the $s \times (d+1)$-tuple elements of $I_{s,d}$, by $(t_{\cdot,\cdot})$. Figure 5 shows the structure of $I_{s,d}$, for $s = 4$ and $d = 7$. The locations of $t_{j,\ell}$ are indicated by the horizontal lines within the diagram.

With this notation, (5.3.3) becomes

$$\frac{V_n^r - \nu_r n}{\sqrt{n}} = \max_{(t_{\cdot,\cdot})\in I_{r,m}} \Big\{ \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \left(\hat{B}_n^\ell(t_{j,\ell}) - \hat{B}_n^\ell(t_{j,\ell-1}) \right)$$
$$+ \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sqrt{n}(\pi_\ell - \pi_{\tau(j)}) \left(t_{j,\ell} - t_{j,\ell-1}\right) \Big\}. \qquad (5.3.7)$$

Our analysis of (5.3.7) will yield the following theorem, whose proof we defer to the conclusion of the section. This theorem gives, in particular, a full characterization of the limiting shape of the Young tableau in the non-uniform iid case.

Theorem 5.3.2 *Let $(X_n)_{n\geq 0}$ be an irreducible, aperiodic, homogeneous Markov chain with finite state space $\mathcal{A}_m = \{\alpha_1 < \cdots < \alpha_m\}$, transition matrix P, and stationary distribution $(\pi_1, \pi_2, \ldots, \pi_m)$. Let $\Sigma = (\sigma_{r,s})_{1\leq r,s\leq m}$ be the associated asymptotic covariance matrix, as given in* (5.2.12)*, and let the law of X_0 be given by the stationary distribution. Let τ be the permutation of $\{1, 2, \ldots, m\}$ such that $\pi_{\tau(i)} \geq \pi_{\tau(i+1)}$, and $\tau(i) < \tau(j)$ whenever $\pi_{\tau(i)} = \pi_{\tau(j)}$ and $i < j$. For each $1 \leq r \leq m$, let V_n^r be the sum of the lengths of the first r rows of the associated Young tableau, and let $\nu_r = \sum_{j=1}^{r} \pi_{\tau(j)}$. Finally, let d_r be the multiplicity of $\pi_{\tau(r)}$, and let*

$$m_r = \begin{cases} 0, & \text{if } \pi_{\tau(r)} = \pi_{\tau(1)}, \\ \max\{i : \pi_{\tau(i)} > \pi_{\tau(r)}\}, & \text{otherwise.} \end{cases}$$

Then, for each $1 \leq r \leq m$,

$$\frac{V_n^r - \nu_r n}{\sqrt{n}} \Rightarrow V_\infty^r := \sum_{i=1}^{m_r} \sigma_{\tau(i)} \tilde{B}^{\tau(i)}(1)$$
$$+ \max_{I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(d_r+m_r-r+j)} \sigma_{\tau(m_r+\ell)} \left(\tilde{B}^{\tau(m_r+\ell)}(t_{j,\ell}) - \tilde{B}^{\tau(m_r+\ell)}(t_{j,\ell-1}) \right), \quad (5.3.8)$$

where the first sum on the right-hand side of (5.3.8) *is understood to be* 0*, if* $m_r = 0$*. Above,* $\sigma_r^2 = \sigma_{r,r}$*, and* $(\tilde{B}^1(t), \tilde{B}^2(t), \dots, \tilde{B}^m(t))$ *is an* m*-dimensional Brownian motion, with covariance matrix* $\tilde{\Sigma} = (\tilde{\sigma}_{r,s})_{1 \le r,s \le m}$ *given by*

$$(\tilde{\sigma}_{r,s}) = t(\sigma_{r,s})/\sigma_r\sigma_s, \tag{5.3.9}$$

for $1 \le r, s \le m$*. Moreover, for any* $1 \le k \le m$*,*

$$\left(\frac{V_n^1 - \nu_1 n}{\sqrt{n}}, \frac{V_n^2 - \nu_2 n}{\sqrt{n}}, \dots, \frac{V_n^k - \nu_k n}{\sqrt{n}}\right) \Rightarrow \left(V_\infty^1, V_\infty^2, \dots, V_\infty^k\right). \tag{5.3.10}$$

Remark 5.3.1 *The critical indices* d_r *and* m_r *in Theorem 5.3.2 are chosen so that*

$$\pi_{\tau(m_r)} > \pi_{\tau(m_r+1)} = \pi_{\tau(r)} = \cdots = \pi_{\tau(m_r+d_r)} > \pi_{\tau(m_r+d_r+1)}.$$

Thus, the functional in (5.3.8) *consists of a sum of* m_r *Gaussian random variables and a maximal functional involving only* d_r *of the* m *one-dimensional Brownian motions.*

Remark 5.3.2 *Another, perhaps more natural, way of describing the covariance structure of the* m*-dimensional Brownian motion in Theorem 5.3.2 is to note that* $(\sigma_1\tilde{B}^1(t), \sigma_2\tilde{B}^2(t), \dots, \sigma_m\tilde{B}^m(t))$ *has covariance matrix* $t\Sigma$*.*

Let us now examine the case $r = 1$. Here, as previously noted, $V_n^1 = LI_n$. Since $m_1 = 0$, (5.3.8) becomes

$$\frac{LI_n - \pi_{max} n}{\sqrt{n}} \Rightarrow \max_{(t_{\cdot,\cdot}) \in I_{1,d_1}} \sum_{\ell=1}^{d_1} \sigma_{\tau(\ell)} \left(\tilde{B}^{\tau(\ell)}(t_{1,\ell}) - \tilde{B}^{\tau(\ell)}(t_{1,\ell-1})\right), \tag{5.3.11}$$

where we have written π_{max} for $\pi_{\tau(1)}$. The functional in (5.3.11) is similar to the one obtained in the iid case in (2.2.39), namely $\tilde{H}_m$, the essential difference being, not in the form of the Brownian functional, but rather in the covariance structure of the Brownian motions.

To see precisely where this difference comes into play, note that if the transition matrix P is cyclic, then the covariance matrix of the Brownian motion is also cyclic. Consider then the 3-letter aperiodic, homogeneous, cyclic Markov case. Since the Brownian covariance matrix is symmetric, and, moreover, degenerate, an additional cyclicity constraint forces it to have the permutation-symmetric structure seen in the iid uniform case. In particular, LI_n will have, up to a scaling factor, the same limiting distribution as in the iid uniform case:

$$\frac{LI_n - n/3}{\sqrt{n}} \Rightarrow \sigma \max_{(t_{\cdot,\cdot}) \in I_{1,3}} \sum_{\ell=1}^{3} \left(\tilde{B}^{\ell}(t_{1,\ell}) - \tilde{B}^{\ell}(t_{1,\ell-1}) \right), \tag{5.3.12}$$

where $\sigma = \sigma_\ell$, for all $1 \leq \ell \leq 3$, and with the Brownian covariance matrix given by

$$\tilde{\Sigma} = t \begin{pmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{pmatrix},$$

and where we have used the fact that $\tau(\ell) = \ell$, for all $1 \leq \ell \leq 3$.

However, when $m \geq 4$, the cyclicity constraint does *not* force the Brownian covariance matrix to have the permutation-symmetric structure, as the following example shows for $m = 4$.

Example 5.3.1 *Consider the following doubly-stochastic, aperiodic, cyclic transition matrix:*

$$P = \begin{pmatrix} 0.4 & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.4 & 0.3 & 0.2 \\ 0.2 & 0.1 & 0.4 & 0.3 \\ 0.3 & 0.2 & 0.1 & 0.4 \end{pmatrix}. \tag{5.3.13}$$

While the doubly-stochastic nature of P ensures that the stationary distribution is uniform, the covariance matrix of the limiting Brownian motion, at three-decimal *accuracy, is computed to be*

$$\tilde{\Sigma} = t \begin{pmatrix} 1.000 & -0.357 & -0.287 & -0.357 \\ -0.357 & 1.000 & -0.357 & -0.287 \\ -0.287 & -0.357 & 1.000 & -0.357 \\ -0.357 & -0.287 & -0.357 & 1.000 \end{pmatrix}, \tag{5.3.14}$$

and $\sigma_r^2 = \sigma^2 := 0.263$*, for each* $1 \leq r \leq 4$*. Thus, the limiting distribution of* LI_n *is given by*

$$\frac{LI_n - n/4}{\sqrt{n}} \Rightarrow \sigma \max_{(t_{\cdot,\cdot}) \in I_{1,4}} \sum_{\ell=j}^{4} \left(\tilde{B}^{\ell}(t_{1,\ell}) - \tilde{B}^{\ell}(t_{1,\ell-1}) \right), \tag{5.3.15}$$

for $1 \leq r \leq 4$*. However, while the form of the functional is the same as in the iid uniform case (up to the constant), the covariance structure of the Brownian motion in* (5.3.14) *differs from that of the uniform iid case, i.e., from*

$$t \begin{pmatrix} 1 & -1/3 & -1/3 & -1/3 \\ -1/3 & 1 & -1/3 & -1/3 \\ -1/3 & -1/3 & 1 & -1/3 \\ -1/3 & -1/3 & -1/3 & 1 \end{pmatrix}, \tag{5.3.16}$$

and so the limiting distribution in (5.3.15) *is not that of the uniform iid case.*

We thus see that Kuperberg's conjecture regarding the shape of the Young tableau for random sequences generated by aperiodic, homogeneous, and cyclic matrices [32] is not true for general m-alphabets. By simply extending the first-row analysis above to the second and third rows, we see that it is true for $m = 3$. However, as could have been anticipated by (5.3.12), it fails for $m \geq 4$, as the previous example showed. Furthermore, in the next section we shall see that for the cyclic case the structure of Σ can be described in an elegant manner which delineates precisely when we obtain the uniform iid limiting law.

In the more general doubly stochastic case, we have the following corollary:

Corollary 5.3.1 *Let the transition matrix P of Theorem 5.3.2 be doubly stochastic. Then, for every $1 \le r \le m$, $m_r = 0, d_r = m$, and*

$$\frac{V_n^r - rn/m}{\sqrt{n}} \Rightarrow \max_{(t_{\cdot,\cdot}) \in I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \left(\tilde{B}^\ell(t_{j,\ell}) - \tilde{B}^\ell(t_{j,\ell-1}) \right). \tag{5.3.17}$$

If, moreover, the matrix P has all entries of $1/m$ (i.e., in the iid uniform alphabet case), then

$$\frac{V_n^r - rn/m}{\sqrt{n}} \Rightarrow \frac{\sqrt{m-1}}{m} \max_{(t_{\cdot,\cdot}) \in I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \left(\tilde{B}^\ell(t_{j,\ell}) - \tilde{B}^\ell(t_{j,\ell-1}) \right) \tag{5.3.18}$$

and the covariance matrix in (5.3.9) *has all its off-diagonals equal to* $-1/(m-1)$.

Proof. For each $1 \le r \le m$, $\pi_r = 1/m$, and so $\nu_r = r/m$, $m_r = 0$, and the multiplicity $d_r = m$. Moreover, the permutation τ is simply the identity permutation. This proves (5.3.17). If, moreover, all the transition probabilities are $1/m$, then the multinomial nature of the underlying combinatorial quantities a_k^r tells us that $\sigma_r^2 = (1/m)(1 - 1/m)$, for each $1 \le r \le m$, and that $\rho_{r_1,r_2} = -1/(m-1)$, for each $r_1 \ne r_2$, thus proving (5.3.18). ∎

To see that the functional in (5.3.17) is generally different from the uniform iid case, even for $m = 3$, consider the following non-cyclic example:

Example 5.3.2 *Let a doubly-stochastic (but non-cyclic), aperiodic Markov chain have transition matrix*

$$P = \begin{pmatrix} 0.4 & 0.6 & 0.0 \\ 0.6 & 0.0 & 0.4 \\ 0.0 & 0.4 & 0.6 \end{pmatrix}. \tag{5.3.19}$$

As in Example 5.3.1, the doubly-stochastic nature of P ensures that the stationary distribution is uniform. In the present example, the asymptotic covariance matrix, at three-decimal accuracy, is computed to be

$$\begin{pmatrix} 0.459 & 0.049 & -0.506 \\ 0.049 & 0.086 & -0.136 \\ -0.506 & -0.136 & 0.642 \end{pmatrix}. \tag{5.3.20}$$

Note that, even though we have a uniform stationary distribution, the asymptotic variances (i.e., the diagonals of (5.3.20)*) have dramatically different values. Moreover, according to Remark 5.2.2, in the uniform iid case, the only possibility for the Brownian covariance matrix is that the off-diagonals have value $-1/2$. However, the Brownian motion covariance matrix obtained from* (5.3.20) *is*

$$t\begin{pmatrix} 1.000 & 0.246 & -0.935 \\ 0.246 & 1.000 & -0.577 \\ -0.935 & -0.577 & 1.000 \end{pmatrix}. \tag{5.3.21}$$

Not only are the off-diagonals different from $-1/2$, but in some cases are even positive. In short, the functional in (5.3.17) *has a distribution which differs from any iid case (even non-uniform).*

Remark 5.3.3 *Generalizing a result of Baryshnikov [6] and of Gravner, Tracy, and Widom [22] on the representation of the maximal eigenvalue of an $m \times m$ element of the GUE, Doumerc [16] found a Brownian functional expression for all the eigenvalues of an $m \times m$ element of the GUE. Our expression in* (5.3.18) *is similar, with the exception that our m-dimensional Brownian motion is constrained by a zero-sum condition, and, moreover, has a different covariance structure. (We note, moreover, that the parameters over which his Brownian functional is maximized in [16] might be intended to range over a slightly larger set which corresponds to our $I_{r,m}$.) Using* *a path-transformation technique relating the joint distribution of a certain transformation of* n *continuous processes to the joint distribution of the processes conditioned never to leave the Weyl chamber, O'Connell and Yor [37] employed queuing-theoretic arguments to obtain Brownian functional representations for the entire spectrum of the* $m \times m$ *element of the GUE. In a study of much more general transformations of this type, Bougerol and Jeulin [9] were able to obtain this result as a special case.*

If $d_r = 1$, *i.e.*, if the r^{th} most probable state is unique, then the following result can be viewed as lying at the other extreme from Corollary 5.3.1:

Corollary 5.3.2 *Let* $1 \le r \le m$*, and let* $d_r = 1$ *in Theorem 5.3.2. Then*

$$\frac{V_n^r - \nu_r n}{\sqrt{n}} \Rightarrow \sum_{i=1}^{r} \sigma_{\tau(i)} \tilde{B}^{\tau(i)}(1). \tag{5.3.22}$$

Proof. If $d_r = 1$, then $m_r = r - 1$, and so the maximal term of (5.3.8) contains only one summand, namely $\sigma_{\tau(m_r+1)} \tilde{B}^{\tau(m_r+1)}(1) = \sigma_{\tau(r)} \tilde{B}^{\tau(r)}(1)$. Including this term in the first summation term of (5.3.8) proves (5.3.22). ■

Remark 5.3.4 *The maximal term of the functional in* (5.3.8) *is that of the doubly-stochastic,* d_r*-letter case. Indeed, the maximal term involves precisely* d_r *Brownian motions over the* $r - m_r$ *rows. Such a functional would arise in a doubly-stochastic* d_r*-letter situation with a covariance matrix consisting of the sub-matrix of the original* Σ *corresponding to the* d_r *Brownian motions, as in Corollary 5.3.1. The Gaussian term corresponds to the functional of Corollary 5.3.2. That is, in some sense, the limiting law of* (5.3.8) *interpolates between these two extreme cases.*

Proof. (Theorem 5.3.2) Since the $r = m$ case is trivial (V_n^m is then identically equal to n), assume that $r < m$. Recall the approximating functional (5.3.7):

$$\frac{V_n^r - \nu_r n}{\sqrt{n}} = \max_{I_{r,m}} \Bigg\{ \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \left(\hat{B}_n^\ell(t_{j,\ell}) - \hat{B}_n^\ell(t_{j,\ell-1}) \right) + \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sqrt{n}(\pi_\ell - \pi_{\tau(j)}) \left(t_{j,\ell} - t_{j,\ell-1}\right) \Bigg\}. \tag{5.3.23}$$

Introducing the notation $\Delta t_{j,\ell} := [t_{j,\ell-1}, t_{j,\ell-1}]$ and $M_n^\ell(\Delta t_{j,\ell}) := M_n^\ell(t_{j,\ell}) - M_n^\ell(t_{j,\ell-1})$, for any m-dimensional process $M(t) = (M^1(t), M^2(t), \ldots, M^m(t))$, $t \in [0,1]$, we can rewrite (5.3.23) more compactly as

$$\frac{V_n^r - \nu_r n}{\sqrt{n}} = \max_{I_{r,m}} \Bigg\{ \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \hat{B}_n^\ell(\Delta t_{j,\ell}) - \sqrt{n} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} (\pi_{\tau(j)} - \pi_\ell) |\Delta t_{j,\ell}| \Bigg\}. \tag{5.3.24}$$

The main idea of the proof to follow will be to show that the second summation of (5.3.24) can, in effect, be eliminated by choosing the $(\Delta t_{j,\ell})$ in an appropriate manner. Now some of the coefficients $(\pi_{\tau(j)} - \pi_\ell)$ are zero; such terms do not cause any problems. Intuitively, however, the remaining terms should have $|\Delta t_{j,\ell}| = 0$. Defining the restricted set of parameters $I_{r,m}^* = \{(t_{j,\ell}) \in I_{r,m} : \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} (\pi_\ell - \pi_{\tau(j)}) |\Delta t_{j,\ell}| = 0, 1 \le \ell \le m\}$, we see that, provided $I_{r,m}^* \neq \emptyset$,

$$\max_{I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \left(\sigma_\ell \hat{B}_n^\ell(\Delta t_{j,\ell}) - \sqrt{n} \left(\pi_{\tau(j)} - \pi_\ell\right) |\Delta t_{j,\ell}| \right) \geq \max_{I_{r,m}^*} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \hat{B}_n^\ell(\Delta t_{j,\ell}). \tag{5.3.25}$$

Moreover, by the Invariance Principle and the Continuous Mapping Theorem,

$$\max_{I_{r,m}^*} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \hat{B}_n^\ell(\Delta t_{j,\ell}) \Rightarrow \max_{I_{r,m}^*} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \tilde{B}^\ell(\Delta t_{j,\ell}). \tag{5.3.26}$$

We claim that, indeed, $I_{r,m}^* \neq \emptyset$, and that, moreover,

$$\max_{I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \left(\sigma_\ell \hat{B}_n^\ell(\Delta t_{j,\ell}) - \sqrt{n}\left(\pi_{\tau(j)} - \pi_\ell\right) |\Delta t_{j,\ell}| \right)$$

$$\Rightarrow \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \tilde{B}^\ell(\Delta t_{j,\ell}). \tag{5.3.27}$$

We will prove that $I^*_{r,m} \neq \emptyset$ by creating a bijection between $I^*_{r,m}$ and I_{r-m_r,d_r}. To this end, for $1 \le i \le m_r$, let $\tilde{I}_{\tau(i),i} = [u_{\tau(i),i-1}, u_{\tau(i),i}] = [0,1]$. Next, choose any $(u_{\cdot,\cdot}) \in I_{r-m_r,d_r}$, and define further intervals $\tilde{I}_{\tau(m_r+j),\ell} = \Delta u_{j,\ell}$, for $1 \le j \le r - m_r$ and $1 \le \ell \le d_r$.

We now create a partition of these intervals in a manner which relies on the ideas used in the proof of Theorem 5.3.1. Consider the *set* of points $\{u_{j,\ell}\}_{(1 \le j \le r-m_r, 1 \le \ell \le d_r)}$, and order them as $s_0 := 0 < s_1 < \cdots < s_{\kappa-1} < s_\kappa := 1$, for some integer κ, and let $\Delta s_q = [s_{q-1}, s_q]$, for all $1 \le q \le \kappa$.

Trivially, for each $1 \le q \le \kappa$, and for each $1 \le i \le m_r$, $\Delta s_q \subset \tilde{I}_{\tau(i),i}$. Moreover, for each $1 \le j \le r - m_r$, there exists a unique $\ell(j,q)$ such that $\Delta s_q \subset \tilde{I}_{\tau(m_r+j),\ell(j,q)}$. For each q, consider the set of indices $A_q := \{\tau(1), \ldots, \tau(m_r)\} \cup \{\tau(m_r + \ell(1,q)), \ldots, \tau(m_r + \ell(r - m_r, q))\}$, and order these r elements of A_q as $1 \le \tilde{\ell}(1,q) < \cdots < \tilde{\ell}(r,q) \le m$.

Using these partitions, we examine, with foresight, the following functional of a general m-dimensional process $(M(t))_{t \ge 0}$:

$$\sum_{i=1}^{m_r} M^{\tau(i)}(1) + \sum_{j=1}^{(r-m_r)} \sum_{\ell=j}^{(r-m_r+d_r-1)} M^{\tau(m_r+\ell)}(\Delta u_{j,\ell}) \tag{5.3.28}$$

$$= \sum_{i=1}^{m_r} \left(\sum_{q=1}^{\kappa} M^{\tau(i)}(\Delta s_q) \right) + \sum_{j=1}^{(r-m_r)} \sum_{\ell=j}^{(r-m_r+d_r-1)} \left(\sum_{q:\Delta s_q \subset \tilde{I}_{\tau(m_r+j),\ell}} M^{\tau(m_r+\ell)}(\Delta s_q) \right)$$

$$= \sum_{q=1}^{\kappa} \left(\sum_{i=1}^{m_r} M^{\tau(i)}(\Delta s_q) + \sum_{j=1}^{(r-m_r)} M^{\tau(m_r+\ell(j,q))}(\Delta s_q) \right)$$
$$= \sum_{q=1}^{\kappa} \sum_{j=1}^{r} M^{\tilde{\ell}(j,q)}(\Delta s_q) = \sum_{j=1}^{r} \sum_{q=1}^{\kappa} M^{\tilde{\ell}(j,q)}(\Delta s_q)$$
$$= \sum_{j=1}^{r} \sum_{\ell=1}^{r} M^{\tilde{\ell}(j,q)}(\Delta t_{j,\ell}), \tag{5.3.29}$$

where, for each $1 \le j \le r$, and for each $1 \le \ell \le m$, $t_{j,\ell} := \max\{s_q : \ell \ge \tilde{\ell}(j,q)\}$. (That is, for each j, we collapse together intervals Δs_q corresponding to the same component M^{ℓ}.) Now, since our functional in (5.3.29) has non-trivial summands only for ℓ such that $\pi_{\tau(\ell)} \ge \pi_{\tau(r)}$, we have shown that $(t_{\cdot,\cdot}) \in I^*_{r,m}$.

The following example illustrates this argument. Suppose we have an alphabet of size $m = 8$, with

$$(\pi_1, \pi_2, \ldots, \pi_8) = (0.07, 0.1, 0.2, 0.06, 0.2, 0.06, 0.1, 0.2).$$

Then,

$$\pi_{\tau(1)} = \pi_{\tau(2)} = \pi_{\tau(3)} = 0.2, \quad m_1 = m_2 = m_3 = 0, \quad d_1 = d_2 = d_3 = 3,$$

$$\pi_{\tau(4)} = \pi_{\tau(5)} = 0.1, \quad m_4 = m_5 = 3, \quad d_4 = d_5 = 2,$$

$$\pi_{\tau(6)} = 0.07, \quad m_6 = 5, \quad d_6 = 1,$$

$$\pi_{\tau(7)} = \pi_{\tau(8)} = 0.06, \quad m_7 = m_8 = 6, \quad d_7 = d_8 = 2.$$

In particular, note that the two largest, distinct probability values are 0.2 and 0.1, of multiplicities 3 and 2, respectively. Next, consider the case $r = 4$. We now show how $I_{r-m_r,d_r} = I_{4-3,2} = I_{1,2}$ corresponds to an element of $I^*_{r,m} = I^*_{4,8}$. Figure 6 shows a typical element of the unconstrained index set $I_{4,8}$.

Now $\tau(1) = 3, \tau(2) = 5, \tau(3) = 8, \tau(4) = 2$, and $\tau(5) = 7$. Our construction begins with the amalgamation of $m_r = m_4 = 3$ rows, corresponding to the three indices for which π_i is strictly less than $\pi_{\tau(r)} = \pi_{\tau(4)} = 0.1$, with $I_{1,2}$. This is shown in Figure 7.

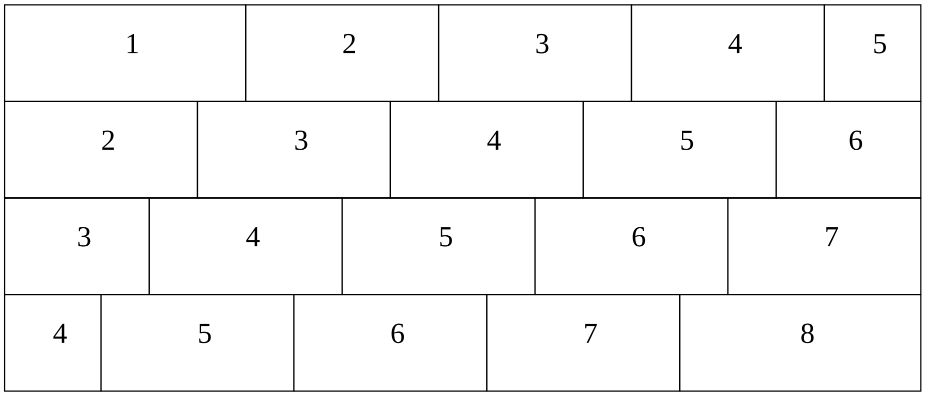

Figure 6: A typical element of $I_{4,8}$.

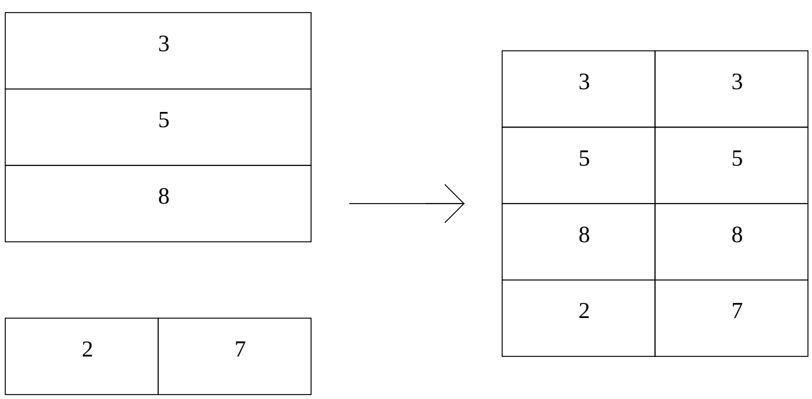

Figure 7: Amalgamating 3 rows with $I_{1,2}$.

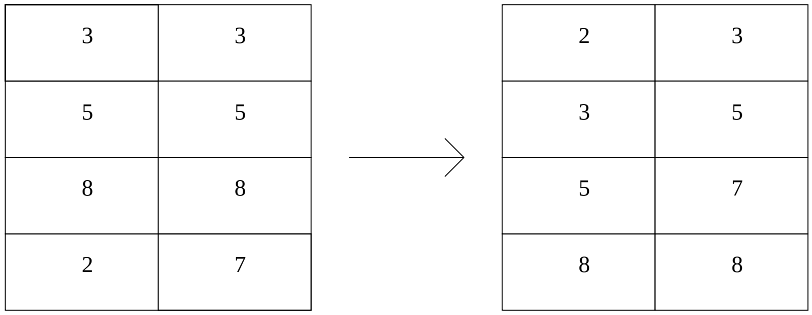

Figure 8: Reordering vertically to obtain an element in $I^*_{4,8}$.

Finally, we simply reorder each vertical column in the original order of the indices, as shown in Figure 8. We see that, first of all, we have constructed an element of $I_{4,8}$. Moreover, since we have three rows whose indices are associated with the maximum value, and a remaining row whose indices are associated with $\pi_{\tau(4)}$, we indeed have an element of $I^*_{4,8}$. Note that the $4 \times 4 = 16$ free indices in $I_{4,8}$ (corresponding to the locations of the 16 vertical bars in Figure 6) have been reduced to a *single* one in $I^*_{4,8}$.

In addition, we may essentially reverse this construction, starting with an element of $I^*_{r,m}$ ($\neq \emptyset$), and so obtain an element of I_{r-m_r,d_r}. Indeed, from the definitions of $I^*_{r,m}$ and ν_r we know that

$$\nu_r = \sum_{j=1}^{r} \pi_{\tau(j)} = \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \pi_\ell |\Delta t_{j,\ell}|,$$

for any $(t_{\cdot,\cdot}) \in I^*_{r,m}$. However, we also have

$$\begin{aligned}
\sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \pi_\ell |\Delta t_{j,\ell}| &= \mathbf{1}_{\{m_r>0\}} \Big(\sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \mathbf{1}_{\{\pi_{\tau(\ell)} \geq \pi_{\tau(m_r)}\}} \pi_\ell |\Delta t_{j,\ell}| \\
&\qquad + \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \mathbf{1}_{\{\pi_{\tau(\ell)} < \pi_{\tau(m_r)}\}} \pi_\ell |\Delta t_{j,\ell}| \Big) \\
&\qquad + \mathbf{1}_{\{m_r=0\}} \pi_{\tau(1)} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} |\Delta t_{j,\ell}| \\
&\leq \mathbf{1}_{\{m_r>0\}} \big((\pi_{\tau(1)} + \cdots + \pi_{\tau(m_r)}) + (r - m_r) \pi_{\tau(r)} \big)
\end{aligned}$$

$$+ \mathbf{1}_{\{m_r=0\}} r\pi_{\tau(1)}$$

$$= \nu_r,$$

with equality holding throughout if and only if $m_r = 0$ or $m_r > 0$ and $\sum_{j=1}^{r} |\Delta t_{j,\ell}| = 1$, for all ℓ such that $\pi_{\tau(\ell)} \geq \pi_{\tau(m_r)}$, and that, moreover, $\sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \mathbf{1}_{\{\pi_{\tau(\ell)}=\pi_{\tau(r)}\}}$ $|\Delta t_{j,\ell}| = r - m_r$. If $m_r > 0$, then, for any $(t_{\cdot,\cdot}) \in I^*_{r,m}$, we may start with (5.3.29), and use again the permutation of the indices employed there. We thus obtain the first term of (5.3.28), which corresponds to the condition $\sum_{j=1}^{r} |\Delta t_{j,\ell}| = 1$, for all ℓ such that $\pi_{\tau(\ell)} \geq \pi_{\tau(m_r)}$, and also the second term of (5.3.28), which corresponds to the other condition $\sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \mathbf{1}_{\{\pi_{\tau(\ell)}=\pi_{\tau(r)}\}}|\Delta t_{j,\ell}| = r - m_r$. If $m_r = 0$ the same reasoning holds, except that the first term in (5.3.28) is taken to be zero.

Having thus established a bijection between $I^*_{r,m}$ and I_{r-m_r,d_r}, we may thus maximize over these two parameter sets, and so, for any process $(M(t))_{t\geq 0}$, obtain the general result

$$\sum_{i=1}^{m_r} M^{\tau(i)}(1) + \max_{I_{r-m_r,d_r}} \sum_{j=1}^{(r-m_r)} \sum_{\ell-j}^{(r-m_r+d_r-1)} M^{\tau(m_r+\ell)}(\Delta u_{j,\ell})$$
$$= \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\tilde{\ell}(j,q)}(\Delta t_{j,\ell}). \tag{5.3.30}$$

We now proceed to show that (5.3.27) holds. First, fix $c > 0$, and, for each $1 \leq \ell \leq m$, set

$$c_\ell = \begin{cases} c, & \text{if } \pi_\ell < \pi_{\tau(r)}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.3.31}$$

Next, let $\widehat{M}^\ell_n(t) = \sigma_\ell \hat{B}^\ell_n(t) - c_\ell t$, and let $M^\ell(t) = \sigma_\ell \tilde{B}^\ell(t) - c_\ell t$. Then, for n large enough, namely, for $n > c/(\pi_{\tau(r)} - \pi_{\tau(r+1)})$, we have that, almost surely, for any $t_{\cdot,\cdot} \in I_{r,m}$,

$$\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j}\widehat{M}_n^{\ell}(\Delta t_{j,\ell})$$
$$\geq \sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j}\Big(\sigma_\ell \hat{B}_n^{\ell}(\Delta t_{j,\ell}) - \sqrt{n}\left(\pi_{\tau(j)}-\pi_\ell\right)|\Delta t_{j,\ell}|\Big). \tag{5.3.32}$$

Hence, almost surely, both

$$\max_{I_{r,m}}\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j}\widehat{M}_n^{\ell}(\Delta t_{j,\ell})$$
$$\geq \max_{I_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j}\Big(\sigma_\ell \hat{B}_n^{\ell}(\Delta t_{j,\ell}) - \sqrt{n}\left(\pi_{\tau(j)}-\pi_\ell\right)|\Delta t_{j,\ell}|\Big), \tag{5.3.33}$$

and

$$\max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j}\widehat{M}_n^{\ell}(\Delta t_{j,\ell}) = \max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j}\sigma_\ell \hat{B}_n^{\ell}(\Delta t_{j,\ell}). \tag{5.3.34}$$

Now choose any $z > 0$. Then

$$\mathbb{P}\bigg(\max_{I_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j}\Big(\sigma_\ell \hat{B}_n^{\ell}(\Delta t_{j,\ell}) - \sqrt{n}\left(\pi_{\tau(j)}-\pi_\ell\right)|\Delta t_{j,\ell}|\Big)$$
$$- \max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j}\sigma_\ell \hat{B}_n^{\ell}(\Delta s_{j,\ell}) > z\bigg)$$
$$\leq \mathbb{P}\left(\max_{I_{r,m}}\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j}\widehat{M}_n^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j}\widehat{M}_n^{\ell}(\Delta t_{j,\ell}) > z\right), \tag{5.3.35}$$

so that

$$\limsup_{n\to\infty}\mathbb{P}\bigg(\max_{I_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j}\Big(\sigma_\ell \hat{B}_n^{\ell}(\Delta t_{j,\ell}) - \sqrt{n}\left(\pi_{\tau(j)}-\pi_\ell\right)|\Delta t_{j,\ell}|\Big)$$
$$- \max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j}\sigma_\ell \hat{B}_n^{\ell}(\Delta s_{j,\ell}) > z\bigg)$$

$$\leq \limsup_{n\to\infty} \mathbb{P}\Big(\max_{I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \widehat{M}_n^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \widehat{M}_n^{\ell}(\Delta t_{j,\ell}) > z\Big)$$

$$= \mathbb{P}\left(\max_{I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) > z\right), \tag{5.3.36}$$

by the Invariance Principle and the Continuous Mapping Theorem. Next, for any $0 \leq \varepsilon \leq 1$, let

$$I_{r,m}(\varepsilon) = \{(t_{j,\ell}) \in I_{r,m} : \sum_{j,\ell} |\Delta t_{j,\ell}| \mathbf{1}_{\{\pi_\ell < \pi_{\tau(r)}\}} \leq \varepsilon r\}.$$

Thus, $I^*_{r,m} = I_{r,m}(0) \subset I_{r,m}(\varepsilon) \subset I_{r,m}(1) = I_{r,m}$. We bound (5.3.36) using this family of subsets as follows:

$$\begin{aligned}
&\mathbb{P}\left(\max_{I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) > z\right) \\
&\leq \mathbb{P}\left(\max_{I_{r,m}(\varepsilon)} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) > z\right) \\
&\quad + \mathbb{P}\left(\max_{I_{r,m}\setminus I_{r,m}(\varepsilon)} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) > z\right) \\
&\leq \mathbb{P}\left(\max_{I_{r,m}(\varepsilon)} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \tilde{B}^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \tilde{B}^{\ell}(\Delta s_{j,\ell}) > z\right) \\
&\quad + \mathbb{P}\left(\max_{I_{r,m}\setminus I_{r,m}(\varepsilon)} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \tilde{B}^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \tilde{B}^{\ell}(\Delta s_{j,\ell}) > z + \varepsilon rc\right) \\
&\leq \mathbb{P}\left(\max_{I_{r,m}(\varepsilon)} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \tilde{B}^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \tilde{B}^{\ell}(\Delta s_{j,\ell}) > z\right) \\
&\quad + \mathbb{P}\Big(\max_{I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \tilde{B}^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}} \sum_{j=1}^{r} \sum_{\ell=1}^{m-r+j} \tilde{B}^{\ell}(\Delta s_{j,\ell}) > z + \varepsilon rc\Big).
\end{aligned} \tag{5.3.37}$$

We can now take the limsup in (5.3.37), as $c \to \infty$, and then, as $\varepsilon \to 0$, and so establish convergence to zero in probability. Moreover, since

$$\mathbb{P}\left(\max_{I_{r,m}}\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) - \max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) \geq 0\right) = 1,$$

we have in fact shown, with the help of (5.3.36), that with probability one,

$$\max_{I_{r,m}}\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}) = \max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=1}^{m-r+j} M^{\ell}(\Delta t_{j,\ell}),$$

and thus

$$\max_{I_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j} \left(\sigma_\ell \hat{B}^{\ell}_n(\Delta t_{j,\ell}) - \sqrt{n}\left(\pi_{\tau(j)} - \pi_\ell\right)|\Delta t_{j,\ell}|\right) - \max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j} \sigma_\ell \hat{B}^{\ell}_n(\Delta s_{j,\ell}) \xrightarrow{\mathbb{P}} 0. \tag{5.3.38}$$

Since

$$\max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j} \sigma_\ell \hat{B}^{\ell}_n(\Delta s_{j,\ell}) \Rightarrow \max_{I^*_{r,m}}\sum_{j=1}^{r}\sum_{\ell=j}^{m-r+j} \sigma_\ell \tilde{B}^{\ell}(\Delta s_{j,\ell}), \tag{5.3.39}$$

by the Converging Together Lemma, we have proved (5.3.27). Equation (5.3.8) of the theorem follows from the bijection between $I^*_{r,m}$ and I_{r-m_r,d_r} described in the general result (5.3.30).

Finally, we can obtain the convergence of the joint distribution in (5.3.10) in the following manner. Given any $(\theta_1, \theta_2, \ldots, \theta_r) \in \mathbb{R}^r$, we have

$$\begin{aligned}
&\sum_{k=1}^{r}\theta_k\left(\frac{V^k_n - \nu_k n}{\sqrt{n}}\right) \\
&\quad = \sum_{k=1}^{r}\theta_k\left(\max_{I_{k,m}}\sum_{j=1}^{k}\sum_{\ell=j}^{m-k+j}\left(\sigma_\ell \hat{B}^{\ell}_n(\Delta t_{j,\ell}) - \sqrt{n}\left(\pi_{\tau(j)} - \pi_\ell\right)|\Delta t_{j,\ell}|\right)\right) \\
&\quad = \sum_{k=1}^{r}\theta_k\left(\max_{I_{k,m}}\sum_{j=1}^{k}\sum_{\ell=j}^{m-k+j}\left(\sigma_\ell \hat{B}^{\ell}_n(\Delta t_{j,\ell}) - \sqrt{n}\left(\pi_{\tau(j)} - \pi_\ell\right)|\Delta t_{j,\ell}|\right)\right. \\
&\quad\quad \left. - \max_{I^*_{k,m}}\sum_{j=1}^{k}\sum_{\ell=j}^{m-k+j}\sigma_\ell \hat{B}^{\ell}_n(\Delta s_{j,\ell})\right) + \sum_{k=1}^{r}\theta_k\left(\max_{I^*_{k,m}}\sum_{j=1}^{k}\sum_{\ell=j}^{m-k+j}\sigma_\ell \hat{B}^{\ell}_n(\Delta s_{j,\ell})\right).
\end{aligned} \tag{5.3.40}$$

Now from (5.3.38), the first summation on the right-hand side of (5.3.40) converges to zero in probability, as $n \to \infty$. Moreover, the second summation is a continuous functional of $(\hat{B}_n^1, \hat{B}_n^2, \ldots, \hat{B}_n^m)$, and so, by the Invariance Principle and Continuous Mapping Theorem, converges. Then the Converging Together Lemma, along with the bijection result (5.3.30), gives

$$\sum_{k=1}^{r} \theta_k \left(\frac{V_n^k - \nu_k n}{\sqrt{n}} \right)$$
$$\Rightarrow \sum_{k=1}^{r} \theta_k \left(\max_{I_{k,m}^*} \sum_{j=1}^{k} \sum_{\ell=j}^{m-k+j} \sigma_\ell \tilde{B}^\ell(\Delta s_{j,\ell}) \right) = \sum_{k=1}^{r} \theta_k V_\infty^k. \tag{5.3.41}$$

Since (5.3.41) holds for arbitrary $(\theta_1, \theta_2, \ldots, \theta_r) \in \mathbb{R}^r$, by the Cramér-Wold Theorem, we have the joint convergence result (5.3.10). ■

Since the shape of the Young tableau is more naturally expressed in terms of the R_n^k, rather than of the V_n^k, we may restate the results of the previous theorem as follows:

Theorem 5.3.3 *Let $(X_n)_{n \geq 0}$ be an irreducible, aperiodic, homogeneous Markov chain with finite state space $\mathcal{A}_m = \{\alpha_1 < \cdots < \alpha_m\}$, and with stationary distribution $(\pi_1, \pi_2, \ldots, \pi_m)$. Then, in the notations of Theorem 5.3.2,*

$$\left(\frac{R_n^1 - \pi_{\tau(1)} n}{\sqrt{n}}, \frac{R_n^2 - \pi_{\tau(2)} n}{\sqrt{n}}, \ldots, \frac{R_n^m - \pi_{\tau(m)} n}{\sqrt{n}} \right) \Rightarrow (R_\infty^1, R_\infty^2, \ldots, R_\infty^m), \tag{5.3.42}$$

where

$$R_\infty^1 = \max_{I_{1,d_1}} \sum_{\ell=1}^{d_1} \sigma_{\tau(\ell)} \left(\tilde{B}^{\tau(\ell)}(t_{1,\ell}) - \tilde{B}^{\tau(\ell)}(t_{1,\ell-1}) \right), \tag{5.3.43}$$

and, for each $2 \leq k \leq m$,

$$
\begin{aligned}
R_\infty^k = &\sum_{i=m_{k-1}+1}^{m_k} \sigma_{\tau(i)} \tilde{B}^{\tau(i)}(1) \\
&+ \max_{I_{k-m_k,d_k}} \sum_{j=1}^{k-m_k} \sum_{\ell=j}^{(d_k+m_k-k+j)} \sigma_{\tau(m_k+\ell)} \tilde{B}^{\tau(m_k+\ell)}(\Delta t_{j,\ell}) \\
&- \max_{I_{k-1-m_{k-1},d_{k-1}}} \sum_{j=1}^{k-1-m_{k-1}} \sum_{\ell=j}^{(d_{k-1}+m_{k-1}-k+1+j)} \sigma_{\tau(m_{k-1}+\ell)} \tilde{B}^{\tau(m_{k-1}+\ell)}(\Delta t_{j,\ell}),
\end{aligned} \tag{5.3.44}
$$

where we use the notation $\tilde{B}^s(\Delta t_{j,\ell}) = \tilde{B}^s(t_{j,\ell}) - \tilde{B}^s(t_{j,\ell-1})$, for any $1 \le s \le m$, $1 \le j \le k$, and $1 \le \ell \le m$, and where the first sum on the right-hand side of (5.3.44) *is understood to be* 0*, if $m_k = m_{k-1}$.*

Proof. First, $R_n^1 = V_n^1$, and, for each $2 \le k \le m$, $R_n^k = V_n^k - V_n^{k-1}$. Expressing these equalities at the multivariate level, we have

$$
\begin{aligned}
&\left(\frac{R_n^1 - \pi_{\tau(1)} n}{\sqrt{n}}, \frac{R_n^2 - \pi_{\tau(2)} n}{\sqrt{n}}, \dots, \frac{R_n^m - \pi_{\tau(m)} n}{\sqrt{n}} \right) \\
&\qquad = \left(\frac{V_n^1 - \pi_{\tau(1)} n}{\sqrt{n}}, \frac{V_n^2 - V_n^1 - \pi_{\tau(2)} n}{\sqrt{n}}, \dots, \frac{V_n^m - V_n^{m-1} - \pi_{\tau(m)} n}{\sqrt{n}} \right) \\
&\qquad = \left(\frac{V_n^1 - \nu_1 n}{\sqrt{n}}, \frac{V_n^2 - \nu_2 n}{\sqrt{n}}, \dots, \frac{V_n^m - \nu_m n}{\sqrt{n}} \right) \\
&\qquad\qquad - \left(0, \frac{V_n^1 - \nu_1 n}{\sqrt{n}}, \dots, \frac{V_n^m - \nu_{m-1} n}{\sqrt{n}} \right) \\
&\qquad \Rightarrow (V_\infty^1, V_\infty^2, \dots, V_\infty^m) - (0, V_\infty^1, \dots, V_\infty^m) \\
&\qquad := (R_\infty^1, R_\infty^2, \dots, R_\infty^m),
\end{aligned} \tag{5.3.45}
$$

where the weak convergence follows immediately from the Continuous Mapping Theorem, since the transformation is linear.

Equations (5.3.43) and (5.3.44) follow simply from the Brownian expressions for $(V_\infty^1, V_\infty^2, \dots, V_\infty^m)$ in Theorem 5.3.2. ∎

If all m letters have unique stationary probabilities, then we have the following corollary to Theorem 5.3.3:

Corollary 5.3.3 *If the stationary distribution of Theorem 5.3.3 is such that each π_r is unique, then*

$$\left(\frac{R_n^1 - \pi_{\tau(1)}n}{\sqrt{n}}, \frac{R_n^2 - \pi_{\tau(2)}n}{\sqrt{n}}, \ldots, \frac{R_n^m - \pi_{\tau(m)}n}{\sqrt{n}}\right) \Rightarrow N((0,0,\ldots,0),\Sigma). \qquad (5.3.46)$$

In other words, the limiting distribution is identical in law to the spectrum of the diagonal matrix $D = \operatorname{diag}\{Z_1, Z_2, \ldots, Z_m\}$, where $(Z_1, Z_2, \ldots, Z_m)$ is a centered normal random vector with covariance matrix Σ.

Proof. Now, for all $1 \le k \le m$, $d_k = 1$, and $m_k = k-1$, so that

$$\begin{aligned} R_\infty^1 &= \max_{I_{1,d_1}} \sum_{\ell=1}^{d_1} \sigma_{\tau(\ell)} \left(\tilde{B}^{\tau(\ell)}(t_{1,\ell}) - \tilde{B}^{\tau(\ell)}(t_{1,\ell-1})\right) \\ &= \sigma_{\tau(1)}\tilde{B}^{\tau(1)}(1), \end{aligned}$$

and, for each $2 \le k \le m$,

$$\begin{aligned} R_\infty^k = &\sum_{i=m_{k-1}+1}^{m_k} \sigma_{\tau(i)}\tilde{B}^{\tau(i)}(1) \\ &+ \max_{I_{k-m_k,d_k}} \sum_{j=1}^{k-m_k} \sum_{\ell=j}^{(d_k+m_k-k+j)} \sigma_{\tau(m_k+\ell)}\tilde{B}^{\tau(m_k+\ell)}(\Delta t_{j,\ell}) \\ &- \max_{I_{k-1-m_{k-1},d_{k-1}}} \sum_{j=1}^{k-1-m_{k-1}} \sum_{\ell=j}^{(d_{k-1}+m_{k-1}-k+1+j)} \sigma_{\tau(m_{k-1}+\ell)}\tilde{B}^{\tau(m_{k-1}+\ell)}(\Delta t_{j,\ell}) \\ = &\sum_{i=k-1}^{k-1} \sigma_{\tau(i)}\tilde{B}^{\tau(i)}(1) \\ &+ \max_{I_{1,1}} \sum_{j=1}^{1} \sum_{\ell=j}^{j} \sigma_{\tau(k-1+\ell)}\tilde{B}^{\tau(k-1+\ell)}(\Delta t_{j,\ell}) \end{aligned}$$

$$- \max_{I_{1,1}} \sum_{j=1}^{1} \sum_{\ell=j}^{j} \sigma_{\tau(k-2+\ell)} \tilde{B}^{\tau(k-2+\ell)}(\Delta t_{j,\ell})$$

$$= \sigma_{\tau(k-1)} \tilde{B}^{\tau(k-1)}(1) + \sigma_{\tau(k)} \tilde{B}^{\tau(k)}(1) - \sigma_{\tau(k-1)} \tilde{B}^{\tau(k-1)}(1)$$

$$= \sigma_{\tau(k)} \tilde{B}^{\tau(k)}(1).$$

Moreover, the joint law result for $(R^1_\infty, R^2_\infty, \dots, R^m_\infty)$ holds as well, and this is clearly a multivariate normal distribution, with mean $(0, 0, \dots, 0)$ and covariance matrix Σ. Since the spectrum of a diagonal matrix consists of its diagonal elements, the final claim of the corollary holds. ∎

Remark 5.3.5 *We know that the joint law of* $(R^1_\infty, R^2_\infty, \dots, R^m_\infty)$ *in the iid uniform alphabet case is identical to the joint law of the eigenvalues of an* $m \times m$ *traceless GUE matrix. Corollary 5.3.3 also gives a spectral characterization for the unique probability case, in particular, for a non-uniform iid alphabet with unique stationary probabilities. This is consistent with the characterization of the limiting law of* LI_n *in the non-uniform iid case, due to Its, Tracy, and Widom [28, 29], as that of the largest eigenvalue of the block associated with the most probable letters among a direct sum of independent GUE matrices whose dimensions correspond to the multiplicities* d_r *of Theorems 5.3.2 and 5.3.3, subject to the condition that* $\sum_{r=1}^m \sqrt{\pi_{\tau(r)}} X_r = 0$, *where* $X_1, X_2, \dots, X_m$ *are the diagonal elements of the random matrix. More generally, the joint law in Corollary 5.3.3 is a special case of the non-uniform iid result of Xu [46].*

Remark 5.3.6 *The difference between the zero-trace condition* $\sum_{r=1}^m X_r = 0$ *and the generalized traceless condition* $\sum_{r=1}^m \sqrt{\pi_{\tau(r)}} X_r = 0$ *amounts to nothing more than a difference in the choice of scaling for each row* R^r_n. *We will find it more natural to express our results using the normalization associated with the zero-trace condition* $\sum_{r=1}^m X_r = 0$

5.4 Fine Structure of the Brownian Functional

So far, we have seen that the limiting shape of the random Young tableau generated by an aperiodic, irreducible, homogeneous Markov chain can be expressed as a Brownian functional. The form of this functional is similar to the iid case; the essential difference is in the covariance structure of the Brownian motion. We begin our study of the consequences of this difference.

In the iid uniform m-alphabet case, Johansson [31] proved that the limiting shape of the Young tableau had a joint law which is that of the spectrum of an $m \times m$ traceless GUE matrix. An immediate consequence of this result is that the limiting shape of the Young tableau contains simple symmetries, *e.g.,* for each $1 \leq r \leq m$, $R_\infty^r \stackrel{\mathcal{L}}{=} -R_\infty^{m-r}$. Now, as was seen in Corollary 5.3.1 of Theorem 5.3.2, the form of the Brownian functional in the doubly stochastic case involved only the maximal term. We will see that that there is also a pleasing symmetry to the limiting shape of Young tableaux in the doubly stochastic case by examining a natural bijection between the parameter set $I_{r,m}$ and $I_{m-r,m}$, for any $1 \leq r \leq m-1$. Indeed, this result will follow as a corollary to the following, more general, theorem:

Theorem 5.4.1 *The limiting functionals of Theorem 5.3.2 enjoy the following symmetry property: for every* $1 \leq r \leq m-1$*,*

$$\begin{aligned} V_\infty^r &:= \sum_{i=1}^{m_r} \sigma_{\tau(i)} \tilde{B}^{\tau(i)}(1) \\ &\quad + \max_{t(\cdot,\cdot) \in I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(m_r+d_r-r+j)} \sigma_{\tau(m_r+\ell)} \tilde{B}^{\tau(m_r+\ell)}(\Delta t_{j,\ell}) \\ &\stackrel{\mathcal{L}}{=} \sum_{i=m_r+d_r+1}^{m} \sigma_{\tau(i)} \tilde{B}^{\tau(i)}(1) \\ &\quad + \max_{u(\cdot,\cdot) \in I_{m_r+d_r-r,d_r}} \sum_{j=1}^{m_r+d_r-r} \sum_{\ell=j}^{r-m_r+j} \sigma_{\tau(m_r+\ell)} \tilde{B}^{\tau(m_r+\ell)}(\Delta u_{j,\ell}), \end{aligned} \tag{5.4.1}$$

where $\tilde{B}^\ell(\Delta) := \tilde{B}^\ell(t) - \tilde{B}^\ell(s)$, for $\Delta = [s,t]$, and where the non-maximal terms on the left and right-hand sides of (5.4.1) *are identically zero if $m_r = 0$, or $m_r + d_r = m$, respectively.*

Remark 5.4.1 *Recall that, from the definitions of m_r and d_r, the non-maximal summation terms on the left and right-hand sides of* (5.4.1) *reflect the letters which have, respectively, greater and smaller stationary probabilities than $\pi_{\tau(r)}$. Recall, moreover, that the maximal terms are associated with the indices having the same stationary probability as $\pi_{\tau(r)}$. The maximal term on the left-hand side of* (5.4.1) *involves a summation over $r - m_r$ rows, while the one on the right-hand side involves $m_{r+1} - r$ rows. Thus, in a sense, the two maximal terms in* (5.4.1) *split $d_r = m_{r+1} - m_r$ rows between themselves. In summary, the functional on the right-hand side of* (5.4.1) *corresponds to the sum of the $m - r$ bottom rows of the Young tableau.*

Proof. Without loss of generality, we may assume that $\tau(j) = j$, for all $1 \leq j \leq m$. Fix $1 \leq r \leq m-1$, and for any point t in the index set I_{r-m_r,d_r}, define $\Delta t_{j+m_r,\ell} = [t_{j,\ell-1}, t_{j,\ell}]$, for $1 \leq j \leq r - m_r$ and $1 \leq \ell \leq d_r$. Furthermore, for each $1 \leq j \leq m_r$ or $m_{r+1} < j \leq m$, set $\Delta t_{j,\ell} = [0,1]$, for $j = \ell$, $\Delta t_{j,\ell} = \{0\}$, for $0 \leq \ell < j$, and $\Delta t_{j,\ell} = \{1\}$, for $j < \ell \leq m$. Next, as in the proof of Theorem 5.3.2, consider the *set* of points $\{t_{j,\ell}\}_{(1 \leq j \leq r-m_r, 1 \leq \ell \leq d_r)}$, and order them as $s_0 := 0 < s_1 < \cdots < s_{\kappa-1} < s_\kappa := 1$, for some integer κ, and let $\Delta s_q = [s_{q-1}, s_q]$, for each $1 \leq q \leq \kappa$.

Now, for each $1 \leq q \leq \kappa$, let A_q consist of the indices ℓ for which $\Delta s_q \cap \Delta t_{j,\ell} \neq \emptyset$. Then, almost surely,

$$\begin{aligned}
\sum_{i=1}^{m_r} \sigma_i \tilde{B}^i(1) + \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(m_r+d_r-r+j)} & \sigma_{m_r+\ell} \tilde{B}^{m_r+\ell}(\Delta t_{j,\ell}) \\
&= \sum_{j=1}^{r} \sum_{\ell=1}^{m} \sigma_\ell \tilde{B}^\ell(\Delta t_{j,\ell})
\end{aligned}$$

$$= \sum_{j=1}^{r} \sum_{q=1}^{\kappa} \sum_{\ell=1}^{m} \sigma_\ell \tilde{B}^\ell(\Delta t_{j,\ell} \cap \Delta s_q)$$
$$= \sum_{j=1}^{r} \sum_{q=1}^{\kappa} \sum_{\ell \in A_q} \sigma_\ell \tilde{B}^\ell(\Delta s_q). \tag{5.4.2}$$

Now by the "stairstep" properties of $I_{r,m}$ there are precisely r elements in each A_q. Letting $\tilde{A}_q = \{1, \ldots, m\} \setminus A_q$, for each $1 \leq q \leq \kappa$, we thus see that each $\tilde{A}_q$ contains exactly $m - r$ elements. Let $\tilde{\ell}_{j,q}$ be the j^{th} smallest element of $\tilde{A}_q$. We claim that for each $1 \leq j \leq m - r$, the sequence $\tilde{\ell}_{j,1}, \tilde{\ell}_{j,2}, \ldots, \tilde{\ell}_{j,\kappa}$. is weakly decreasing.

Indeed, fix $1 \leq j \leq m - r$ and $1 \leq q \leq \kappa - 1$, and suppose that $\tilde{\ell}_{j,q}$ is less than all the elements of A_q. Then, by the properties of $I_{r,m}$, the least element of A_{q+1} is no smaller, so that the j^{th} smallest element of $\tilde{A}_q$, $\tilde{\ell}_{j,q+1}$ is also $\tilde{\ell}_{j,q}$. Next, suppose that $\tilde{\ell}_{j,q}$ is greater than $k \geq 1$ elements of A_q. Thus, $\tilde{\ell}_{j,q} = j + k$. Then there are at most k elements of A_{q+1} which are less than or equal to $\tilde{\ell}_{j,q}$, by the properties of $I_{r,m}$. But this implies that there are at least j elements of $\tilde{A}_{q+1}$ which are less than or equal to $\tilde{\ell}_{j,q}$. Thus, $\tilde{\ell}_{j,q+1} \leq \tilde{\ell}_{j,q}$, and the claim is proved.

Moreover, since each A_q contains $\{1, 2, \ldots, m_r\}$, we see that necessarily each $\tilde{A}_q$ contains $\{m_r + d_r + 1, m_r + d_r + 2, \ldots, m\}$.

For each $1 \leq j \leq m - r$, we may now amalgamate the intervals Δs_q to obtain a partition of the unit interval. Specifically, for each $1 \leq j \leq m - r$, and each $1 \leq \ell \leq m$, let $\tilde{u}_{j,\ell}$ be the smallest s_q such that $\tilde{\ell}_{j,q+1} \leq \ell$. (We define $\tilde{u}_{j,0} = 1$, for all $1 \leq j \leq m - r$.)

Finally, and most crucially, recall that $\sum_{\ell=1}^{m} \sigma_\ell \tilde{B}^\ell(t) = 0$, for all t. Then since $(\tilde{B}^1, \tilde{B}^2, \ldots, \tilde{B}^m) \stackrel{\mathcal{L}}{=} (-\tilde{B}^1, -\tilde{B}^2, \ldots, -\tilde{B}^m)$,

$$\sum_{j=1}^{r} \sum_{q=1}^{\kappa} \sum_{\ell \in A_q} \sigma_\ell \tilde{B}^\ell(\Delta s_q)$$

$$
\begin{aligned}
&= \sum_{j=1}^{m-r} \sum_{q=1}^{\kappa} \sum_{\ell \in \tilde{A}_q} \left(-\sigma_\ell \tilde{B}^\ell(\Delta s_q)\right) \\
&= - \sum_{i=m_r+d_r+1}^{m} \sigma_i \tilde{B}^i(1) - \sum_{j=1}^{m_r+d_r-r} \sum_{\ell=1}^{m} \sigma_{m_r+\ell} \tilde{B}^{m_r+\ell}(\Delta u_{j,\ell}) \\
&\stackrel{\mathcal{L}}{=} \sum_{i=m_r+d_r+1}^{m} \sigma_i \tilde{B}^i(1) + \sum_{j=1}^{m_r+d_r-r} \sum_{\ell=1}^{m} \sigma_{m_r+\ell} \tilde{B}^{m_r+\ell}(\Delta u_{j,\ell}),
\end{aligned} \tag{5.4.3}
$$

where $\Delta u_{j,\ell} = [u_{j,\ell-1}, u_{j,\ell}]$. But, by the way we ordered each A_q, we must have $\Delta u_{j_1,\ell} \cap \Delta u_{j_2,\ell} = \emptyset$, for any $j_1 \neq j_2$. Thus, $u \in I_{m_r+d_r-r,d_r}$, and so we may restrict the summation over ℓ in (5.4.3) to $\ell = j, \ldots, r - m_r + j$, since the remaining terms are zero. Equation (5.4.1) follows immediately by taking the maxima over I_{r-m_r,d_r} and $I_{m_r+d_r-r,d_r}$ over the left-hand and right-hand sides, respectively, of (5.4.3). ■

For doubly stochastic transition matrices, the symmetry is even more apparent:

Corollary 5.4.1 *Let the transition matrix P of Theorem 5.3.2 be doubly stochastic. Then, for every $1 \leq r \leq m-1$,*

$$
\begin{aligned}
V_\infty^r &:= \max_{t(\cdot,\cdot) \in I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{m-r+j} \sigma_\ell \left(\tilde{B}^\ell(t_{j,\ell}) - \tilde{B}^\ell(t_{j,\ell-1})\right) \\
&\stackrel{\mathcal{L}}{=} \max_{u(\cdot,\cdot) \in I_{m-r,m}} \sum_{j=1}^{m-r} \sum_{\ell=j}^{r+j} \sigma_\ell \left(\tilde{B}^\ell(u_{j,\ell}) - \tilde{B}^\ell(u_{j,\ell-1})\right) := V_\infty^{m-r},
\end{aligned} \tag{5.4.4}
$$

and so

$$
\lim_{n\to\infty} \frac{\sum_{j=1}^{r} R_n^j - rn/m}{\sqrt{n}} \stackrel{\mathcal{L}}{=} \lim_{n\to\infty} \frac{rn/m - \sum_{j=m-r+1}^{m} R_n^j}{\sqrt{n}}. \tag{5.4.5}
$$

Moreover,

$$
(V_\infty^1, \ldots, V_\infty^r) \stackrel{\mathcal{L}}{=} (V_\infty^{m-1}, \ldots, V_\infty^{m-r}). \tag{5.4.6}
$$

Proof. Since $m_r = 0$ and $d_r = m$ for all $1 \le r \le m$, the non-maximal terms on both sides of (5.4.1) disappear, and we have (5.4.4).

To prove (5.4.5), recall that $V_n^m = \sum_{j=1}^m R_n^j = n$. Then, from the result just proved,

$$\begin{aligned}\frac{V_n^{m-r} - (m-r)n/m}{\sqrt{n}} &= \frac{\sum_{j=1}^{m-r} R_n^j - (m-r)n/m}{\sqrt{n}} \\ &= \frac{\left(n - \sum_{j=m-r+1}^{m} R_n^j\right) - (m-r)n/m}{\sqrt{n}} \\ &= \frac{rn/m - \sum_{j=m-r+1}^{m} R_n^j}{\sqrt{n}} \\ &\Rightarrow V_\infty^{m-r} \stackrel{\mathcal{L}}{=} V_\infty^r, \end{aligned} \tag{5.4.7}$$

and we have established the claimed symmetry.

Finally, the extension of (5.4.4) to (5.4.6) follows from a standard Cramér-Wold argument. ■

Remark 5.4.2 *Since $R_\infty^m = -V_\infty^{m-1}$, almost surely, Corollary 5.4.1 states that $R_\infty^m \stackrel{\mathcal{L}}{=} -R_\infty^1$. From the symmetry of the Brownian motion, we thus see that R_∞^m may be represented as a minimal Brownian functional:*

$$R_\infty^m = \min_{I_{1,m}} \sum_{\ell=1}^{m} \sigma_\ell \left(\tilde{B}^\ell(t_{1,\ell}) - \tilde{B}^\ell(t_{1,\ell-1})\right).$$

Turning again to the cyclic case, recall that, for $m \ge 4$, the limiting shape of the Young tableau in general differs from that of the iid uniform case. The following theorem characterizes the asymptotic covariance matrices of such Markov chains.

Theorem 5.4.2 *Let P be the $m \times m$ transition matrix of an aperiodic, irreducible, cyclic Markov chain on an m-letter, ordered alphabet, $\mathcal{A}_m = \{\alpha_1 < \alpha_2 < \cdots < \alpha_m\}$, with*

$$P = \begin{pmatrix} a_1 & a_m & \cdots & a_3 & a_2 \\ a_2 & a_1 & \ddots & & a_3 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ a_{m-1} & & \ddots & a_1 & a_m \\ a_m & a_{m-1} & \cdots & a_2 & a_1 \end{pmatrix}. \tag{5.4.8}$$

Then, for $1 \leq j \leq m$, $\lambda_j = \sum_{k=1}^{m} a_k \omega^{(k-1)(j-1)}$ *is an eigenvalue of* P, *where* $\omega = exp(2\pi i/m)$ *is the* m^{th} *principal root of unity. Moreover, letting* $\gamma_j = \lambda_j/(1-\lambda_j)$, *for* $2 \leq j \leq m$, *and* $\beta_j = \cos(2\pi j/m)$, *for* $0 \leq j \leq m$, *the asymptotic covariance matrix* Σ *is given by*

$$\Sigma = \frac{m-1}{m^2} M^{(1)} + \frac{4}{m^2} \sum_{j=2}^{m_0+1} Re(\gamma_j) M^{(j)}, \qquad m = 2m_0 + 1, \tag{5.4.9}$$

and

$$\Sigma = \frac{m-1}{m^2} M^{(1)} + \frac{4}{m^2} \sum_{j=2}^{m_0} Re(\gamma_j) M^{(j)} + \frac{2}{m^2} \gamma_{m_0+1} M^{(m_0+1)}, \qquad m = 2m_0, \tag{5.4.10}$$

where $M^{(j)}$ *is an* $m \times m$ *Toeplitz matrix with entries* $(M^{(j)})_{k,\ell} = \beta_{(j-1)|k-\ell|}$, *for* $2 \leq j \leq m$, *and* $(M^{(1)})_{k,\ell} = \delta_{k,\ell} - (1-\delta_{k,\ell})/(m-1)$, *for* $j = 1$.

Proof. It is straightforward, and classical, to verify that, for each $1 \leq j \leq m$, $(1, \omega^{j-1}, \omega^{2(j-1)}, \ldots, \omega^{(m-1)(j-1)})$ is a left eigenvector of P, with eigenvalue $\lambda_j = \sum_{k=1}^{m} a_k \omega^{(k-1)(j-1)}$. We can thus write our standard diagonalization of P as $P = S^{-1}\Lambda S$, where $\Lambda = \mathrm{diag}(1, \lambda_2, \ldots, \lambda_m)$,

$$
S = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{m-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(m-1)} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & \omega^{m-1} & \omega^{2(m-1)} & \cdots & \omega^{(m-1)^2} \end{pmatrix}, \tag{5.4.11}
$$

and

$$
S^{-1} = \frac{1}{m} \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \cdots & \omega^{-(m-1)} \\ 1 & \omega^{-2} & \omega^{-4} & \cdots & \omega^{-2(m-1)} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & \omega^{-(m-1)} & \omega^{-2(m-1)} & \cdots & \omega^{-(m-1)^2} \end{pmatrix}. \tag{5.4.12}
$$

In the present cyclic, and hence, doubly stochastic case, we know that $\Sigma = (1/m)(I + S^{-1}DS + (S^{-1}DS)^T)$, where, as usual, we have $D = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_m)$ $= \mathrm{diag}(-1/2, \lambda_2/(1-\lambda_2), \ldots, \lambda_m/(1-\lambda_m))$. We can then compute the entries of the covariance matrix Σ $S^{-1}DS$ as follows:

$$
\begin{aligned}
(S^{-1}DS)_{j_1,j_2} &= \sum_{k,\ell} (S^{-1})_{j_1,k} (D)_{k,\ell} (S)_{\ell,j_2} \\
&= \sum_{k,\ell} \frac{1}{m} (\omega^{-(j_1-1)(k-1)})(\delta_{k,\ell}\gamma_k)(\omega^{(j_2-1)(\ell-1)}) \\
&= \sum_{k=1}^{m} \frac{\gamma_k}{m} \omega^{(j_2-j_1)(k-1)} \\
&= \frac{1}{m} \left(-\frac{1}{2} + \sum_{k=2}^{m} \gamma_k \omega^{(j_2-j_1)(k-1)} \right),
\end{aligned} \tag{5.4.13}
$$

for all $1 \leq j_1, j_2, \leq m$. The entries of the asymptotic covariance matrix can thus be written as

$$
\begin{aligned}
\sigma_{j_1,j_2} &= \frac{1}{m}\left(\delta_{j_1,j_2} + (S^{-1}DS)_{j_1,j_2} + (S^{-1}DS)_{j_2,j_1}\right) \\
&= \frac{1}{m}\left(\delta_{j_1,j_2} + \frac{1}{m}\left(-1 + \sum_{k=2}^{m}\gamma_k\big(\omega^{(j_2-j_1)(k-1)} + \omega^{(j_1-j_2)(k-1)}\big)\right)\right) \\
&= \frac{m-1}{m^2}M^{(1)}_{j_1,j_2} + \frac{2}{m^2}\sum_{k=2}^{m}\gamma_k\beta_{|j_2-j_1|(k-1)}, \qquad (5.4.14)
\end{aligned}
$$

for all $1 \leq j_1, j_2, \leq m$.

Next, note that since $\lambda_{m+2-k} = \bar{\lambda_k}$, *i.e.*, the complex conjugate of λ_{m+2-k}, we have $\gamma_{m+2-k} = \bar{\gamma}_k$, for all $2 \leq k \leq m$. Moreover, since $\beta_{|j_2-j_1|(k-1)} = \beta_{|j_2-j_1|((m+2-k)-1)}$, we can write (5.4.14) more symmetrically as (5.4.9) or (5.4.10), depending on whether m is odd or even, respectively, and in the latter case, we also use that γ_{m_0+1} is real, since $\omega^{m_0} = -1$. ■

Let us again examine the cases $m = 3$ and $m = 4$. In the former case, we have

$$
M^{(1)} = \begin{pmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{pmatrix}.
$$

But for $m = 3$, $\beta_1 = -1/2 = \beta_2$, and so $M^{(1)} = M^{(2)}$. Hence

$$
\Sigma = \frac{2}{9}M^{(1)} + \frac{4}{9}Re(\gamma_2)M^{(2)} = \frac{2}{9}(1 + 2Re(\gamma_2))M^{(1)}. \qquad (5.4.15)
$$

Hence, for $m = 3$, cyclicity *always* produces a rescaled version of the uniform iid case, with the rescaling factor given by $1 + 2Re(\gamma_2)$.

For $m = 4$, however,

$$
M^{(1)} = \begin{pmatrix} 1 & -1/3 & -1/3 & -1/3 \\ -1/3 & 1 & -1/3 & -1/3 \\ -1/3 & -1/3 & 1 & -1/3 \\ -1/3 & -1/3 & -1/3 & 1 \end{pmatrix},
$$

and $\beta_1 = 0$, $\beta_2 = -1$, and $\beta_3 = 0$. Thus,

$$M^{(2)} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix},$$

and

$$M^{(3)} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix}.$$

In this case, we have

$$\Sigma = \frac{3}{16}M^{(1)} + \frac{4}{16}Re(\gamma_2)M^{(2)} + \frac{2}{16}\gamma_3 M^{(3)}.$$

Next, note that $2M^{(2)} + M^{(3)} = 3M^{(1)}$. Then, if $Re(\gamma_2) = \gamma_3$,

$$\begin{aligned}\Sigma &= \frac{3}{16}M^{(1)} + \frac{4}{16}Re(\gamma_2)M^{(2)} + \frac{2}{16}\gamma_3 M^{(3)} \\ &= \frac{3}{16}M^{(1)} + \frac{2}{16}(2Re(\gamma_2)M^{(1)}) \\ &= \frac{3}{16}(1 + 2Re(\gamma_2))M^{(1)}, \end{aligned} \tag{5.4.16}$$

so that there is still a rescaled version of the iid case in a non-iid cyclic setting. Indeed, since we know that $\lambda_2 = a_1 + ia_2 - a_3 - ia_4 = (a_1 - a_3) + i(a_2 - a_4)$ and $\lambda_3 = a_1 - a_2 + a_3 - a_4$, we find that

$$Re(\gamma_2) = \frac{1 - a_2 - 2a_3 - a_4}{(a_2 + 2a_3 + a_4)^2 + (a_2 - a_4)^2} - 1,$$

and $\gamma_3 = 1/(2(a_2 + a_4)) - 1$. A short calculation then shows that $Re(\gamma_2) = \gamma_3$ if and only if $a_3^2 = a_2 a_4$. We thus have a complete characterization of all 4-letter, cyclic

Markov chains whose Young tableaux have the same limiting shape as the uniform iid case. In particular, choosing $a_2 = a_4 = a$, for some $0 < a < 1/3$, leads to $a_3 = a$ and $a_1 = 1 - 3a$. If, moreover, $a = 1/4$, we have again the iid uniform case. For $a \neq 1/4$, however, we may view the Markov chain as a "lazy" version of the uniform iid case.

Note that the scaling factor in both (5.4.15) and (5.4.16) is $1 + 2Re(\gamma_2)$. The following theorem shows that, in fact, such a scaling factor occurs for general m, and gives a spectral characterization of all transition matrices which lead to an iid limiting shape.

Theorem 5.4.3 *Let P be the $m \times m$ transition matrix of an aperiodic, irreducible, cyclic Markov chain on an m-letter, ordered alphabet given in Theorem 5.4.2. Then the asymptotic covariance matrix Σ is a rescaled version of the iid uniform covariance matrix $\Sigma_{iidu} := ((m-1)/m^2)M^{(1)}$ if and only if the constants $\gamma_j = \lambda_j/(1-\lambda_j)$, for $2 \leq j \leq m$, satisfy the condition*

$$Re(\gamma_j) = \gamma, \qquad \text{for all } 2 \leq j \leq m, \tag{5.4.17}$$

for some real constant γ. Moreover, the scaling is then given by

$$\Sigma = (1 + 2\gamma)\Sigma_{iidu}. \tag{5.4.18}$$

Proof. We first claim that the system of matrix equations

$$\sum_{j=2}^{m} b_j M^{(j)} = M^{(1)} \tag{5.4.19}$$

has a unique solution $b_j = 1/(m-1)$, for all $2 \leq j \leq m$. Indeed, revisiting (5.4.14), we can express each $M^{(j)}$ as

$$
\begin{aligned}
M^{(j)} &= \tilde{M}^{(j)} + \tilde{M}^{(-j)} \\
&= \tilde{M}^{(j)} + \tilde{M}^{(m-j+1)},
\end{aligned}
\tag{5.4.20}
$$

where $(\tilde{M}^{(j)})_{k,\ell} = \omega^{(j-1)(\ell-k)}/2$, for all $1 \le k, \ell \le m$, so that (5.4.19) becomes

$$
\begin{aligned}
M^{(1)} &= \sum_{j=2}^{m} b_j \left(\tilde{M}^{(j)} + \tilde{M}^{(m-j+1)} \right) \\
&= \sum_{j=2}^{m} (b_j + b_{m-j+1}) \tilde{M}^{(j)} \\
&= \sum_{j=2}^{m} \tilde{b}_j \tilde{M}^{(j)},
\end{aligned}
\tag{5.4.21}
$$

where $\tilde{b}_j := (b_j + b_{m-j+1})/2$, for $2 \le j \le m$.

Now, clearly, each $\tilde{M}^{(j)}$ is cyclic, so that in solving (5.4.21) we need only examine the m entries in the first rows of the matrices. We can thus reduce (5.4.21) to the $m \times (m-1)$ system of equations

$$
\begin{pmatrix}
1 & 1 & 1 & \cdots & 1 \\
\omega & \omega^2 & \omega^3 & \cdots & \omega^{m-1} \\
\omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(m-1)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\omega^{m-1} & \omega^{2(m-1)} & \omega^{3(m-1)} & \cdots & \omega^{(m-1)^2}
\end{pmatrix}
\begin{pmatrix}
\tilde{b}_2 \\ \tilde{b}_3 \\ \vdots \\ \tilde{b}_m
\end{pmatrix}
=
\begin{pmatrix}
1 \\ \frac{-1}{m-1} \\ \frac{-1}{m-1} \\ \vdots \\ \frac{-1}{m-1}
\end{pmatrix}.
\tag{5.4.22}
$$

Since each of the last $m-1$ rows of the matrix in (5.4.22) sums to -1, it is clear that $\tilde{b}_j = 1/(m-1)$ is a solution to the system. To see that this solution is, in fact, unique, consider the $(m-1) \times (m-1)$ sub-matrix consisting of the last $m-1$ rows of the matrix in (5.4.22), namely,

$$\begin{pmatrix} \omega & \omega^2 & \omega^3 & \cdots & \omega^{m-1} \\ \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(m-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega^{m-1} & \omega^{2(m-1)} & \omega^{3(m-1)} & \cdots & \omega^{(m-1)^2} \end{pmatrix}. \tag{5.4.23}$$

Now this matrix, which is very closely related to the Fourier matrix which arises in discrete Fourier transform problems, is in fact invertible, and can be shown to have one eigenvalue of -1, and $m-2$ eigenvalues of the form $\pm\sqrt{m}$ and $\pm i\sqrt{m}$, so that the modulus of the determinant is $m^{(m-2)/2} \neq 0$. Thus, the solution $\tilde{b}_j = 1/(m-1)$ is unique, and since $b_j = (b_j + b_{m-j+1})/2 = b_{m-j+1}$, for all $2 \leq j \leq m$, we conclude that $b_j = 1/(m-1)$ as well, for all $2 \leq j \leq m$, and the claim is proved.

We can now use Theorem 5.4.2 to simplify the asymptotic covariance matrix decomposition as follows:

$$\begin{aligned} \Sigma &= \frac{m-1}{m^2} M^{(1)} + \frac{2}{m^2} \sum_{k=2}^{m} \gamma_k M^{(k)} \\ &= \frac{m-1}{m^2} M^{(1)} + 2\gamma \frac{1}{m^2} \sum_{k=2}^{m} M^{(k)} \\ &= \frac{m-1}{m^2} M^{(1)} + 2\gamma \frac{m-1}{m^2} M^{(1)} \\ &= (1+2\gamma) \frac{m-1}{m^2} M^{(1)} \\ &= (1+2\gamma)\Sigma_{iidu}, \end{aligned} \tag{5.4.24}$$

where $\gamma = Re(\gamma_j)$, for all $2 \leq j \leq m$. If the real parts of γ_j are not all identical, then the uniqueness of the solution of (5.4.19) implies that no such simplification is possible, and the theorem is proved. ■

Remark 5.4.3 *To see that the condition in* (5.4.17) *is not vacuous for any* m*, recall that for* $m = 4$*, the "lazy" chain has the iid limiting shape. This is true for general*

m: if $a_2 = a_3 = \cdots = a_m = a$, for some $0 < a < 1/(m-1)$, then $\lambda_j = 1 - (m-1)a$, for all $2 \le j \le m$. Trivially, then, $\gamma_j = 1/((m-1)a) - 1 := \gamma$, for all $2 \le j \le m$, so that the conditions of Theorem 5.4.3 are satisfied, and the scaling factor is given by $1 + 2\gamma = (2 - (m-1)a)/((m-1)a)$. Even in the $m = 4$ case, however, we saw that there were other, more general, cyclic transition matrices which gave rise to the iid limiting distribution.

The previous theorem indicates precisely when we may expect the limiting shape of a cyclic Markov chain to be identical to that of the iid uniform case. Now the first-order behavior of all rows of the Young tableau is $n/m + O(\sqrt{n})$ for cyclic Markov chains. Although this differs from the first-order behavior in the non-uniform iid case, one may still ask whether the limiting shape for some cyclic Markov chains might still be that of some non-uniform iid case. In fact, this can never occur: cyclicity ensures that the asymptotic covariance matrix is also cyclic, and thus cannot be equal to the asymptotic covariance matrix of any non-uniform iid case.

Still, we may ask how to relate the iid non-uniform limiting shape to that of a general Markov chain having the same stationary distribution. The following interpolation result describes the asymptotic covariance matrix for a Markov chain whose transition matrix is a convex combination of an iid (uniform or non-uniform) transition matrix and another arbitrary transition matrix having the same stationary distribution:

Theorem 5.4.4 *For any $m \ge 3$, let P_0 be the $m \times m$ transition matrix of an irreducible, aperiodic, homogeneous Markov chain, and let its associated asymptotic covariance matrix be given by*

$$\Sigma_0 = \Pi_0 + \Pi_0(S_0^{-1} D_0 S_0) + (S_0^{-1} D_0 S_0)^T \Pi_0, \tag{5.4.25}$$

in the standard notations of Theorem 5.2.1. Then, for $0 < \delta \le 1$, the transition *matrix* $P = (1-\delta)I_m + \delta P_0$ *has an asymptotic covariance matrix given by*

$$\Sigma = \frac{1}{\delta}\left(\Sigma_0 + (1-\delta)\Sigma_{\Pi_0}\right), \tag{5.4.26}$$

where Σ_{Π_0} *is the covariance matrix associated with the iid Markov chain having the same stationary distribution as* P_0*.*

Proof. Using the standard notations of Theorem 5.2.1, we will write

$$\Sigma = \Pi + \Pi(S^{-1}DS) + (S^{-1}DS)^T\Pi$$

in terms of the decomposition Σ_0 in (5.4.25). Now, clearly, the stationary distribution under P is that of P_0, so that $\Pi = \Pi_0$. We will thus write the stationary distribution simply as $(\pi_1, \pi_2, \ldots, \pi_m)$. Moreover, the eigenvectors are also unchanged, so that $S = S_0$. However, for each eigenvalue $\lambda_{k,0}$ of P_0, we have that $\lambda_k = (1-\delta) + \delta\lambda_{k,0}$ is an eigenvalue of P, for $1 \le k \le m$. Thus, for each $2 \le k \le m$, the diagonal entries of D are given by

$$\begin{aligned}
\gamma_k &:= \frac{\lambda_k}{1-\lambda_k} \\
&= \frac{(1-\delta) + \delta\lambda_{k,0}}{\delta(1-\lambda_{k,0})} \\
&= \frac{1-\delta}{\delta} + \gamma_{k,0},
\end{aligned}$$

where $\gamma_{k,0}$ are the diagonal entries of D_0. We can thus decompose D as follows:

$$\begin{aligned}
D &= \mathrm{diag}(-1/2, \gamma_2, \ldots, \gamma_m) \\
&= \mathrm{diag}(-1/2, 0, \ldots, 0) + \left(\frac{1-\delta}{\delta}\right)\mathrm{diag}(0, 1, \ldots, 1) \\
&\quad + \left(\frac{1}{\delta}\right)\mathrm{diag}(0, \gamma_{2,0}, \ldots, \gamma_{m,0})
\end{aligned}$$

$$= \mathrm{diag}\left(-\left(\frac{1-\delta}{2\delta}\right), 0, \ldots, 0\right) + \left(\frac{1-\delta}{\delta}\right) I_m + \left(\frac{1}{\delta}\right) D_0. \tag{5.4.27}$$

Next, recall from Proposition 5.2.1 that the first column of S^{-1} is $(1, 1, \ldots, 1)^T$. Hence,

$$\begin{aligned}
S^{-1}DS &= S_0^{-1}DS_0 \\
&= \begin{pmatrix} 1 & * & \cdots & * \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 1 & * & \cdots & * \end{pmatrix} \begin{pmatrix} -\frac{1-\delta}{2\delta} & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_m \\ * & * & \cdots & * \\ \vdots & \vdots & \cdots & \vdots \\ * & * & \cdots & * \end{pmatrix} \\
&\quad + \left(\frac{1-\delta}{\delta}\right) S_0^{-1} I_m S_0 + \left(\frac{1}{\delta}\right) S_0^{-1} D_0 S_0 \\
&= -\left(\frac{1-\delta}{2\delta}\right) \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_m \\ \pi_1 & \pi_2 & \cdots & \pi_m \\ \vdots & \vdots & \cdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_m \end{pmatrix} + \left(\frac{1-\delta}{\delta}\right) I_m + \left(\frac{1}{\delta}\right) S_0^{-1} D_0 S_0,
\end{aligned} \tag{5.4.28}$$

which gives us

$$\begin{aligned}
\Pi S^{-1}DS &= \Pi_0 S^{-1}DS \\
&= -\left(\frac{1-\delta}{2\delta}\right) \begin{pmatrix} \pi_1 & 0 & \cdots & 0 \\ 0 & \pi_2 & \cdots & \vdots \\ 0 & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \pi_m \end{pmatrix} \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_m \\ \pi_1 & \pi_2 & \cdots & \pi_m \\ \vdots & \vdots & \cdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_m \end{pmatrix} \\
&\quad + \left(\frac{1-\delta}{\delta}\right) \Pi_0 + \left(\frac{1}{\delta}\right) \Pi_0 S_0^{-1} D_0 S_0
\end{aligned}$$

$$= -\left(\frac{1-\delta}{2\delta}\right)\begin{pmatrix} \pi_1^2 & \pi_1\pi_2 & \cdots & \pi_1\pi_m \\ \pi_2\pi_1 & \pi_2^2 & \cdots & \pi_2\pi_m \\ \vdots & \vdots & \ddots & \vdots \\ \pi_m\pi_1 & \pi_m\pi_2 & \cdots & \pi_m^2 \end{pmatrix}$$

$$+\left(\frac{1-\delta}{\delta}\right)\Pi_0 + \left(\frac{1}{\delta}\right)\Pi_0 S_0^{-1} D_0 S_0. \tag{5.4.29}$$

Finally, we can express Σ as

$$\begin{aligned}
\Sigma &= \Pi + \Pi(S^{-1}DS) + (S^{-1}DS)^T\Pi \\
&= \left(\frac{1}{\delta}\right)\Pi_0 + \left(1-\frac{1}{\delta}\right)\Pi_0 + \Pi_0(S^{-1}DS) + (\Pi_0(S^{-1}DS))^T \\
&= \left(\frac{1}{\delta}\right)\Sigma_0 + \left(1-\frac{1}{\delta}\right)(\Pi_0 - 2\Pi_0) \\
&\qquad + \left(1-\frac{1}{\delta}\right)\begin{pmatrix} \pi_1^2 & \pi_1\pi_2 & \cdots & \pi_1\pi_m \\ \pi_2\pi_1 & \pi_2^2 & \cdots & \pi_2\pi_m \\ \vdots & \vdots & \ddots & \vdots \\ \pi_m\pi_1 & \pi_m\pi_2 & \cdots & \pi_m^2 \end{pmatrix} \\
&= \left(\frac{1}{\delta}\right)\Sigma_0 + \left(1-\frac{1}{\delta}\right)(-\Sigma_{\Pi_0}) \\
&= \frac{1}{\delta}\left(\Sigma_0 + (1-\delta)\Sigma_{\Pi_0}\right),
\end{aligned} \tag{5.4.30}$$

and we are done. ■

Thus far we have expressed our limiting laws in terms of Brownian functionals whose Brownian motions have a non-trivial covariance structure arising directly from the specific nature of the transition matrix. It is of interest to instead express the limiting laws in terms of *standard* Brownian motions.

Since the asymptotic covariance matrix Σ is non-negative definite, we can find an $m \times m$ matrix C such that $\Sigma = CC^T$. (The matrix C is not unique, since $(CQ)(CQ)^T = CC^T = \Sigma$ for any orthogonal matrix Q.) Clearly, we then have

$$(\sigma_1 \tilde{B}^1(t), \sigma_2 \tilde{B}^2(t), \dots, \sigma_m \tilde{B}^m(t))^T = C(B^1(t), B^2(t), \dots, B^m(t))^T, \tag{5.4.31}$$

where $(B^1(t), B^2(t), \dots, B^m(t))^T$ is a standard, m-dimensional Brownian motion, since

$$\begin{aligned}
&\mathbb{E}\big[(\sigma_1 \tilde{B}^1(t), \sigma_2 \tilde{B}^2(t), \dots, \sigma_m \tilde{B}^m(t))^T (\sigma_1 \tilde{B}^1(t), \sigma_2 \tilde{B}^2(t), \dots, \sigma_m \tilde{B}^m(t))\big] \\
&\quad = \mathbb{E}\big[C(B^1(t), B^2(t), \dots, B^m(t))^T\big]\big[(C(B^1(t), B^2(t), \dots, B^m(t))^T\big]^T \\
&\quad = C\big[\mathbb{E}(B^1(t), B^2(t), \dots, B^m(t))^T)(B^1(t), B^2(t), \dots, B^m(t))\big]C^T \\
&\quad = C(tI_m)C^T \\
&\quad = t\Sigma.
\end{aligned}$$

Next, we can, without loss of generality, assume that $\tau(\ell) = \ell$, for all ℓ, and so write our main result (5.3.8) in Theorem 5.3.2 as

$$\begin{aligned}
\frac{V_n^r - \nu_r n}{\sqrt{n}} &\Rightarrow \sum_{k=1}^{m_r} \sigma_k \tilde{B}^k(1) + \max_{I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(d_r+m_r-r+j)} \sigma_{m_r+\ell} \tilde{B}^{m_r+\ell}(\Delta t_{j,\ell}) \\
&:= V_\infty^r.
\end{aligned} \tag{5.4.32}$$

Simply substituting (5.4.31) into (5.4.32) immediately yields

$$
\begin{aligned}
V_\infty^r &= \sum_{k=1}^{m_r}\left(\sum_{i=1}^{m} C_{k,i}B^i(1)\right)\\
&\quad + \max_{I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)}\left(\sum_{i=1}^{m} C_{m_r+\ell,i}B^i(\Delta t_{j,\ell})\right)\\
&= \sum_{i=1}^{m}\sum_{k=1}^{m_r} C_{k,i}B^i(1)\\
&\quad + \max_{I_{r-m_r,d_r}} \sum_{i=1}^{m}\sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)} C_{m_r+\ell,i}B^i(\Delta t_{j,\ell}). \qquad (5.4.33)
\end{aligned}
$$

Now the first term in (5.4.33) is simply a Gaussian term whose variance can be computed explicitly. Unfortunately, the maximal term does not in general succumb to any significant simplifications. However, in the iid case, we can further simplify (5.4.33) in a very satisfying way.

Indeed, since, in the iid case, we have $\sigma_k^2 = \pi_k(1-\pi_k)$ and, for $k \neq \ell$, $\sigma_{k,\ell} = -\pi_k\pi_\ell$, one can quickly check that C can be chosen so that $C_{k,k} = \sqrt{\pi_k} - \sqrt{\pi_k}\pi_k$, and, for $k \neq \ell$, $C_{k,\ell} = -\sqrt{\pi_\ell}\pi_k$. Moreover, for all $m_r + 1 \leq k \leq m_r + d_r$, $\pi_k = \pi_{m_r+1} = \pi_r$. Then, within the maximal term, $C_{m_r+\ell,i} = \sqrt{\pi_r} - \pi_r\sqrt{\pi_r}$, for $i = m_r + \ell$, and $C_{m_r+\ell,i} = -\pi_r\sqrt{\pi_i}$, for $i \neq m_r + \ell$. With the convention that $\nu_0 = 0$, we can then express (5.4.33) as

$$
\begin{aligned}
V_\infty^r &= \sum_{i=1}^{m_r}\sqrt{\pi_i}B^i(1) + \sum_{i=1}^{m}\sum_{k=1}^{m_r}(-\sqrt{\pi_i}\pi_k)B^i(1)\\
&\quad + \max_{I_{r-m_r,d_r}}\left\{\sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)}\sqrt{\pi_r}B^{m_r+\ell}(\Delta t_{j,\ell})\right.\\
&\qquad \left. + \sum_{i=1}^{m}\sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)}(-\pi_r\sqrt{\pi_i})B^i(\Delta t_{j,\ell})\right\}\\
&= \sum_{i=1}^{m_r}\sqrt{\pi_i}B^i(1) - \sum_{i=1}^{m}\sqrt{\pi_i}B^i(1)\sum_{k=1}^{m_r}\pi_k\\
&\quad + \sqrt{\pi_r}\max_{I_{r-m_r,d_r}}\left\{\sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)} B^{m_r+\ell}(\Delta t_{j,\ell})\right.
\end{aligned}
$$

$$
\begin{aligned}
&\qquad - \sqrt{\pi_r} \sum_{i=1}^{m} \sqrt{\pi_i} \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(d_r+m_r-r+j)} B^i(\Delta t_{j,\ell}) \Big\} \\
&= \Big\{ \sum_{i=1}^{m_r} \sqrt{\pi_i} B^i(1) - \nu_{m_r} \sum_{i=1}^{m} \sqrt{\pi_i} B^i(1) - \pi_r r \sum_{i=1}^{m} \sqrt{\pi_i} B^i(1) \Big\} \\
&\qquad + \sqrt{\pi_r} \max_{I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(d_r+m_r-r+j)} B^{m_r+\ell}(\Delta t_{j,\ell}) \\
&= \Big\{ \sum_{i=1}^{m_r} \sqrt{\pi_i} B^i(1) - (\nu_{m_r} + \pi_r r) \sum_{i=1}^{m} \sqrt{\pi_i} B^i(1) \Big\} \\
&\quad + \sqrt{\pi_r} \max_{I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(d_r+m_r-r+j)} B^{m_r+\ell}(\Delta t_{j,\ell}) \\
&= \Big\{ (1 - \nu_{m_r} - \pi_r r) \sum_{i=1}^{m_r} \sqrt{\pi_i} B^i(1) \\
&\qquad - (\nu_{m_r} + \pi_r r) \sum_{i=m_r+d_r+1}^{m} \sqrt{\pi_i} B^i(1) \Big\} \\
&\quad + \sqrt{\pi_r} \Big\{ -(\nu_{m_r} + \pi_r r) \sum_{i=m_r+1}^{m_r+d_r} B^i(1) \\
&\qquad + \max_{I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(d_r+m_r-r+j)} B^{m_r+\ell}(\Delta t_{j,\ell}) \Big\}. \qquad (5.4.34)
\end{aligned}
$$

Note that the first two Gaussian term of (5.4.34) are independent of the remaining two Gaussian-maximal expression terms.

Following Glynn and Whitt[20] and Barishnykov[6], who studied the Brownian functional

$$D_m = \max_{I_{1,m}} \sum_{\ell=1}^{m} B^\ell(\Delta t_\ell),$$

we define the following, more general, Brownian functional:

$$D_{r,m} := \max_{I_{r,m}} \sum_{j=1}^{r} \sum_{\ell=j}^{(m-r+j)} B^\ell(\Delta t_{j,\ell}), \qquad (5.4.35)$$

where $1 \leq r \leq m$. Clearly, the maximal term in (5.4.34) has just such a form. We also remark that $D_{r,m}$ corresponds to the sum of the r largest eigenvalues of an $m \times m$ GUE matrix.

To better understand (5.4.34), we may, without much loss in generality, focus on the first block, that is, values of r such that $m_r = 0$. The first Gaussian term of (5.4.34) thus vanishes, and, writing π_{max} for π_r, we have

$$\begin{aligned} V_\infty^r = & -r\pi_{max} \sum_{i=d_1+1}^{m} \sqrt{\pi_i} B^i(1) \\ & + \sqrt{\pi_{max}} \left(-r\pi_{max} \sum_{i=1}^{d_1} B^i(1) + D_{r,d_r} \right). \end{aligned} \tag{5.4.36}$$

In the uniform iid case, the first Gaussian term of (5.4.36) itself vanishes, since $d_r = d_1 = m$, and we have

$$\begin{aligned} V_\infty^r &= \frac{1}{\sqrt{m}} \left(-\frac{r}{m} \sum_{i=1}^{m} B^i(1) + D_{r,m} \right) \\ &:= \frac{H_{r,m}}{\sqrt{m}}. \end{aligned} \tag{5.4.37}$$

For $r = 1$, this result corresponds to Theorem 2.3.1. Furthermore, and still specializing (5.4.36) to $r = 1$,

$$\begin{aligned} \frac{LI_n - \pi_{max} n}{\sqrt{n}} \Rightarrow & -\pi_{max} \sum_{i=d_1+1}^{m} \sqrt{\pi_i} B^i(1) \\ & + \sqrt{\pi_{max}} \left(-\pi_{max} \sum_{i=1}^{d_1} B^i(1) + D_{1,d_1} \right) \\ = & -\pi_{max} \sum_{i=d_1+1}^{m} \sqrt{\pi_i} B^i(1) \\ & + \sqrt{\pi_{max}} \left(\frac{1}{d_1} - \pi_{max} \right) \sum_{i=1}^{d_1} B^i(1) \end{aligned}$$

$$+\sqrt{\pi_{max}}H_{1,d_1}. \tag{5.4.38}$$

One can easily compute the variance of the Gaussian terms in (5.4.38) to be $\pi_{max}(1-d_1\pi_{max})/d_1$, which is consistent with Proposition 2.3.1.

The iid development above suggests that we can find additional cases which yield simple functionals of standard Brownian motions. Indeed, the first property of the matrix C in the iid case that allowed the functionals to be simplified was that $C_{k,\ell} = c_\ell$, for all $k \neq \ell, m_r+1 \leq k \leq m_r+d_r$, and $1 \leq \ell \leq m$, where $c_1, c_2, \ldots, c_m$ were real numbers. Then, writing the diagonal terms of C as $C_{k,k} = b_k + c_k$, for $m_r+1 \leq k \leq m_r+d_r$, we may revisit (5.4.33), and write

$$\begin{aligned}
V_\infty^r &= \sum_{i=1}^{m}\sum_{k=1}^{m_r} C_{k,i}B^i(1) + \max_{I_{r-m_r,d_r}} \sum_{i=1}^{m}\sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)} C_{m_r+\ell,i}B^i(\Delta t_{j,\ell}) \\
&= \sum_{i=1}^{m}\sum_{k=1}^{m_r} C_{k,i}B^i(1) + \max_{I_{r-m_r,d_r}} \left\{ \sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)} b_{m_r+\ell}B^{m_r+\ell}(\Delta t_{j,\ell}) \right. \\
&\qquad \left. + \sum_{i=1}^{m}\sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)} c_i B^i(\Delta t_{j,\ell}) \right\} \\
&= \sum_{i=1}^{m}\sum_{k=1}^{m_r} C_{k,i}B^i(1) + r\sum_{i=1}^{m} c_i B^i(1) \\
&\qquad + \max_{I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r}\sum_{\ell=j}^{(d_r+m_r-r+j)} b_{m_r+\ell}B^{m_r+\ell}(\Delta t_{j,\ell})
\end{aligned} \tag{5.4.39}$$

Except for the fact that we have written the functional in terms of standard Brownian motions, the maximal term in (5.4.39) is no simpler than that of our original functional. However, the second property of the iid case that yielded further simplifications was that $b_k = b$, for all $m_r+1 \leq k \leq m_r+d_r$. In this case, (5.4.39) becomes

$$V_\infty^r = \sum_{i=1}^{m} \sum_{k=1}^{m_r} C_{k,i} B^i(1) + r \sum_{i=1}^{m} c_i B^i(1)$$
$$+ b \max_{I_{r-m_r,d_r}} \sum_{j=1}^{r-m_r} \sum_{\ell=j}^{(d_r+m_r-r+j)} B^{m_r+\ell}(\Delta t_{j,\ell}) \tag{5.4.40}$$

Again, by focusing on the first block, we no longer have the initial Gaussian term, and (5.4.40) becomes

$$\begin{aligned}
V_\infty^r &= r \sum_{i=d_1+1}^{m} c_i B^i(1) \\
&\quad + r \sum_{i=1}^{d_1} c_i B^i(1) + b \max_{I_{r,d_1}} \sum_{j=1}^{r} \sum_{\ell=j}^{(d_1-r+j)} B^{\ell}(\Delta t_{j,\ell}) \\
&= r \sum_{i=d_1+1}^{m} c_i B^i(1) + \left(r \sum_{i=1}^{d_1} c_i B^i(1) + b D_{r,d_1} \right) \\
&= r \sum_{i=d_1+1}^{m} c_i B^i(1) + r \sum_{i=1}^{d_1} \left(c_i + \frac{b}{d_1} \right) B^i(1) \\
&\quad + b \left(-\frac{r}{d_1} \sum_{i=1}^{d_1} B^i(1) + D_{r,d_1} \right) \\
&= r \sum_{i=d_1+1}^{m} c_i B^i(1) + r \sum_{i=1}^{d_1} \left(c_i + \frac{b}{d_1} \right) B^i(1) + b H_{r,d_1}.
\end{aligned} \tag{5.4.41}$$

We restate these results in the following theorem:

Theorem 5.4.5 *Assume, without loss of generality, that $\tau(\ell) = \ell$, for all $1 \leq \ell \leq m$, in the notations of Theorem 5.3.2. Moreover, let the asymptotic covariance matrix be given by $\Sigma = CC^T$, where C is an $m \times m$ matrix whose first d_1 rows are given by*

$$\begin{cases} C_{k,\ell} = c_\ell, & k \neq \ell, 1 \leq k \leq d_1, 1 \leq \ell \leq m \\ C_{k,k} = b + c_k, & 1 \leq k \leq d_1, \end{cases} \tag{5.4.42}$$

for some real constants $c_1, c_2, \ldots, c_m$ and b. Then, for $1 \leq r \leq d_1$,

$$V_\infty^r = r \sum_{i=d_1+1}^{m} c_i B^i(1) + r \sum_{i=1}^{d_1} \left(c_i + \frac{b}{d_1} \right) B^i(1) + bH_{r,d_1}, \tag{5.4.43}$$

where H_{r,d_1} is the maximal functional

$$H_{r,d_1} := \frac{1}{\sqrt{d_1}} \left(-\frac{r}{d_1} \sum_{i=1}^{d_1} B^i(1) + \max_{I_{r,d_1}} \sum_{j=1}^{r} \sum_{\ell=j}^{(d_1-r+j)} B^\ell(\Delta t_{j,\ell}) \right).$$

Remark 5.4.4 *One can generalize Theorem 5.4.5 to non-initial blocks (i.e., to $r > d_1$) by extending the conditions in* (5.4.42) *to non-initial blocks and then applying the theorem to $V_\infty^r - V_\infty^{m_r}$.*

To better understand which asymptotic covariance matrices Σ can be decomposed in this manner, the conditions $C_{k,\ell} = c_\ell$, for all $k \neq \ell, 1 \leq k \leq d_1, 1 \leq \ell \leq m$, and $b_k = b$, for all $1 \leq k \leq d_1$, imply that

$$\sigma_k^2 = b^2 + 2bc_k + \sum_{i=1}^{m} c_i^2, \tag{5.4.44}$$

for $1 \leq k \leq d_1$, and

$$\sigma_{k,\ell} = bc_k + bc_\ell + \sum_{i=1}^{m} c_i^2, \tag{5.4.45}$$

for $1 \leq k < \ell \leq d_1$.

If we let $(Z_1, Z_2, \ldots, Z_m)$ be a centered Gaussian random vector with covariance matrix Σ, then (5.4.44) and (5.4.45) give us

$$\begin{aligned} \mathbb{E}(Z_k - Z_\ell)^2 &= \sigma_k^2 - 2\sigma_{k,\ell} + \sigma_\ell^2 \\ &= 2b^2, \end{aligned} \tag{5.4.46}$$

for all $1 \leq k < \ell \leq d_1$. That is, the L^2-distance between any pair (Z_k, Z_ℓ) is the same, for $1 \leq k < \ell \leq d_1$.

Notice that if $\sigma_k^2 = \sigma^2$, for all $1 \leq k \leq d_1$, then in fact (5.4.46) implies that $\rho_{k,\ell} = \sigma_{k,\ell}/\sigma_k\sigma_\ell = 1 - b^2/\sigma^2 := \rho$, for all $1 \leq k < \ell \leq d_1$. That is, the $d_1 \times d_1$ submatrix of Σ must be permutation-symmetric.

Next, we note that, for $1 \leq k < \ell \leq d_1$,

$$\sigma_k^2 - \sigma_\ell^2 = 2b(c_k - c_\ell), \tag{5.4.47}$$

so that $c_k = \sigma_k^2/(2b) + c_0$, for some constant c_0. Substituting this expression into (5.4.44) and, writing $\Gamma = \sum_{i=d_1+1}^m c_i^2$, we obtain

$$\begin{aligned}
\sigma_k^2 &= b^2 + 2b\left(\frac{\sigma_k^2}{2b} + c_0\right) + \sum_{i=1}^{d_1}\left(\frac{\sigma_i^2}{2b} + c_0\right)^2 + \Gamma \\
&= b^2 + \sigma_k^2 + 2bc_0 + \sum_{i=1}^{d_1}\left(\frac{\sigma_i^2}{2b} + c_0\right)^2 + \Gamma.
\end{aligned} \tag{5.4.48}$$

Writing $\overline{\sigma^r} = (\sum_{i=1}^{d_1}\sigma_i^r)/d_1$, for any $r > 0$, (5.4.48) gives us

$$\begin{aligned}
b^2 &+ 2bc_0 + \sum_{i=1}^{d_1}\left(\frac{\sigma_i^2}{2b} + c_0\right)^2 + \Gamma \\
&= d_1 c_0^2 + \left(2b + \frac{d_1}{b}\overline{\sigma^2}\right)c_0 + \left(b^2 + \frac{d_1\overline{\sigma^4}}{4b^2} + \Gamma\right) \\
&= 0.
\end{aligned} \tag{5.4.49}$$

In order for c_0 to be a real number, the discriminant of the quadratic equation in (5.4.49) must satisfy

$$\left(2b + \frac{d_1}{b}\overline{\sigma^2}\right)^2 - 4d_1\left(b^2 + \frac{d_1\overline{\sigma^4}}{4b^2} + \Gamma\right) \geq 0, \tag{5.4.50}$$

which leads to the inequality

$$(d_1 - 1)b^4 - d_1(\overline{\sigma^2} - \Gamma) + \frac{d_1^2}{4}\left(\overline{\sigma^4} - \left(\overline{\sigma^2}\right)^2\right) \leq 0. \tag{5.4.51}$$

This inequality, in turn, gives us constraints on b^2. Indeed, the necessary and sufficient condition needed for such a b^2 to exist is given by examining the quadratic in b^2 in (5.4.51) at its extremal point, namely, at $b^2 = d_1(\overline{\sigma^2} - \Gamma))/(2(d_1 - 1))$. Doing so leads to the condition

$$-\left(\frac{d_1^2(\overline{\sigma^2} - \Gamma)^2}{4(d_1 - 1)}\right) + \frac{d_1^2}{4}\left(\overline{\sigma^4} - \left(\overline{\sigma^2}\right)^2\right) \leq 0, \tag{5.4.52}$$

or simply,

$$\overline{\sigma^4} - \left(\overline{\sigma^2}\right)^2 \leq \frac{(\overline{\sigma^2} - \Gamma)^2}{d_1 - 1}, \tag{5.4.53}$$

since $\overline{\sigma^4} \geq \left(\overline{\sigma^2}\right)^2$. The closer that Γ is to $\overline{\sigma^2}$, the more similar that the d_1 variances must be. Thus, (5.4.53) functions as a bound on the variability among these d_1 variances. Provided that the variances satisfy (5.4.53), the condition on b^2 is given by

$$\begin{aligned} b^2 \in \Bigg(&\frac{d_1}{2(d_1 - 1)}\left\{(\overline{\sigma^2} - \Gamma) - \sqrt{(\overline{\sigma^2} - \Gamma)^2 - (d_1 - 1)\left(\overline{\sigma^4} - \left(\overline{\sigma^2}\right)^2\right)}\right\}, \\ &\frac{d_1}{2(d_1 - 1)}\left\{(\overline{\sigma^2} - \Gamma) + \sqrt{(\overline{\sigma^2} - \Gamma)^2 - (d_1 - 1)\left(\overline{\sigma^4} - \left(\overline{\sigma^2}\right)^2\right)}\right\}\Bigg). \end{aligned} \tag{5.4.54}$$

Now consider the doubly stochastic case, where $d_1 = m$. Applying the general fact that each row of Σ must necessarily sum to zero, we use (5.4.44) and (5.4.45) to find that, for each $1 \leq k \leq m$,

$$\begin{aligned}
\sum_{\ell=1}^{m} \sigma_{k,\ell} &= \sigma_k^2 + \sum_{\ell \neq k} \sigma_{k,\ell} \\
&= \left(b^2 + 2bc_k + \sum_{i=1}^{m} c_i^2 \right) + \sum_{\ell \neq k} \left(bc_k + bc_\ell + \sum_{i=1}^{m} c_i^2 \right) \\
&= b^2 + b \left(mc_k + \sum_{\ell=1}^{m} c_\ell \right) + m \sum_{i=1}^{m} c_i^2 \\
&= 0, \qquad (5.4.55)
\end{aligned}$$

so that $c_k = c \in \mathbb{R}$, for all $1 \leq k \leq m$. Substituting c back into (5.4.55) gives us

$$\begin{aligned}
\sum_{\ell=1}^{m} \sigma_{k,\ell} &= b^2 + b(mc + mc) + m(mc^2) \\
&= (b + mc)^2 = 0, \qquad (5.4.56)
\end{aligned}$$

so that $b = -mc$. This then implies that $\sigma_k^2 = m(m-1)c^2$ and $\sigma_{k,\ell} = -mc^2$, for all $1 \leq k \leq m$, $\ell \neq k$. But this is precisely a permutation-symmetric covariance matrix, which in the iid case corresponds to the class of Markov chains having a uniform stationary distribution.

We summarize these results in the following:

Theorem 5.4.6 *In order that the asymptotic covariance matrix* Σ *have a decomposition* $\Sigma = CC^T$*, where*

$$\begin{cases} C_{k,\ell} = c_\ell, & k \neq \ell, \quad 1 \leq k \leq d_1, \quad 1 \leq \ell \leq m, \\ C_{k,k} = b + c_k, & 1 \leq k \leq d_1, \end{cases} \qquad (5.4.57)$$

for some real constants $c_1, c_2, \ldots, c_m$ *and* b*, it is necessary and sufficient that*

$$\overline{\sigma^4} - \left(\overline{\sigma^2}\right)^2 \leq \frac{(\overline{\sigma^2} - \Gamma)^2}{d_1 - 1}, \qquad (5.4.58)$$

where $\Gamma = \sum_{i=d_1+1}^{m} c_i^2$, *and* $\overline{\sigma^r} = (\sum_{i=1}^{d_1} \sigma_i^r)/d_1$, *for any* $r > 0$. *In this case,*

$$b^2 \in \left(\frac{d_1}{2(d_1-1)} \left\{ (\overline{\sigma^2} - \Gamma) - \sqrt{(\overline{\sigma^2} - \Gamma)^2 - (d_1 - 1)\left(\overline{\sigma^4} - \left(\overline{\sigma^2}\right)^2\right)} \right\}, \right.$$
$$\left. \frac{d_1}{2(d_1-1)} \left\{ (\overline{\sigma^2} - \Gamma) + \sqrt{(\overline{\sigma^2} - \Gamma)^2 - (d_1 - 1)\left(\overline{\sigma^4} - \left(\overline{\sigma^2}\right)^2\right)} \right\} \right). \qquad (5.4.59)$$

In particular, if $d_1 = m$, *the asymptotic covariance matrix must be permutation-symmetric, with* $c_k = c$, *for all* k, *and* $b = -mc$, *so that the common variance is* $m(m-1)c^2$ *and the common covariances are all* $-mc^2$.

5.5 Connections to Random Matrix Theory

For iid uniform m-letter alphabets, the limiting law of the Young tableau corresponds to the joint distribution of the eigenvalues of an $m \times m$ matrix from the traceless GUE [31]. In the non-uniform iid case, we further noted that Xu [46] has extended the first-row results of Its, Tracy, and Widom [28, 29] to that of the entire Young tableau by described the limiting shape as that of the joint distribution of the eigenvalues of a random matrix consisting of independent diagonal blocks, each of which is a matrix from the GUE. The size of each block depends upon the multiplicity of the corresponding stationary probability. In addition, there is a zero-trace condition involving the stationary probabilities on the composite matrix.

As a first step in extending these connections between Brownian functionals and spectra of random matrices, recall the general case when the stationary probabilities are all distinct (see Remark 5.3.5). Our Brownian functionals then have no true maximal terms, so that the limiting shape, $(R_\infty^1, R_\infty^2, \ldots, R_\infty^m)$ is simply multivariate normal, with covariance matrix Σ (or, more precisely, the matrix obtained by permuting the rows and columns of Σ using τ, the permutation of $\{1, 2, \ldots, m\}$ previously defined). Trivially, this limiting law corresponds to the spectrum of a diagonal matrix whose elements are multivariate normal with the same covariance matrix Σ.

We can see that this general result is consistent with the non-uniform iid case having distinct probabilities. Indeed, each block is of size 1, and is rescaled so that the variance is $\pi_{\tau(i)}(1-\pi_{\tau(i)})$, for $1 \le i \le m$. Because of this rescaling, instead of having a generalized zero-trace condition, as in the non-rescaled matrices used in [28, 29], our condition is rather a true zero-trace condition. This zero-trace condition is clear, since the covariance matrix for *any* iid case (uniform and non-uniform alike) is that of a multinomial distribution with parameters $(n = 1; \pi_{\tau(1)}, \pi_{\tau(2)}, \ldots, \pi_{\tau(m)})$, and any $(Y_1, Y_2, \ldots, Y_m)$ having such a distribution of course satisfies $\sum_{i=1}^m Y_i = 1$, so that $\mathrm{Var}(\sum_{i=1}^m Y_i) = 0$, which implies the zero-trace condition for $(R_\infty^1, R_\infty^2, \ldots, R_\infty^m)$.

Next, consider the case when each stationary probability has multiplicity no greater than 2. One may conjecture that the limiting shape $(R_\infty^1, R_\infty^2, \ldots, R_\infty^m)$ is that of the spectrum of a direct sum of certain 2×2 and/or 1×1 random matrices. Specifically, let $\kappa \le m$ be the number of distinct probabilities among the stationary distributions. Then the composite matrix consists of a direct sum of κ GUE matrices which are as follows. First, the overall diagonal $(X_1, X_2, \ldots, X_m)$ of the matrix has a $N(0, \Sigma)$ distribution. Next, if $d_r = 1$, then the GUE matrix is simply the 1×1 matrix (X_r). Finally, if $d_r = 2$, then the GUE matrix is the 2×2 matrix

$$\begin{pmatrix} X_{m_r+1} & Y_{m_r+1} + iZ_{m_r+1} \\ Y_{m_r+1} - iZ_{m_r+1} & X_{m_r+2} \end{pmatrix},$$

whose off-diagonal random variables Y_{m_r+1} and Z_{m_r+1} are iid, centered, normal random variables, independent of all other random variables in the overall matrix, with variance

$$(\sigma_{m_r+1}^2 - 2\rho_{m_r+1,m_r+2}\sigma_{m_r+1}\sigma_{m_r+2} + \sigma_{m_r+2}^2)/4.$$

If such a conjecture were true, it would imply the following, more modest marginal result regarding a single block of such a matrix, which without loss of generality we take to be the first block. Specifically, if $d_1 = 2$ and $\tau(r) = r$, for all $1 \le r \le m$, we claim that $(R^1_\infty, R^2_\infty) = (V^1_\infty, V^2_\infty - V^1_\infty)$ is distributed as the spectrum (λ_1, λ_2) of the 2×2 Gaussian Hermitian matrix

$$A_1 := \begin{pmatrix} X_1 & Y_1 + iZ_1 \\ Y_1 - iZ_1 & X_2 \end{pmatrix}, \tag{5.5.1}$$

where $\lambda_1 \ge \lambda_2$. Equivalently, we will show that (V^1_∞, V^2_∞) is distributed as $(\lambda_1, \lambda_1 + \lambda_2)$.

Let the 2×2 submatrix Σ_2 of Σ be written as

$$\Sigma_2 = \begin{pmatrix} \tilde{\sigma}_1^2 & \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 \\ \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 & \tilde{\sigma}_2^2 \end{pmatrix}. \tag{5.5.2}$$

Then

$$\begin{aligned}
(V^1_\infty, V^2_\infty) &= \Big(\max_{0 \le t \le 1} \big(\tilde{\sigma}_1 \tilde{B}^1(t) + \tilde{\sigma}_2 \tilde{B}^2(1) - \tilde{\sigma}_2 \tilde{B}^2(t) \big), \\
&\qquad \tilde{\sigma}_1 \tilde{B}^1(1) + \tilde{\sigma}_1 \tilde{B}^2(1) \Big) \\
&= \Big(\tilde{\sigma}_2 \tilde{B}^2(1) + \max_{0 \le t \le 1} \big(\tilde{\sigma}_1 \tilde{B}^1(t) - \tilde{\sigma}_2 \tilde{B}^2(t) \big), \\
&\qquad \tilde{\sigma}_1 \tilde{B}^1(1) + \tilde{\sigma}_1 \tilde{B}^2(1) \Big).
\end{aligned} \tag{5.5.3}$$

We simplify (5.5.3), by introducing new Brownian motions and then decomposing the resulting expression into two independent parts. To do so, begin by defining the new variances and correlation coefficients $\sigma_1^2 := \tilde{\sigma}_2^2$, $\sigma_2^2 := \tilde{\sigma}_1^2 - 2\tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 + \tilde{\sigma}_2^2$, and $\rho := (\tilde{\rho}\tilde{\sigma}_1 - \tilde{\sigma}_2)/\sqrt{\tilde{\sigma}_1^2 - 2\tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 + \tilde{\sigma}_2^2}$. Then it is easily verified that $B^1(t) := \tilde{B}^2(t)$, and $B^2(t) := (\tilde{\sigma}_1 \tilde{B}^1(t) - \tilde{\sigma}_2 \tilde{B}^2(t))/\sigma_2$ are (dependent) standard Brownian motions, and (5.5.3) becomes

$$
\begin{aligned}
(V_\infty^1, V_\infty^2) &= \big(\sigma_1 B^1(1) + \sigma_2 \max_{0\le t\le 1} B^2(t), 2\sigma_1 B^1(1) + \sigma_2 B^2(1)\big) \\
&= \Big((\sigma_1 B^1(1) - \rho\sigma_1 B^2(1)) + \sigma_2\Big(\rho\frac{\sigma_1}{\sigma_2} + \max_{0\le t\le 1} B^2(t)\Big), \\
&\qquad 2(\sigma_1 B^1(1) - \rho\sigma_1 B^2(1)) + (\sigma_2 + 2\rho\sigma_1)B^2(1)\Big). \qquad (5.5.4)
\end{aligned}
$$

Note that $B^1(t) - \rho B^2(t)$ is independent of $B^2(t)$ and has variance $\sigma_1^2(1-\rho^2)$. Introducing the Brownian functional

$$
U(\beta) = \left(\beta - \frac{1}{2}\right) B^2(1) + \max_{0\le t\le 1} B^2(t), \qquad (5.5.5)
$$

$\beta \in \mathbb{R}$, and using σ_1^2, σ_2^2, and ρ above, (5.5.4) becomes

$$
\begin{aligned}
(V_\infty^1, V_\infty^2) &\stackrel{\mathcal{L}}{=} \sigma_1\sqrt{1-\rho^2}Z(1,2) + \left(\sigma_2 U\left(\frac{1}{2} - \rho\frac{\sigma_1}{\sigma_2}\right), (\sigma_2 + 2\rho\sigma_1)B^2(1)\right) \\
&= \frac{\tilde\sigma_1\tilde\sigma_2\sqrt{1-\tilde\rho^2}}{\sqrt{\tilde\sigma_1^2 - 2\tilde\rho\tilde\sigma_1\tilde\sigma_2 + \tilde\sigma_2^2}} Z(1,2) \\
&\quad + \Bigg(\sqrt{\tilde\sigma_1^2 - 2\tilde\rho\tilde\sigma_1\tilde\sigma_2 + \tilde\sigma_2^2} \quad U\Bigg(\frac{\tilde\sigma_1^2 - \tilde\sigma_2^2}{2\sqrt{\tilde\sigma_1^2 - 2\tilde\rho\tilde\sigma_1\tilde\sigma_2 + \tilde\sigma_2^2}}\Bigg), \\
&\qquad 2(\tilde\sigma_1^2 - \tilde\sigma_2^2)B^2(1)\Bigg), \qquad (5.5.6)
\end{aligned}
$$

where Z is a standard normal random variable independent of the sigma-field generated by B^2.

Turning now to the eigenvalues' distributions, we first consider the centered, multivariate normal random variables (W_1, W_2), having covariance matrix

$$
\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},
$$

and let W_3 and W_4 be two iid, centered, normal random variables, independent of (W_1, W_2), with variance σ_2^2. Then it is classical that

$$\left(W_2, \sqrt{W_2^2 + W_3^2 + W_4^2}\right) \stackrel{\mathcal{L}}{=} \sigma_2\big(B(1), 2 \max_{0 \le t \le 1} B(t) - B(1)\big),$$

or, equivalently,

$$\left(W_2, \beta W_2 + \frac{1}{2}\sqrt{W_2^2 + W_3^2 + W_4^2}\right) \stackrel{\mathcal{L}}{=} \sigma_2(B(1), U(\beta)), \tag{5.5.7}$$

where B is a standard Brownian motion, and $U(\beta)$, $\beta \in \mathbb{R}$, is defined in terms of B, rather than in terms of B^2, as in (5.5.5). Then consider the random variable

$$\begin{aligned}\tilde{\lambda} &:= W_1 + \sqrt{W_2^2 + W_3^2 + W_4^2} \\ &= \left(W_1 - \rho\frac{\sigma_1}{\sigma_2}\right) + \left(\rho\frac{\sigma_1}{\sigma_2} + \sqrt{W_2^2 + W_3^2 + W_4^2}\right).\end{aligned} \tag{5.5.8}$$

Using (5.5.7), and noting that the variance of the first term in (5.5.8) is $\sigma_1^2(1-\rho^2)$, it is easy to see that

$$\tilde{\lambda} \stackrel{\mathcal{L}}{=} \sigma_1\sqrt{1-\rho^2}Z + 2\sigma_2 U\left(\frac{\rho\sigma_1}{2\sigma_2}\right), \tag{5.5.9}$$

where Z is a standard normal random variable independent of B.

We now apply this result to the eigenvalues of the matrix A_1 in (5.5.1), namely, to

$$\lambda_1 = \left(\frac{X_1 + X_2}{2}\right) + \sqrt{\left(\frac{X_1 - X_2}{2}\right) + Y_1^2 + Z_1^2}, \tag{5.5.10}$$

and

$$\lambda_2 = \left(\frac{X_1 + X_2}{2}\right) - \sqrt{\left(\frac{X_1 - X_2}{2}\right) + Y_1^2 + Z_1^2}. \tag{5.5.11}$$

Letting $W_1 = (X_1 + X_2)/2$, $W_2 = (X_1 - X_2)/2$, $W_3 = Y_1$, and $W_4 = Z_1$, we have

$$
\begin{aligned}
(\lambda_1, \lambda_1 + \lambda_2) &= \left(W_1 + \sqrt{W_2^2 + W_3^2 + W_4^2}, 2W_1 \right) \\
&= \Bigg(\left(W_1 - \hat{\rho}\frac{\hat{\sigma}_1}{\hat{\sigma}_2} W_2 \right) + 2\left(\hat{\rho}\frac{\hat{\sigma}_1}{2\hat{\sigma}_2} W_2 + \frac{1}{2}\sqrt{W_2^2 + W_3^2 + W_4^2} \right), \\
&\qquad\quad 2\left(W_1 - \hat{\rho}\frac{\hat{\sigma}_1}{\hat{\sigma}_2} W_2 \right) + 2\hat{\rho}\frac{\hat{\sigma}_1}{\hat{\sigma}_2} W_2 \Bigg) \\
&= \left(W_1 - \hat{\rho}\frac{\hat{\sigma}_1}{\hat{\sigma}_2} W_2 \right)(1,2) \\
&\qquad + \left(\hat{\rho}\frac{\hat{\sigma}_1}{2\hat{\sigma}_2} W_2 + \frac{1}{2}\sqrt{W_2^2 + W_3^2 + W_4^2}, 2\hat{\rho}\frac{\hat{\sigma}_1}{\hat{\sigma}_2} W_2 \right),
\end{aligned} \tag{5.5.12}
$$

where $\hat{\sigma}_1^2 = (\tilde{\sigma}_1^2 + 2\tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 + \tilde{\sigma}_2^2)/4$, $\hat{\sigma}_2^2 = (\tilde{\sigma}_1^2 - 2\tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 + \tilde{\sigma}_2^2)/4$, and $\hat{\rho}\hat{\sigma}_1^2\hat{\sigma}_2^2 = (\tilde{\sigma}_1^2 - \tilde{\sigma}_2^2)/4$. Noting that the variance of $W_1 - (\hat{\rho}\hat{\sigma}_1/\hat{\sigma}_2)W_2$ is $\hat{\sigma}_1^2(1-\hat{\rho}^2) = \sigma_1^2(1-\rho^2)$, and that, moreover, $\beta := \hat{\rho}\hat{\sigma}_1/2\hat{\sigma}_2 = (\tilde{\sigma}_1^2 - \tilde{\sigma}_2^2)/(2\sqrt{\tilde{\sigma}_1^2 - 2\tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 + \tilde{\sigma}_2^2})$, we find that

$$
\begin{aligned}
(\lambda_1, \lambda_1 + \lambda_2) &= \hat{\sigma}_1\sqrt{1-\hat{\rho}^2}\, Z(1,2) + \left(2\hat{\sigma}_2 U\left(\frac{\hat{\rho}\hat{\sigma}_1}{2\sigma_2} \right), 2\hat{\rho}\frac{\hat{\sigma}_1}{\hat{\sigma}_2} B^2(1) \right) \\
&= \sigma_1\sqrt{1-\rho^2}\, Z(1,2) + \sigma_2\big(U(\beta), 4\beta B^2(1) \big) \\
&\stackrel{\mathcal{L}}{=} (V_\infty^1, V_\infty^2),
\end{aligned} \tag{5.5.13}
$$

and we have our identity in law.

To illustrate the ways in which additional random matrix interpretations might potentially illuminate other, apparently unrelated, Brownian functionals, consider the following example. Let $(\varepsilon_k)_{k\geq 1}$ be a sequence of positive numbers decreasing to zero. Then it is possible to find an increasing sequence of integers $(m_k)_{k\geq 1}$ so that, for each k, there is a Markov chain on m_k letters such that:

- the maximal stationary probability $\pi_{max}(k)$ is of multiplicity 3, and
- the 3×3 covariance submatrix $\Sigma_3(k)$ governing the associated Brownian functional $V_\infty^1(k)$ is of the form

$$\Sigma_3(k) = \sigma(k)^2 \begin{pmatrix} \varepsilon_k^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \varepsilon_k^2 \end{pmatrix}. \tag{5.5.14}$$

That is, the variance of $B^{\tau(2)}$ becomes arbitrarily large in comparison to that of $B^{\tau(1)}$ and $B^{\tau(3)}$.

Then, since $LI_n(k) = V_n^1(k)$, we have, as $n \to \infty$,

$$\begin{aligned} \frac{LI_n(k) - \pi_{max}(k)}{\sqrt{n}} &\Rightarrow \max_{I_{1,3}} \sum_{\ell=1}^{3} \sigma_{\tau(\ell)} B^{\tau(\ell)}(\Delta t_\ell) \\ &= \sigma(k) \max_{I_{1,3}} \big(\varepsilon_k (B^{\tau(1)}(t_1) - B^{\tau(1)}(0)) + (B^{\tau(2)}(t_2) - B^{\tau(2)}(t_1)) \\ &\qquad + (\varepsilon_k (B^{\tau(3)}(1) - B^{\tau(3)}(t_2))\big) \\ &:= V_\infty^1(k), \end{aligned} \tag{5.5.15}$$

so that, as $k \to \infty$,

$$\frac{V_\infty^1(k)}{\sigma(k)} \Rightarrow \max_{0 \le t_1 \le t_2 \le 1} (B(t_2) - B(t_1)), \tag{5.5.16}$$

where $B(t)$ is a standard Brownian motion. The right-hand side of (5.5.16) is known as the *local score*, and describes the largest positive increase that B makes within the unit interval. Such functionals are of great importance in sequence comparison, particularly in bioinformatics, (*e.g.*, see Daudin, Ettienne, and Vallois [14].) Moreover, (see Revuz and Yor [40])

$$\begin{aligned} \max_{0 \le t_1 \le t_2 \le 1} (B(t_2) - B(t_1)) &= \max_{0 \le t_2 \le 1} (B(t_2) - \min_{0 \le t_1 \le t_2} B(t_1)) \\ &\stackrel{\mathcal{L}}{=} \max_{0 \le t \le 1} |B(t)|. \end{aligned} \tag{5.5.17}$$

The equality in law between the first and last expressions in (5.5.17) follows immediately from the classical result of Lévy, namely, $(|B(t)|)_{t\geq 0} \stackrel{\mathcal{L}}{=} (\max_{0\leq s\leq t} B(s) - B(t))_{t\geq 0}$. Thus, if we have a random matrix connection to $V^1_\infty(k)$, we can extend it to $\max_{0\leq t\leq 1}|B(t)|$, at least in some limiting sense. This is also interesting from the following point of view. Classically, the Brownian functional $\max_{0\leq t\leq 1} B(t) \stackrel{\mathcal{L}}{=} |B(1)|$, and a trivial random matrix connection can be seen by examining the eigenvalues of the random matrix

$$\begin{pmatrix} Z & 0 \\ 0 & -Z \end{pmatrix}, \tag{5.5.18}$$

where Z is a standard normal random variable. Then, clearly, λ_{max} has the half-normal law, since $\lambda_{max} = \max(Z, -Z) = |Z|$. Thus, the maximal Brownian functional $\max_{0\leq t\leq 1} B(t)$ has a random matrix interpretation, one which is considerably simpler than any potential random matrix interpretation for $\max_{0\leq t\leq 1}|B(t)|$.

CHAPTER VI

CONCLUSION

In this paper, we have obtained the limiting shape of Young tableaux generated by an aperiodic, irreducible, homogeneous Markov chain on a finite state alphabet. The following remarks indicate natural directions in which our results in some cases can, and in other cases, may hope to, be extended.

- Our limiting theorems have all been proved assuming that the initial distribution is the stationary one. However, such results as Theorem 2 of Derriennic and Lin [15] extend our framework to initial distributions started at a specified state. Indeed, in this case, *i.e.,* if for some $k = 1, \ldots, m$, $\mathbb{P}(X_0 = \alpha_k) = 1$, the asymptotic covariance matrix is still given by (5.2.12), and, for example, Theorem 5.3.2 remains valid. For an arbitrary initial distribution, what is needed in this non-stationary context is an invariance principle. More generally, our results continue to hold for k^{th}-order Markov chains, and in fact, they extend to *any* sequence for which both an asymptotic covariance matrix and an invariance principle exist.

- Our limiting theorems have only been proved for finite alphabets. However, in Chapter III, it was seen that for countably infinite iid alphabets, LI_n has a limiting law corresponding to that of a non-uniform, finite-alphabet. Hence, for a Markov chain on a countably infinite alphabet (subject to additional constraints), we might still be able to obtain limiting laws of the form developed in this paper.

- By using appropriate existing concentration inequalities, one can expect to establish the convergence of the moments of the rows of the tableau.

- Various other types of subsequence problems can be tackled by the methodologies used in this thesis. To name but a few, comparisons for unimodal sequences, alternating sequences, and sequences with blocks will deserve further similar studies.

- One field in which the connection between Brownian functionals and random matrix theory has been exploited is in Queuing Theory. The development below, following O'Connell and Yor [36], shows how Brownian functionals of the sort we have studied arise as generalizations of standard queuing models.

Let $A(s,t]$ and $S(s,t]$, $-\infty < s < t < \infty$, be two independent Poisson point process on $\mathbb{R}$, with intensity measures λ and μ, respectively, with $0 < \lambda < \mu$. Here A represents the *arrivals* process, and S the *service time* process, at a queue consisting of a single server. The condition $\lambda < \mu$ ensures that the *queue length*

$$Q(t) = \sup_{-\infty < s \leq t} \{A(s,t] - S(s,t]\}, \tag{6.0.19}$$

is a.s. finite, for any $t \in \mathbb{R}$. Then, defining the *departure* process

$$D(s,t] = A(s,t] - (Q(t) - Q(s)), \tag{6.0.20}$$

which is simply the number of arrivals during $(s,t]$ less the change in the queue length during $(s,t]$, the classical problem is to determine the distribution of $D(s,t]$. The answer to this problem is given by *Burke's Theorem* [10] (see Theorem 1 of [36]):

Theorem 6.0.1 *D is a Poisson process with intensity λ, and $\{D(s,t], s \leq t\}$ is independent of $\{Q(s), s \geq t\}$.*

That is, D has the same law as the arrivals process A. Moreover, since,the queue length after time t is independent of the process D up to time t, one may take the departures from the first queue and use them as inputs to a second queue, and observe that the departure process from the second queue also has the law of A. Proceeding in this way, one generalizes to a *tandem queue* of n servers, each taking the departures from the previous queue as its arrivals process.

One can further generalize this model to a *Brownian queue in tandem* in the following manner. Let $B, B^1, B^2, \ldots, B^n$ be independent, standard Brownian motions on $\mathbb{R}$, and write $B^k(s,t) = B^k(t) - B^k(s)$, for each k and $s < t$, and similarly for B. Let $m > 0$ be a constant, and define, in complete analogy to (6.0.19) and (6.0.20),

$$q_1(t) = \sup_{-\infty < s \le t} \left\{ B(s,t) + B^1(s,t) - m(t-s) \right\}, \tag{6.0.21}$$

and, for $s < t$,

$$d_1(s,t) = B(s,t) - (q_1(t) - q_1(s)). \tag{6.0.22}$$

For $k = 2, 3, \ldots, n$, let

$$q_k(t) = \sup_{-\infty < s \le t} \left\{ d_{k-1}(s,t) + B^k(s,t) - m(t-s) \right\}, \tag{6.0.23}$$

and, for $s < t$,

$$d_k(s,t) = d_{k-1}(s,t) - (q_k(t) - q_k(s)). \tag{6.0.24}$$

Here B is the arrivals process for the first queue, d_{k-1} is the arrivals process for the k^{th} queue ($k \ge 2$), and $mt - B^k(t)$ is the service process for the k^{th} queue, for all k. Using the ideas employed in Burke's Theorem, it can be shown that the generalized queue lengths $q_1(0), q_2(0), \ldots, q_n(0)$ are iid random variables. Moreover, they are exponentially distributed with mean $1/m$.

Using the definitions in (6.0.21)-(6.0.24), and a simple inductive argument, one finds that

$$\sum_{k=0}^{n} q_k(0) = \sup_{t>0}\Big\{B(-t,0) - mt + L_n(t)\Big\}, \tag{6.0.25}$$

where

$$L_n(t) = \sup_{\substack{0\le s_1\le\cdots\\ \le s_{m-1}\le t}} \{B^1(-t,-s_{n-1}) + \cdots + B^n(-s_1,0)\}. \tag{6.0.26}$$

By Brownian rescaling, we observe that

$$\begin{aligned} L_n(t) &\stackrel{\mathcal{L}}{=} \sqrt{t} \sup_{\substack{0\le s_1\le\cdots\\ \le s_{m-1}\le 1}} \{B^1(-1,-s_{n-1}) + \cdots + B^n(-s_1,0)\} \\ &\stackrel{\mathcal{L}}{=} \sqrt{t} V^1_\infty, \end{aligned} \tag{6.0.27}$$

where the functional V^1_∞ is as in Theorem 5.3.2, with associated $n \times n$ covariance matrix $\Sigma = tI_n$ and parameter set $I_{1,n}$. Thus, $L_n(t)$ may be thought of as a process version of this V^1_∞.

The generalized Brownian queues in (6.0.21)-(6.0.24) involved *independent* Brownian motions. These may be generalized to Brownian motions $(B^1, \ldots, B^n)$ for which $(\sigma_1 B^1(t), \ldots \sigma_n B^n(t))$ has (nontrivial) covariance matrix $t\Sigma$. Whether or not we keep the initial arrival process B independent of $(B^1,\ldots,B^n)$, we no longer have that $q_1(0), q_2(0), \ldots, q_n(0)$ are iid random variables, due to the dependence among the service times $mt - B^k(t)$, but we do still have the identity (6.0.25) and (6.0.27) relating the total occupancy of the queue at time zero to V^1_∞. More importantly, our generalizations of the Brownian functionals $L_n(t)$ above can be used to describe the joint law of the input/output of each queue.

- An important topic connecting much of random matrix theory to other problems, such as the shape of random Young tableaux, is the field of orthogonal polynomials. (See, *e.g.*, [31].) It would be of great interest to see what, if any, classes of orthogonal polynomials are associated with the limiting laws in this thesis.

REFERENCES

[1] Aldous, D. and Diaconis, P., “Hammersley’s interacting particle process and longest increasing subsequences,” *Probab. Theory Related Fields*, vol. 103, no. 2, pp. 199–213, 1995.

[2] Aldous, D. and Diaconis, P., “Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem,” *Bull. Amer. Math. Soc. (N.S.)*, vol. 36, no. 4, pp. 413–432, 1999.

[3] Baik, J., Deift, P., and Johansson, K., “Addendum to: “On the distribution of the length of the second row of a Young diagram under Plancherel measure”,” *Geom. Funct. Anal.*, vol. 10, no. 6, pp. 1606–1607, 2000.

[4] Baik, J., Deift, P., and Johansson, K., “On the distribution of the length of the second row of a Young diagram under Plancherel measure,” *Geom. Funct. Anal.*, vol. 10, no. 4, pp. 702–731, 2000.

[5] Baik, J., Deift, P., and Johansson, K., “On the distribution of the length of the longest increasing subsequence of random permutations,” *J. Amer. Math. Soc.*, vol. 12, no. 4, pp. 1119–1178, 1999.

[6] Baryshnikov, Y., “GUEs and queues,” *Probab. Theory Related Fields*, vol. 119, no. 2, pp. 256–274, 2001.

[7] Billingsley, P., *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics, New York: John Wiley & Sons Inc., second ed., 1999. A Wiley-Interscience Publication.

[8] Borodin, A., Okounkov, A., and Olshanski, G., “Asymptotics of Plancherel measures for symmetric groups,” *J. Amer. Math. Soc.*, vol. 13, no. 3, pp. 481–515 (electronic), 2000.

[9] Bougerol, P. and Jeulin, T., “Paths in Weyl chambers and random matrices,” *Probab. Theory Related Fields*, vol. 124, no. 4, pp. 517–543, 2002.

[10] Burke, P. J., “The output of a queuing system,” *Operations Res.*, vol. 4, pp. 699–704 (1957), 1956.

[11] Cator, E. and Groeneboom, P., “Hammersley’s process with sources and sinks,” *Ann. Probab.*, vol. 33, no. 3, pp. 879–903, 2005.

[12] Cator, E. and Groeneboom, P., “Second class particles and cube root asymptotics for Hammersley’s process,” *Ann. Probab.*, vol. 34, no. 4, pp. 1273–1295, 2006.

[13] Chistyakov, G. P. and Götze, F., "Distribution of the shape of Markovian random words," *Probab. Theory Related Fields*, vol. 129, no. 1, pp. 18–36, 2004.

[14] Daudin, J.-J., Etienne, M. P., and Vallois, P., "Asymptotic behavior of the local score of independent and identically distributed random sequences," *Stochastic Process. Appl.*, vol. 107, no. 1, pp. 1–28, 2003.

[15] Derriennic, Y. and Lin, M., "The central limit theorem for Markov chains with normal transition operators, started at a point," *Probab. Theory Related Fields*, vol. 119, no. 4, pp. 508–528, 2001.

[16] Doumerc, Y., "A note on representations of eigenvalues of classical Gaussian matrices," in *Séminaire de Probabilités XXXVII*, vol. 1832 of *Lecture Notes in Math.*, pp. 370–384, Berlin: Springer, 2003.

[17] Durrett, R., *Probability: theory and examples.* Belmont, CA: Duxbury Press, second ed., 1996.

[18] Fulton, W., *Young tableaux*, vol. 35 of *London Mathematical Society Student Texts.* Cambridge: Cambridge University Press, 1997. With applications to representation theory and geometry.

[19] Geman, S., "A limit theorem for the norm of random matrices," *Ann. Probab.*, vol. 8, no. 2, pp. 252–261, 1980.

[20] Glynn, P. W. and Whitt, W., "Departures from many queues in series," *Ann. Appl. Probab.*, vol. 1, no. 4, pp. 546–572, 1991.

[21] Gordin, M. I., "The central limit theorem for stationary processes," *Dokl. Akad. Nauk SSSR*, vol. 188, pp. 739–741, 1969.

[22] Gravner, J., Tracy, C. A., and Widom, H., "Limit theorems for height fluctuations in a class of discrete space and time growth models," *J. Statist. Phys.*, vol. 102, no. 5-6, pp. 1085–1132, 2001.

[23] Greene, C., "Some partitions associated with a partially ordered set," *J. Combinatorial Theory Ser. A*, vol. 20, no. 1, pp. 69–79, 1976.

[24] Groeneboom, P., "Ulam's problem and Hammersley's process," *Ann. Probab.*, vol. 29, no. 2, pp. 683–690, 2001.

[25] Groeneboom, P., "Hydrodynamical methods for analyzing longest increasing subsequences," *J. Comput. Appl. Math.*, vol. 142, no. 1, pp. 83–105, 2002. Probabilistic methods in combinatorics and combinatorial optimization.

[26] Hambly, B. M., Martin, J. B., and O'Connell, N., "Concentration results for a Brownian directed percolation problem," *Stochastic Process. Appl.*, vol. 102, no. 2, pp. 207–220, 2002.

[27] Houdré, C., Lember, J., and Matzinger, H., "On the longest common increasing binary subsequence," *C. R. Math. Acad. Sci. Paris*, vol. 343, no. 9, pp. 589–594, 2006.

[28] Its, A. R., Tracy, C. A., and Widom, H., "Random words, Toeplitz determinants, and integrable systems. I," in *Random matrix models and their applications*, vol. 40 of *Math. Sci. Res. Inst. Publ.*, pp. 245–258, Cambridge: Cambridge Univ. Press, 2001.

[29] Its, A. R., Tracy, C. A., and Widom, H., "Random words, Toeplitz determinants and integrable systems. II," *Phys. D*, vol. 152/153, pp. 199–224, 2001. Advances in nonlinear mathematics and science.

[30] Johansson, K., "Shape fluctuations and random matrices," *Comm. Math. Phys.*, vol. 209, no. 2, pp. 437–476, 2000.

[31] Johansson, K., "Discrete orthogonal polynomial ensembles and the Plancherel measure," *Ann. of Math. (2)*, vol. 153, no. 1, pp. 259–296, 2001.

[32] Kuperberg, G., "Random words, quantum statistics, central limits, random matrices," *Methods Appl. Anal.*, vol. 9, no. 1, pp. 99–118, 2002.

[33] Lifshits, M. A., *Gaussian random functions*, vol. 322 of *Mathematics and its Applications*. Dordrecht: Kluwer Academic Publishers, 1995.

[34] Logan, B. F. and Shepp, L. A., "A variational problem for random Young tableaux," *Advances in Math.*, vol. 26, no. 2, pp. 206–222, 1977.

[35] Mehta, M. L., *Random matrices*, vol. 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third ed., 2004.

[36] O'Connell, N. and Yor, M., "Brownian analogues of Burke's theorem," *Stochastic Process. Appl.*, vol. 96, no. 2, pp. 285–304, 2001.

[37] O'Connell, N. and Yor, M., "A representation for non-colliding random walks," *Electron. Comm. Probab.*, vol. 7, pp. 1–12 (electronic), 2002.

[38] Okounkov, A., "Random matrices and random permutations," *Internat. Math. Res. Notices*, no. 20, pp. 1043–1095, 2000.

[39] Pitman, J. W., "One-dimensional Brownian motion and the three-dimensional Bessel process," *Advances in Appl. Probability*, vol. 7, no. 3, pp. 511–526, 1975.

[40] Revuz, D. and Yor, M., *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Berlin: Springer-Verlag, third ed., 1999.

[41] Seppäläinen, T., "A microscopic model for the Burgers equation and longest increasing subsequences," *Electron. J. Probab.*, vol. 1, pp. no. 5, approx. 51 pp. (electronic), 1996.

[42] Seppäläinen, T., "A scaling limit for queues in series," *Ann. Appl. Probab.*, vol. 7, no. 4, pp. 855–872, 1997.

[43] Tracy, C. A. and Widom, H., "Level-spacing distributions and the Airy kernel," *Comm. Math. Phys.*, vol. 159, no. 1, pp. 151–174, 1994.

[44] Tracy, C. A. and Widom, H., "On the distributions of the lengths of the longest monotone subsequences in random words," *Probab. Theory Related Fields*, vol. 119, no. 3, pp. 350–380, 2001.

[45] Veršik, A. M. and Kerov, S. V., "Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of Young tableaux," *Dokl. Akad. Nauk SSSR*, vol. 233, no. 6, pp. 1024–1027, 1977.

[46] Xu, H., *Random Matrices and Subsequences.* PhD thesis, Georgia Institute of Technology, 2008.

VITA

Trevis J. Litherland was born in Olney, Illinois on December 1, 1971, and completed his B.S. in Mathematics at Rose-Hulman Institute of Technology in Terre Haute, Indiana in 1993. As part of his undergraduate education, he took his junior year abroad at the University of Limerick in Ireland. Following the completion of his undergraduate studies, he obtained his M.S. in Applied Mathematics from Georgia Institute of Technology in 1995. Before returning to Georgia Tech to pursue his Ph.D. in Mathematics in 2004, he worked in the Atlanta area as a statistical analyst in the credit scoring industry, where he first began to fully appreciate the power of probabilistic thinking. Trevis married his wife, Sunchin, in 2002, and currently resides in Powder Springs, Georgia. His hobbies include gardening and conducting small-scale, low-tech, civil engineering projects loosely associated with gardening.