2014

# Cycle lengths of θ-biased random permutations

Tongjia Shi
*Harvey Mudd College*

# Cycle Lengths of $\theta$-Biased Random Permutations

**Tongjia Shi**

Nicholas Pippenger, Advisor

Michael Orrison, Reader

HARVEY MUDD COLLEGE

Department of Mathematics

May, 2014

# Abstract

Consider a probability distribution on the permutations of $n$ elements. If the probability of each permutation is proportional to $\theta^K$, where $K$ is the number of cycles in the permutation, then we say that the distribution generates a $\theta$-biased random permutation. A random permutation is a special $\theta$-biased random permutation with $\theta = 1$. The $m^{\text{th}}$ moment of the $r^{\text{th}}$ longest cycle of a random permutation is $\Theta(n^m)$, regardless of $r$ and $\theta$. The joint moments are derived, and it is shown that the longest cycles of a permutation can either be positively or negatively correlated, depending on $\theta$. The $m^{\text{th}}$ moments of the $r^{\text{th}}$ shortest cycle of a random permutation is $\Theta(n^{m-\theta}/(\ln n)^{r-1})$ when $\theta < m$, $\Theta((\ln n)^r)$ when $\theta = m$, and $\Theta(1)$ when $\theta > m$. The exponent of cycle lengths at the $100q^{\text{th}}$ percentile goes to $q$ with zero variance. The exponent of the expected cycle lengths at the $100q^{\text{th}}$ percentile is at least $q$ due to the Jensen's inequality, and the exact value is derived.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Preliminaries

## 1.1 A Quick Probability Review

Informally, a **random variable** assigns a probability to each possible outcome. We call the set of all possible outcomes, $\Omega$, the **sample space** of the random variable. For example, let $Y_1$ denote the random variable of the outcome of throwing a fair dice; then the sample space of $Y_1$ is $\Omega = \{1, 2, 3, 4, 5, 6\}$. The random variable $Y_1$ assigns each of these outcomes a probability of $1/6$. We can write:

$$\Pr(Y_1 = 1) = \Pr(Y_1 = 2) = \cdots = \Pr(Y_1 = 6) = \frac{1}{6}.$$

This is an example of a **discrete random variable**, because the set of possible outcomes is at most countably infinite. On the other hand, a **continuous random variable** is a random variable whose sample space is uncountably infinite. We use a **density function** to define the relative likelihood for each outcome. For example, let $Y_2$ be a random number uniformly chosen between 0 and 1; then $Y_2$ is a continuous random variable. The density function of $Y_2$ is

$$f_{Y_2}(y) = 1, \qquad 0 \le y \le 1.$$

When the set of observations of a random variable $X$ is a subset of the real numbers, we call $X$ a **real-valued random variable**. Both $Y_1$ and $Y_2$ in the above examples are real-valued. We can define the **expected value** or the **first moment** of $X$ to be

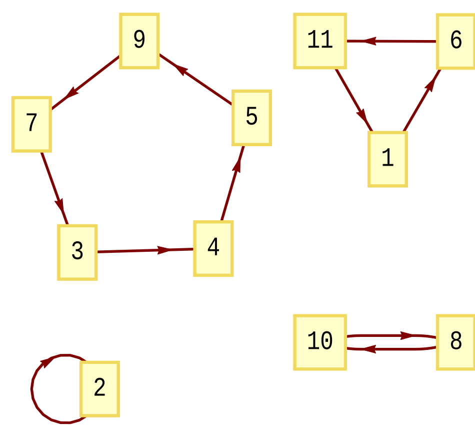
$$\begin{cases} \mathbb{E}[X] = \sum_{\omega \in \Omega} \omega \Pr(X = \omega) & \text{if } X \text{ is discrete,} \\ \mathbb{E}[X] = \int_{\Omega} f_X(\omega)\, d\omega & \text{if } X \text{ is continuous.} \end{cases}$$

The $\mathbf{m^{th}}$ **moment** of a random variable $X$ is defined as $\mathbb{E}[X^m]$.

## 1.2 Random Permutations and Ordered Cycle Lengths

A permutation is a bijection from a set of elements to itself. For example, the following diagram represents a permutation on the set $\{1, 2, 3, \cdots, 11\}$. It maps the element 1 to 6, 6 to 11, 11 to 1, 2 to itself, etc.

**Figure 1.1** A permutation on 11 elements




Since the labeling of the underlying set does not matter, we use "a permutation on $n$ elements" to denote a bijection from $[n] = \{1, 2, \cdots, n\}$ to itself.

We use the notation $K_{ij}$ to denote the **number of cycles** between length $i$ and $j$. For a permutation on $n$ elements, all cycles are between length 0 and $n$, so $K_{0n}$ is simply the number of cycles in the permutation.

A permutation is a type of **decomposable structure** because we can decompose a permutation into cycles. In the example above, we have a decomposable structure of size 11. It has 4 components (cycles), with size $5, 3, 1, 2$, respectively.

The **component frequency spectrum** is a description of the distribution of components (cycles) based on their sizes. For a decomposable structure of size $n$, we let $C_k$ denote the number of its components of size $k$, where $k = 1, 2, \cdots, n$. Then

$$\sum_{k=1}^{n} kC_k = n.$$

We call $C = (C_1, C_2, \cdots, C_n)$ the vector of component counts. The component frequency spectrum is then determined completely by this vector. In the example above, the vector of component counts is $C = (C_1, \cdots, C_{11})$ with

$$C_1 = 1, C_2 = 1, C_3 = 1, C_4 = 0, C_5 = 1, C_6 = C_7 = \cdots = C_{11} = 0.$$

In this thesis we are interested in the ordered cycle lengths of a permutation. The length of the longest cycle of a permutation is the largest $k$ such that $C_k \neq 0$. Similarly we can define the length of the $r^{\text{th}}$ longest cycle, the $r^{\text{th}}$ shortest cycle, and the cycle length at the $100q^{\text{th}}$ percentile. We use $L_r$, $S_r$ and $M_q$ to denote the these values. In our previous example, we have $L_1 = 5$, $L_2 = 3$, $S_1 = 1$, $S_2 = 2$, and $M_{0.5} = 2.5$.

A random permutation on $n$ elements is a probability distribution on all possible $n!$ permutations on the set $\{1, 2, 3, \cdots, n\}$. If the random permutation assigns equal probability to all these $n!$ possible permutations, we call it an **unbiased random permutation**.

The vector of component counts $C = (C_1, \cdots, C_n)$ of a random permutation then becomes a vector of real-valued random variables. The ordered cycle lengths, $L_r$, $S_r$ and $M_q$ are also random variables.

## 1.3 Chinese Restaurant Process

We will study the distribution of ordered cycle lengths for a random permutation as $n$ grows to infinity. But how exactly can we grow from a random permutation on $n$ elements to a random permutation on $n+1$ elements? We can do this through a process called the **Chinese restaurant process**.

Consider a Chinese restaurant where customers go into the restaurant one after another. Let $n$ denote the number of customers in the restaurant. Initially $n = 0$. At time $t = 1$ one customer comes into the restaurant and sits at a new table. At time $t+1$ the $(n+1)^{\text{th}}$ customer comes in. The new customer sits to the right of each of the existing $n$ customers with probability $\dfrac{1}{n+1}$. With probability $\dfrac{1}{n+1}$, she will sit at a new table by herself. Then at each time $t$, the tables in this Chinese restaurant represents the cycles in an unbiased random permutation.

## 1.4 $\theta$-biased Random Permutations

In a previous section we mentioned that an unbiased random permutation assigns equal probability to all possible permutations. In this section we generalize this concept to introduce a family called the $\theta$-biased random permutations.

$\theta$-biased random permutations arises naturally from the Ewens sampling formula, ESF($\theta$) (Ewens, 1972). In a $\theta$-biased random permutation, the density of any permutation is scaled by $\theta^{K_{0n}}$, where $K_{0n}$ is the number of cycles. If $\theta$ is larger than one, then permutations with more cycles get chosen more often, which implies that on average there will be smaller cycles. If $\theta$ is smaller than one, then on average the cycles will be longer. When $\theta \to \infty$ (with $n$ fixed), the permutation becomes the identity map with probability one. When $\theta \to 0$ (with $n$ fixed), the permutation becomes a cycle with probability one.

A $\theta$-biased random permutation can be conveniently generated by a variation of the Chinese restaurant process. At time $t+1$, the $(n+1)^{\text{th}}$ customer chooses to sit to the right of each existing customer with probability $\dfrac{1}{n+\theta}$, and to sit at a new table with probability $\dfrac{\theta}{n+\theta}$. Then at each time $t$, the tables in this Chinese restaurant represents the cycles in a $\theta$-biased random permutation.

Notice that with a larger $\theta$, customers are more likely to sit at a new table, so on average we expect to see more tables, each with fewer customers. When $\theta$ goes to zero, customers almost never sit at a new table, so we are likely to end up with large but fewer tables.

## 1.5 The Conditioning Relation

A common feature for decomposable combinatorial structures is called the **conditioning relation**, which says

$$(C_1, C_2, \cdots, C_n) \sim (Z_1, Z_2, \cdots, Z_n | T_{0n} = n),$$

where $Z_1, Z_2, \cdots, Z_n$ are independent random variables, and

$$T_{0n} = \sum_{k=0}^{n} kZ_k.$$

The random variables $Z_i$ have either Poisson, negative binomial or binomial distributions depending on the class of the structure (assembly, multiset, selection, respectively) (Arratia et al., 2003). Random permutation is a type of assembly and $Z_i \sim \mathrm{Poi}(i)$.

The conditioning relation says that in a ($\theta$-biased) random permutation on $n$ elements, the joint cycle counts are distributed just like the independent Poisson variables, conditioning on the event that the independent Poisson variables happen to form cycles with a total of $n$ elements. The cycle counts have a very complicated joint distribution, and the conditioning relation allows us to investigate it through a much simpler collection of Poisson random variables.

As a side note, $\theta$-biased random permutations are also a type of logarithmic combinatorial structure, which means $Z_i$'s satisfy the **logarithmic condition**:

$$i\,\Pr[Z_i = 1] \to \theta, i\mathbb{E}[Z_i] \to \theta, \text{ as } i \to \infty.$$

# Chapter 2

# Literature Review

The thesis studies the ordered cycle lengths of a $\theta$-biased random permutation on $n$ elements as $n$ goes to infinity. Thus we begin by introducing some results from earlier research.

## 2.1 Longest and Shortest Cycles with $\theta = 1$

Shepp and Lloyd determined the asymptotic behavior of the $m^{\text{th}}$ moment of the size of the $r^{\text{th}}$ longest cycle ($L_r$) and $r^{\text{th}}$ shortest cycle ($S_r$) in a random permutation (Shepp and Lloyd, 1966). Before we summarize their results, we give some definitions. Let $\mathbb{E}_n$ be the expectation taken on a random permutation of size $n$. Let

$$E(x) = \int_x^\infty \frac{e^{-y}}{y}\, dy. \tag{2.1}$$

Then for the longest cycles, we have

$$\lim_{n\to\infty} \mathbb{E}_n \left[ \left( \frac{L_r}{n} \right)^m \right] = G_{r,m},\ m = 0, 1, \cdots, r = 1, 2, \cdots, \tag{2.2}$$

where

$$G_{r,m} = \int_0^\infty x^{m-1} m! \frac{E(x)^{r-1}}{(r-1)!} e^{-E(x)-x}\, dx.$$

These constants have the following asymptotic property:

$$\lim_{r\to\infty} (m+1)^r G_{r,m} = \frac{e^{-m\gamma}}{m!}, \quad m = 0, 1, \cdots,$$

where $\gamma \approx 0.5772156649$ is the Euler's constant.

For the shortest cycles, we have

$$\lim_{n\to\infty} \frac{\mathbb{E}_n[S_r]}{(\log n)^r} = \frac{e^{-\gamma}}{r!}, \quad r \geq 1, \tag{2.3a}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}_n[S_r^m]}{n^{m-1}(\log n)^{r-1}} = \frac{1}{(r-1)!} \int_0^\infty \frac{x^{m-1}}{(m-1)!} e^{E(x)-x}\, dx, \quad m \geq 2, r \geq 1 \tag{2.3b}$$

This is saying that the expected length of $r^{\text{th}}$ longest cycles grows as $\Theta(n)$. The expected length of $r^{\text{th}}$ shortest cycle grows as $\Theta((\log n)^r)$.

## 2.2 Longest and Shortest Cycle with $\theta = 1/2$

Pippenger gave a natural interpretation to a biased random permutation with $\theta = 1/2$. It is called a random cyclation (Pippenger, 2013). Pippenger found the asymptotic behavior of the expected length of the longest and shortest cycle in such a biased random permutation. We summarize the results below.

$$\lim_{n\to\infty} \frac{\mathbb{E}[L_1]}{n} = \int_0^\infty e^{-E(x)/2-x} \tag{2.4}$$

$$\lim_{n\to\infty} \frac{\mathbb{E}[S_1]}{\sqrt{n}} = \frac{\sqrt{\pi}}{2} \int_0^\infty e^{E(x)/2-x}. \tag{2.5}$$

This is saying that the expected length of the longest cycle in such a biased random permutation grows as $\Theta(n)$. The expected length of the shortest cycle grows as $\Theta(\sqrt{n})$.

## 2.3 Longest Cycles for any $\theta$

The longest cycles in a $\theta$-biased ranom permutation are associated with the Poisson-Dirichlet distribution, and thus have been studied more often than smallest cycles. In fact, $(L_1, L_2, \cdots)/n$ converges to the Poisson-Dirichlet distribution with parameter $\theta$ as $n \to \infty$. In 1979, Griffith found the joint moments of Poisson-Dirichlet distribution (Griffiths, 1979). and thus we have the following results for the $\theta$-biased random permutations:

$$\lim_{n\to\infty} \frac{\mathbb{E}_n[L_1^{j_1} \cdots L_r^{j_r}]}{n^j} = \frac{\theta^r \Gamma(\theta)}{\Gamma(\theta+j)} \int y_1^{j_1-1} \cdots y_r^{j_r-1} e^{-\sum_{l=1}^r y_l - \theta E_1(y_r)}\, dy_1 \cdots dy_r, \tag{2.6}$$

where $r \geq 1$ and $j_1 + \cdots + j_r = j$.

For the $r^{\text{th}}$ longest cycle alone,

$$\lim_{n\to\infty} \frac{\mathbb{E}_n[L_r^j]}{n^j} = \frac{\Gamma(\theta+1)}{\Gamma(\theta+j)} \int_0^\infty \frac{(\theta E_1(x))^{r-1}}{(r-1)!} x^{j-1} e^{-x-\theta E_1(x)}\, dx. \tag{2.7}$$

# Chapter 3

# Longest Cycles

## 3.1 Moments of the Length of $r^{\text{th}}$ Longest Cycle

In this section we will re-derive the moments of the length of $r^{\text{th}}$ longest cycle for a $\theta$-biased random permutation, without resorting to the Poisson-Dirichlet distribution. We follow the main idea of the Shepp and Lloyd (1966) paper.

Consider a $\theta$-biased random permutation on $[n]$. According to the conditioning relation, as $n \to \infty$ the cycle structure $C^{(n)} = (C_1^{(n)}, C_2^{(n)}, \cdots)$ is distributed as independent Poisson variables $(Z_1, Z_2, \cdots)$ with parameters $\lambda_j = \theta/j$ for $j = 1, 2, \cdots$, conditioning on $T_{0n} = \sum_{j=1}^{n} jZ_j = n$.

Taking advantage of the Poisson distribution, we will describe the cycle structure in the following way. Consider a unit Poisson process on the positive real line. Let the random variable $Z_j$ then takes the value equal to the number of jumps on the interval $[t_j, t_{j+1})$, where

$$t_j = \theta \sum_{k=1}^{j-1} \frac{1}{k}, \qquad k = 1, 2, 3, \cdots, n.$$

Then

$$(C_1^{(n)}, C_2^{(n)}, \cdots, C_n^{(n)}) \sim (Z_1, Z_2, \cdots, Z_n) | [Z_{n+1} = n].$$

We need to analyze the behavior as $n \to \infty$. But the problem is unbounded ($t_\infty = \infty$). In addition, the conditioning nature of $C^{(n)}$ makes the analysis difficult. Hence we will take $n \to \infty$ in a different way.

Consider the family of random process $C^{(z)} = (C_1^{(z)}, C_2^{(z)}, \cdots)$ (an analogue of the cycle structure) where $0 < z < 1$. $C_j^{(z)}$ equals the number of jumps in the interval $[t_j(z), t_{j+1}(z)]$, where

$$t_j(z) = \theta \sum_{k=1}^{j-1} \frac{z^k}{k}, \qquad k = 1, 2, 3, \cdots .$$

Notice that

$$t_\infty(z) = \lim_{j \to \infty} t_j(z) = \theta \log \frac{1}{1-z}.$$

We will first show that $C^{(z)}$ and $C^{(n)}$ are related in a simple way. Then we will investigate the problem for each $C^{(z)}$. Then we will connect the asymptotic behavior of $C^{(z)}$ with that of $C^{(n)}$ using Tauberian theorems. Throughout the section, we treat $\theta$ as a constant parameter.

### 3.1.1 Relating Functionals on $C^{(z)}$ and $C^{(n)}$

We will first find the distribution of $v_z = \sum_{j=1}^n jC_j^{(z)}$. We start with the PGF of $C_j^{(z)}$:

$$G_{C_j^{(z)}}(x) = \exp(\lambda_j(x-1)) = \exp\left(\frac{\theta z^j (x-1)}{j}\right).$$

Then we have

$$\begin{aligned} G_{v_z}(x) &= \prod_{j=1}^{\infty} \exp\left(\frac{\theta z^j (x^j - 1)}{j}\right) \\ &= \left[\exp\left(\sum_{i=1}^{\infty} \frac{(zx)^j}{j} - \sum_{i=1}^{\infty} \frac{z^j}{j}\right)\right]^\theta \\ &= \left(\frac{1-z}{1-xz}\right)^\theta \end{aligned}$$

This is a negative binomial distribution $\mathrm{NB}(\theta, z)$, where $z$ is the success probability. So we have its pmf:

$$\Pr(v_z = n) = \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} (1-z)^\theta z^n. \tag{3.1}$$

We know

$$\Pr(C_j^{(z)} = a) = e^{-\theta z^j / j} \frac{\theta (z^j / j)^a}{a!}.$$

Thus the joint distribution is

$$\begin{aligned}
\Pr(C^{(z)} = (a_1, \cdots)) &= \prod_{j=1}^{\infty} e^{-\theta z^j/j} \frac{\theta(z^j/j)^{a_j}}{a_j!} \\
&= e^{-\theta \sum z^j/j} \prod_{j=1}^{\infty} \frac{\theta(z^j/j)^{a_j}}{a_j!} \\
&= (1-z)^{\theta} \theta^{v} \prod_{j=1}^{\infty} \frac{(z^j/j)^{a_j}}{a_j!} \\
&= (1-z)^{\theta} \theta^{v} z^{v} \prod_{j=1}^{\infty} \frac{(1/j)^{a_j}}{a_j!}.
\end{aligned}$$

Notice that this is just a scaled version of the distribution of cycle structures in an unbiased random permutation. So the conditional distribution is

$$\Pr(C^{(z)} = (a_1, \cdots) | v_z = n) = \prod_{j=1}^{\infty} \frac{(1/j)^{a_j}}{a_j!}, \qquad \sum_{j=1}^{\infty} j a_j = n.$$

Hence, for any functional on the cycle structure, $\Phi(C^{(z)})$, we have

$$\begin{aligned}
\mathbb{E}_z[\Phi] &= \mathbb{E}[\Phi(C^{(z)})] = \mathbb{E}[\mathbb{E}[\Phi | v]] \\
&= \sum_{n=0}^{\infty} \Pr(v_z = n) \mathbb{E}_n[\Phi] \\
&= \sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} (1-z)^{\theta} z^n E_n(\Phi).
\end{aligned} \tag{3.2}$$

Here, we use $\mathbb{E}_n$ to denote the expectation given that there are a total of $n$ elements in the permutation.

### 3.1.2 Largest Components in $C^{(z)}$

The probability density at $t$ of the $r^{\text{th}}$ last jump of the Poisson process on $[t_0, t_\infty(z)]$ is

$$e^{-(t_\infty(z)-t)} \frac{[t_\infty(z) - t]^{r-1}}{(r-1)!}, \qquad -\infty < t \leq t_\infty(z).$$

The $m^{\text{th}}$ moment of the length of the $r^{\text{th}}$ longest cycle is then

$$\mathbb{E}_z[(L_r)^m] = \sum_{j=1}^{\infty} j^m \int_{t_j(z)}^{t_{j+1}(z)} e^{-(t_\infty(z)-t)} \frac{[t_\infty(z) - t]^{r-1}}{(r-1)!} \, dt.$$

Make a change of variable

$$t_\infty(z) - t = \theta E(x) = \theta \int_x^\infty \frac{e^{-y}}{y}\, dy.$$

Let $z = e^{-s}$, $0 < s < \infty$. Since

$$t_\infty(e^{-s}) - t_j(e^{-s}) = \theta \sum_{k=j}^{\infty} \frac{e^{-ks}}{j}, \qquad j = 1, 2, \cdots$$

and

$$E(js) < \sum_{k=j}^{\infty} \frac{e^{-ks}}{k} \le E((j-1)s), \qquad j = 1, 2, \cdots,$$

there exists $x_j(s)$ such that $(j-1)s \le x_j(s) < js$ and

$$\sum_{k=j}^{\infty} \frac{e^{-ks}}{k} = E(x_j(s)).$$

So we have

$$\mathbb{E}_z[(L_r)^m] = \sum_{j=1}^{\infty} j^m \int_{x_j(s)}^{x_{j+1}(s)} e^{-\theta E(x)} \theta^{r-1} \frac{E(x)^{r-1}}{(r-1)!} \frac{\theta e^{-x}}{x}\, dx.$$

### 3.1.3 Using Tauberian Theorems

Notice that

$$\sum_{j=1}^{\infty} [x_j(s)]^m \mu_j \le s^m \mathbb{E}_z[(L_r)^m] < \sum_{j=1}^{\infty} [x_{j+1}(s)]^m \mu_j,$$

with

$$\mu_j = \int_{x_j(s)}^{x_{j+1}(s)} e^{-\theta E(x)} \frac{E(x)^{r-1}}{(r-1)!} \frac{\theta^r e^{-x}}{x}\, dx.$$

This is an approximation of the Riemann integral as $s \to 0$. So $s/(1-z) \to 1$, $x_1(s) \to 0$ and

$$\lim_{z \to 1} \frac{(1-z)^m}{m!} \mathbb{E}_z^\theta[(L_r)^m] = G_{r,m}^\theta,$$

where

$$G_{r,m}^\theta = \theta^r \int_0^\infty \frac{x^{m-1}}{m!} e^{-\theta E(x) - x} \frac{E(x)^{r-1}}{(r-1)!}\, dx.$$

We will now use Tauberian theorems. First,

$$\begin{aligned}
G^{\theta}_{r,m} &= \lim_{z\to 1} \frac{(1-z)^m}{m!}\mathbb{E}^{\theta}_z[(L_r)^m] \\
&= \lim_{z\to 1} \frac{(1-z)^m}{m!}\sum_{n=0}^{\infty}\frac{\Gamma(n+\theta)}{n!\Gamma(\theta)}(1-z)^{\theta}z^n\mathbb{E}_n[(L_r)^m] \\
&= \lim_{z\to 1}\sum_{n=0}^{\infty}\frac{\Gamma(n+\theta)}{m!n!\Gamma(\theta)}\mathbb{E}_n[(L_r)^m](1-z)^{m+\theta}z^n.
\end{aligned}$$

Since the coefficients of $z^n$ are nonnegative, we have (with $\gamma = m+\theta$ in de Bruijn (1958: p. 147))

$$\sum_{k=0}^{n}\frac{\Gamma(n+\theta)}{m!n!\Gamma(\theta)}\mathbb{E}_n[(L_r)^m] \sim \Gamma(m+\theta+1)^{-1}n^{m+\theta}G^{\theta}_{r,m}, \qquad n\to\infty.$$

Now, we will show that $\dfrac{\Gamma(n+\theta)}{m!n!\Gamma(\theta)}E_n[(L_r)^m]$ is nondecreasing in $n$. This is equivalent to

$$(n+1)\mathbb{E}_n[L_r^m] \le (n+\theta)\mathbb{E}_{n+1}[L_r^m].$$

Recall the notation that $\mathbb{E}_n[L_r^m] = \mathbb{E}[(L_r^{(n)})^m]$. By the double expectation theorem,

$$\mathbb{E}[(L_r^{(n+1)})^m] = \mathbb{E}[\mathbb{E}[(L_r^{(n+1)})^m]|L_r^{(n)}].$$

From the Chinese restaurant, we know that for a $\theta$-biased random permutation on $n$ elements, the next element will be added to the longest cycle with probability at least $\dfrac{L_r}{n+\theta}$ (the "at least" is because there can be multiple longest cycles), in which case $L_r^{(n+1)} = L_r^{(n)}$. Therefore,

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[(L_r^{(n+1)})^m]|L_r^{(n)}] &\ge \mathbb{E}\left[\frac{L_r}{n+\theta}(L_r+1)^m + \frac{n+\theta-L_n}{n+\theta}L_r^m\right] \\
&= \mathbb{E}\left[L_r^m + \frac{L_r}{n+\theta}(L_r+1)^m - \frac{L_n}{n+\theta}L_r^m\right] \\
&= \mathbb{E}[L_r^m] + \mathbb{E}\left[\frac{L_r[(L_r+1)^m - L_r^m]}{n+\theta}\right] \\
&\ge \mathbb{E}[L_r^m] + \mathbb{E}\left[\frac{L_r(mL_r^{m-1})}{n+\theta}\right] \\
&\ge \mathbb{E}_n[L_r^m] + \frac{m\mathbb{E}_n[L_r^m]}{n+\theta}.
\end{aligned}$$

Thus,

$$(n+\theta)\mathbb{E}_{n+1}[L_r^m] \geq (n+\theta+m)\mathbb{E}_n[L_r^m] \geq (n+0+1)\mathbb{E}_n[L_r^m],$$

as desired. Hence (de Bruijn, 1958: p. 139)

$$\frac{\Gamma(n+\theta)}{m!n!\Gamma(\theta)} E_n[(L_r)^m] \sim \Gamma(m+\theta)^{-1} n^{m+\theta-1} G_{r,m}^{\theta}, \qquad n \to \infty.$$

$$E_n\left[\left(\frac{L_r}{n}\right)^m\right] \sim \frac{m!n!\Gamma(\theta)}{\Gamma(n+\theta)\Gamma(m+\theta)} n^{\theta-1} G_{r,m}^{\theta}, \qquad n \to \infty.$$

Notice that

$$\lim_{n\to\infty} \frac{n!n^{\theta-1}}{\Gamma(n+\theta)} = 1.$$

So the expression simplifies to

$$E_n\left[\left(\frac{L_r}{n}\right)^m\right] \sim \frac{m!\Gamma(\theta)}{\Gamma(m+\theta)} G_{r,m}^{\theta}, \qquad n \to \infty.$$

In fact the right hand side is a constant, so

$$\lim_{n\to\infty} E_n\left[\left(\frac{L_r}{n}\right)^m\right] = \frac{m!\Gamma(\theta)}{\Gamma(m+\theta)} \theta^r \int_0^\infty \frac{x^{m-1}}{m!} e^{-\theta E(x)-x} \frac{E(x)^{r-1}}{(r-1)!}\, dx.$$

Rewriting this we get

$$\lim_{n\to\infty} E_n\left[\left(\frac{L_r}{n}\right)^m\right] = \theta^{r-1} \frac{\Gamma(1+\theta)}{\Gamma(m+\theta)} \int_0^\infty x^{m-1} e^{-\theta E(x)-x} \frac{E(x)^{r-1}}{(r-1)!}\, dx.$$

Or equivalently,

$$\lim_{n\to\infty} E_n\left[\left(\frac{L_r}{n}\right)^m\right] = \theta^{r-1}(\theta+1)\cdots(\theta+m-1) \int_0^\infty x^{m-1} e^{-\theta E(x)-x} \frac{E(x)^{r-1}}{(r-1)!}\, dx. \tag{3.3}$$

### 3.1.4 Special Cases

When $\theta = 1$. The $m^{\text{th}}$ moment of the $r^{\text{th}}$ longest cycle (in an unbiased random permutation) is

$$\lim_{n\to\infty} E_n\left[\left(\frac{L_r}{n}\right)^m\right] = \int_0^\infty \frac{x^{m-1}}{m!} e^{-E(x)-x} \frac{E(x)^{r-1}}{(r-1)!}\, dx,$$

agreeing with Shepp and Lloyd's result.

When $\theta = 1/2$. The expected length of the longest cycle is

$$\lim_{n\to\infty} E_n\left[\left(\frac{L_1}{n}\right)^1\right] = \int_0^\infty e^{-E(x)/2-x}\,dx,$$

agreeing with Pippenger's result.

The expected length of the longest cycle is

$$\lim_{n\to\infty} E_n\left[\frac{L_1}{n}\right] = \int_0^\infty e^{-\theta E(x)-x}\,dx.$$

Hence,

$$\begin{aligned}
\left.\frac{\partial E_n\left[\frac{L_1}{n}\right]}{\partial\theta}\right|_{\theta=0} &= \int_0^\infty e^{-x}\left(-\int_x^\infty \frac{e^{-y}}{y}\,dy\right)\,dx \\
&= -\int_0^\infty \frac{e^{-y}}{y}\left(\int_0^y e^{-x}\,dx\right)\,dy \\
&= -\int_0^\infty \frac{e^{-y}(1-e^{-y})}{y}\,dy \\
&= -\ln 2
\end{aligned}$$

So the expected length of longest cycle is $(1-\theta\ln 2)n$ near $\theta = 0$ as $n \to \infty$.

## 3.2 Numerical Results

In this section we summarizes some numerical results based on the derivation above.

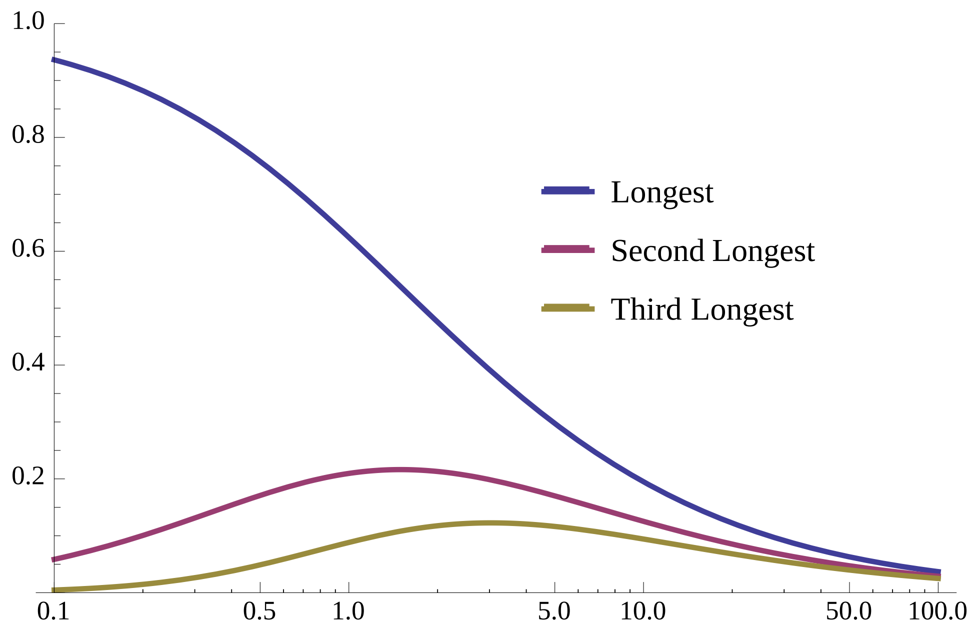
**Table 3.1** The expected length of longest cycles for $\theta = 1.0, 0.8, 0.5$.

| | 1.0 | 0.8 | 0.5 |
|---|---|---|---|
| 1 | 0.62432998854355 | 0.536108373545422 | 0.378911505634246 |
| 2 | 0.20958087428419 | 0.160284846024997 | 0.085454809929831 |
| 3 | 0.08831609888315 | 0.060408169200032 | 0.024448712684230 |
| 4 | 0.04034198873687 | 0.024676360768897 | 0.007572860699943 |
| 5 | 0.01914548402332 | 0.010462396440102 | 0.002428938443347 |
| 6 | 0.00927494376258 | 0.004523280173624 | 0.000792675331955 |
| 7 | 0.00454696522865 | 0.001977125526407 | 0.000261075394740 |
| 8 | 0.00224517570820 | 0.000869766839161 | 0.000086425509984 |
| 9 | 0.00111356578567 | 0.000384105270082 | 0.000028692496835 |
| 10 | 0.00055387022318 | 0.000170030817358 | 0.000009541484700 |
| 11 | 0.00027598637365 | 0.000075378090968 | 0.000003176029182 |
| 12 | 0.00013768220553 | 0.000033447535664 | 0.000001057792923 |
| 13 | 0.00006873870825 | 0.000014850347479 | 0.000000352422288 |
| 14 | 0.00003433552970 | 0.000006595837194 | 0.000000117439213 |
| 15 | 0.00001715656503 | 0.000002930256391 | 0.000000039139454 |
| 16 | 0.00000857456764 | 0.000001301987044 | 0.000000013045098 |
| 17 | 0.00000428605007 | 0.000000578561445 | 0.000000004348089 |
| 18 | 0.00000214261490 | 0.000000257110062 | 0.000000001449308 |
| 19 | 0.00000107117102 | 0.000000114263049 | 0.000000000483092 |
| 20 | 0.00000053554010 | 0.000000050781269 | 0.000000000161028 |

**Table 3.2** The natural log of the expected length of longest cycles for $\theta = 0.25, 0.5, 0.75, 1, 2, 5, 10$.

| | 0.25 | 0.5 | 0.75 | 1 | 2 | 5 | 10 |
|---|---|---|---|---|---|---|---|
| 1 | -0.15343 | -0.27731 | -0.38132 | -0.47108 | -0.7431 | -1.21305 | -1.63535 |
| 2 | -2.14468 | -1.76662 | -1.62438 | -1.56265 | -1.54648 | -1.77137 | -2.07665 |
| 3 | -3.89321 | -3.01803 | -2.63482 | -2.42683 | -2.14032 | -2.15059 | -2.36161 |
| 4 | -5.56484 | -4.19004 | -3.56481 | -3.21036 | -2.65596 | -2.46095 | -2.58643 |
| 5 | -7.20466 | -5.32715 | -4.45789 | -3.95569 | -3.13191 | -2.73451 | -2.77873 |
| 6 | -8.82966 | -6.44695 | -5.33177 | -4.68044 | -3.58473 | -2.98521 | -2.95048 |
| 7 | -10.4473 | -7.55755 | -6.19496 | -5.3933 | -4.023 | -3.22044 | -3.10801 |
| 8 | -12.0611 | -8.66308 | -7.05193 | -6.09897 | -4.45164 | -3.44459 | -3.2551 |
| 9 | -13.673 | -9.76573 | -7.9052 | -6.80019 | -4.87372 | -3.66051 | -3.39421 |
| 10 | -15.2837 | -10.8667 | -8.75621 | -7.49858 | -5.29122 | -3.87014 | -3.52704 |
| 11 | -16.8939 | -11.9667 | -9.60583 | -8.19516 | -5.70547 | -4.07485 | -3.65479 |
| 12 | -18.5037 | -13.0662 | -10.4546 | -8.89056 | -6.11739 | -4.27567 | -3.77838 |
| 13 | -20.1133 | -14.1653 | -11.3028 | -9.5852 | -6.52761 | -4.47337 | -3.89849 |
| 14 | -21.7229 | -15.2642 | -12.1507 | -10.2793 | -6.93659 | -4.66852 | -4.01567 |
| 15 | -23.3324 | -16.363 | -12.9984 | -10.9731 | -7.34467 | -4.86159 | -4.13035 |
| 16 | -24.9419 | -17.4617 | -13.8459 | -11.6667 | -7.75208 | -5.05295 | -4.24288 |
| 17 | -26.5513 | -18.5604 | -14.6933 | -12.3601 | -8.15899 | -5.24289 | -4.35355 |
| 18 | -28.1608 | -19.659 | -15.5407 | -13.0535 | -8.56553 | -5.43165 | -4.4626 |
| 19 | -29.7702 | -20.7577 | -16.3881 | -13.7468 | -8.9718 | -5.61942 | -4.57024 |
| 20 | -31.3797 | -21.8563 | -17.2354 | -14.44 | -9.37787 | -5.80637 | -4.67662 |

**Figure 3.1** The expected length of the longest cycle, the second longest cycle, and the third longest cycle, for $0.1 < \theta < 100$



# Chapter 4

# Shortest Cycles

## 4.1 Moments of the Length of $r^{\text{th}}$ Shortest Cycle when $m > \theta$

We will follow the same approach as in the previous section.

### 4.1.1 Smallest Components in $C^{(z)}$

The $m^{\text{th}}$ moment of the length of the $r^{\text{th}}$ shortest cycle is

$$\mathbb{E}_z[(S_r)^m] = \sum_{j=1}^{\infty} j^m \int_{t_j(z)}^{t_{j+1}(z)} e^{-t} \frac{t^{r-1}}{(r-1)!}\, dt.$$

Make the same change of variable:

$$t_\infty(z) - t = \theta E(x) = \theta \int_x^\infty \frac{e^{-y}}{y}\, dy.$$

Let $z = e^{-s}$, $0 < s < \infty$. There exists $x_j(s)$ such that $(j-1)s \leq x_j(s) < js$ and

$$\sum_{k=j}^{\infty} \frac{e^{-ks}}{k} = E(x_j(s)).$$

So we have

$$\begin{aligned}
\mathbb{E}_z[(S_r)^m] &= \sum_{j=1}^{\infty} j^m \int_{x_j(s)}^{x_{j+1}(s)} e^{\theta E(x) - t_\infty} \frac{[t_\infty - \theta E(x)]^{r-1}}{(r-1)!} \frac{\theta e^{-x}}{x}\, dx \\
&= (1-z)^\theta \sum_{j=1}^{\infty} j^m \int_{x_j(s)}^{x_{j+1}(s)} \frac{[t_\infty - \theta E(x)]^{r-1}}{(r-1)!} \frac{\theta e^{\theta E(x) - x}}{x}\, dx.
\end{aligned}$$

Notice that

$$\sum_{j=1}^{\infty}[x_j(s)]^m \mu_j \leq \mathbb{E}_z[(S_r)^m] < \sum_{j=1}^{\infty}[x_{j+1}(s)]^m \mu_j,$$

with

$$\mu_j = \int_{x_j(s)}^{x_{j+1}(s)} e^{\theta E(x)} \frac{[t_\infty - E(x)]^{r-1}}{(r-1)!} \frac{\theta e^{-x}}{x}\, dx.$$

As we will see later, the only dominating term in $[t_\infty - E(x)]^{r-1}$ will be the leading $t_\infty^{r-1}$ term. Assuming $\theta < m$. This is an approximation of the Riemann integral as $s \to 0$. So $s/(1-z) \to 1$, $x_1(s) \to 0$ and

$$\lim_{z\to 1} \frac{(1-z)^m}{m!}\mathbb{E}_z^\theta[(S_r)^m] = (1-z)^\theta \ln\left(\frac{1}{1-z}\right)^{r-1} H_{r,m}^\theta.$$

where

$$H_{r,m}^\theta = \frac{\theta}{(r-1)!}\int_0^\infty \frac{x^{m-1}}{m!} e^{\theta E(x) - x}\, dx.$$

We rewrite this as

$$\lim_{z\to 1} \frac{(1-z)^{m-\theta}}{m!} \ln\left(\frac{1}{1-z}\right)^{-(r-1)} \mathbb{E}_z^\theta[(S_r)^m] = H_{r,m}^\theta.$$

We will now use Tauberian theorems. First,

$$\begin{aligned}
H_{r,m}^\theta &= \lim_{z\to 1} \frac{(1-z)^{m-\theta}}{m!} \ln\left(\frac{1}{1-z}\right)^{-(r-1)} \mathbb{E}_z^\theta[(S_r)^m] \\
&= \lim_{z\to 1} \frac{(1-z)^{m-\theta}}{m!} \ln\left(\frac{1}{1-z}\right)^{-(r-1)} \sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)}(1-z)^\theta z^n \mathbb{E}_n[(S_r)^m] \\
&= \lim_{z\to 1} \sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{m!n!\Gamma(\theta)}\mathbb{E}_n[(S_r)^m](1-z)^m \ln\left(\frac{1}{1-z}\right)^{-(r-1)} z^n.
\end{aligned}$$

Since the coefficients of $z^n$ are nonnegative, we have

$$\sum_{k=0}^{n} \frac{\Gamma(n+\theta)}{m!n!\Gamma(\theta)}\mathbb{E}_n[(S_r)^m] \sim \Gamma(m+1)^{-1} n^m \ln(n)^{r-1} H_{r,m}^\theta, \qquad n \to \infty.$$

Now, we will show that $\dfrac{\Gamma(n+\theta)}{m!n!\Gamma(\theta)} E_n[(S_r)^m]$ is nondecreasing in $n$. This is equivalent to

$$(n+1)\mathbb{E}_n[S_r^m] \leq (n+\theta)\mathbb{E}_{n+1}[S_r^m].$$

Follow the same argument as in the previous section,

$$\begin{aligned}
\mathbb{E}_n[S_r^m] &= \mathbb{E}_{n+1}[\mathbb{E}_n[S_r^m]] \\
&\leq \mathbb{E}_{n+1}\left[\frac{S_r}{n+\theta+1}(S_r-1)^m + \frac{n+\theta+1-S_r}{n+\theta+1}S_r^m\right] \\
&= \mathbb{E}_{n+1}[S_r^m] + \frac{\mathbb{E}_{n+1}[S_r[(S_r-1)^m - S_r^m]]}{n+\theta+1} \\
&\leq \frac{n+\theta+1-m}{n+\theta+1}\mathbb{E}_{n+1}[S_r^m]
\end{aligned}$$

So we have

$$\begin{aligned}
(n+1)\mathbb{E}_n[S_r^m] &\leq (n+\theta+1)\mathbb{E}_n[S_r^m] \\
&\leq (n+\theta+1-m)\mathbb{E}_{n+1}[S_r^m] \\
&\leq (n+\theta)\mathbb{E}_{n+1}[S_r^m],
\end{aligned}$$

as desired. Hence

$$\frac{\Gamma(n+\theta)}{m!n!\Gamma(\theta)}E_n[(S_r)^m] \sim \Gamma(m+1)^{-1}[mn^{m-1}\ln(n)^{r-1} + n^{m-1}\ln(n)^{r-2}]H_{r,m}^{\theta}, \quad n \to \infty.$$

$$\frac{n^{\theta}}{(\ln n)^{r-1}}E_n\left[\left(\frac{S_r}{n}\right)^m\right] \sim m\Gamma(\theta)H_{r,m}^{\theta}, \qquad n \to \infty.$$

In fact the right hand side is a constant, so

$$\lim_{n\to\infty}\frac{n^{\theta}}{(\ln n)^{r-1}}E_n\left[\left(\frac{S_r}{n}\right)^m\right] = m\Gamma(\theta)\frac{\theta}{(r-1)!}\int_0^{\infty}\frac{x^{m-1}}{m!}e^{\theta E(x)-x}\,dx.$$

Rewriting this we get

$$\lim_{n\to\infty}\frac{n^{\theta}}{(\ln n)^{r-1}}\mathbb{E}_n\left[\left(\frac{S_r}{n}\right)^m\right] = \frac{\Gamma(1+\theta)}{(r-1)!}\int_0^{\infty}\frac{x^{m-1}}{(m-1)!}e^{\theta E(x)-x}\,dx.$$

### 4.1.2 Special Cases

When $\theta = 1$ and $m \neq 1$. The $m^{\text{th}}$ moment of the $r^{\text{th}}$ shortest cycle (in a unbiased random permutation) satisfies

$$\lim_{n\to\infty}\frac{n}{(\ln n)^{r-1}}\mathbb{E}_n\left[\left(\frac{S_r}{n}\right)^m\right] = \frac{1}{(r-1)!}\int_0^{\infty}\frac{x^{m-1}}{(m-1)!}e^{E(x)-x}\,dx.$$

agreeing with Shepp and Lloyd's result.

When $\theta = 1/2$. The expected length of the shortest cycle is $(m = 1, r = 1)$

$$\lim_{n\to\infty} \sqrt{n}\mathbb{E}_n \left[ \left( \frac{S_r}{n} \right) \right] = \frac{\sqrt{\pi}}{2} \int_0^\infty e^{E(x)/2-x}\, dx.$$

agreeing with Pippenger's result.

### 4.1.3 Limitations

The argument above only works for $m > \theta$. Let's first investigate how it breaks down when $m \geq \theta$.

Recall that in the previous section the constant associated with the asymptotic form is

$$H^\theta_{r,m} = \frac{\theta}{(r-1)!} \int_0^\infty \frac{x^{m-1}}{m!} e^{\theta E(x)-x}\, dx.$$

When $m \geq \theta$, one can prove that this integral diverges. We need to analyze how the integral behaves as a function of $z$, but we lost that information when we did the Riemann-sum approximation, so we need to go back a step and approximate more accurately, as will be in done in the next section.

## 4.2 Moments of the Length of $r^{\text{th}}$ Shortest Cycle when $m = \theta$

Recall that we have

$$\mathbb{E}_z[(S_r)^m] = (1-z)^\theta \sum_{j=1}^\infty j^m \int_{x_j(s)}^{x_{j+1}(s)} \frac{[t_\infty - \theta E(x)]^{r-1}}{(r-1)!} \frac{\theta e^{\theta E(x)-x}}{x}\, dx.$$

As before, we will see that the only dominant term in $[t_\infty - \theta E(x)]^{r-1}$ is $t_\infty^{r-1}$. So

$$\mathbb{E}_z[(S_r)^m] \sim \frac{\theta^r}{(r-1)!} (1-z)^\theta \ln\left(\frac{1}{1-z}\right)^{r-1} \sum_{j=1}^\infty j^m \int_{x_j(s)}^{x_{j+1}(s)} \frac{e^{\theta E(x)-x}}{x}\, dx.$$

The integrand is a strictly decreasing function on the positive reals. Since $x_{j+1}(s) - x_j(s) \leq 2s$, we can make the following approximation:

$$\sum_{j=1}^\infty (js)^m \int_{x_j(s)}^{x_{j+1}(s)} \frac{e^{\theta E(x)-x}}{x}\, dx = \int_{x_1(s)}^\infty x^{m-1} e^{\theta E(x)-x}\, dx + \text{error}.$$

Recall that

$$E(x_1(s)) = \ln \frac{1}{1-z},$$

so we have the following lemma.

**Lemma 4.1**

$$\int_{x_1(s)}^{\infty} x^{m-1} e^{\theta E(x) - x}\, dx \sim \begin{cases} e^{-m\gamma} \ln\left(\frac{1}{1-z}\right) & \theta = m \\ \frac{e^{-\theta\gamma}}{\theta - m} \left(\frac{1}{1-z}\right)^{\theta - m} & \theta > m \end{cases}.$$

The error term is on the order of $(1-z)^m \ln(1/(1-z))$ if $\theta = m$, so it goes to zero as $s \to 0$. When $\theta > m$, the error term is on the order of $(1-z)^{2m-\theta}$, which does not always go to zero. So we first study the case $\theta = m$.

When $\theta = m$ we have

$$(1-z)^m \mathbb{E}_z[(S_r)^m] \sim \frac{\theta^r}{(r-1)!}(1-z)^\theta \ln\left(\frac{1}{1-z}\right)^{r-1} e^{-m\gamma} \ln\left(\frac{1}{1-z}\right).$$

This simplifies to

$$\mathbb{E}_z[(S_r)^m] \sim e^{-m\gamma} \frac{\theta^r}{(r-1)!} \ln\left(\frac{1}{1-z}\right)^r.$$

Therefore,

$$\begin{aligned} e^{-m\gamma} \frac{\theta^r}{(r-1)!} &= \lim_{z\to 1} \ln\left(\frac{1}{1-z}\right)^{-r} \mathbb{E}_z[(S_r)^m] \\ &= \lim_{z\to 1} \ln\left(\frac{1}{1-z}\right)^{-r} \sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)}(1-z)^\theta z^n \mathbb{E}_n[(S_r)^m] \\ &= \lim_{z\to 1} \sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)}(1-z)^\theta \ln\left(\frac{1}{1-z}\right)^{-r} z^n \mathbb{E}_n[(S_r)^m]. \end{aligned}$$

Since the coefficients of $z^n$ are nonnegative, we have

$$\sum_{k=0}^{n} \frac{\Gamma(k+\theta)}{k!\Gamma(\theta)} \mathbb{E}_n[(S_r)^m] \sim \Gamma(\theta+1)^{-1} n^\theta \ln(n)^r \frac{e^{-m\gamma}\theta^r}{(r-1)!}, \qquad n \to \infty.$$

We have shown before that $\frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} \mathbb{E}_n[(S_r)^m]$ is nondecreasing in $n$, so

$$\frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} \mathbb{E}_n[(S_r)^m] \sim \Gamma(\theta+1)^{-1} r n^{\theta-1} \ln(n)^r \frac{e^{-m\gamma}\theta^r}{(r-1)!}, \qquad n \to \infty.$$

Simplifying the expression we have

$$\mathbb{E}_n[(S_r)^m] \sim \frac{e^{-m\gamma}\theta^{r-1}}{r!}(\ln n)^r \qquad n \to \infty.$$

## 4.3 Moments of the Length of $r^{\text{th}}$ Shortest Cycle when $m < \theta$

For $m < \theta$, we will take a different approach. Let

$$T_{bn} = \sum_{j=b+1}^{n} jZ_j, \qquad 0 \le b < n.$$

**Lemma 4.13** Suppose that $m = m_n \in \mathbb{Z}_+$ satisfies $m/n \to y \in (0, \infty)$ as $n \to \infty$, and that $b = b_n = o(n)$. Then

$$n \Pr[T_{bn} = m] \sim p_\theta(y), \quad n \to \infty.$$

We look for the probability that the $r^{\text{th}}$ shortest cycle is greater than $b$.

$$\begin{aligned}\Pr[S_r > b] &= \Pr[C_1 + \cdots + C_b < r] \\ &= \Pr[Z_1 + \cdots + Z_b < r]\frac{\Pr[T_{bn} = n - T_{0b}]}{\Pr[T_{0n} = n]}\end{aligned}$$

The first term goes to

$$e^{-\lambda}\sum_{s=0}^{r-1}\frac{\lambda^s}{s!}, \qquad \lambda = \theta\sum_{k=1}^{b}\frac{1}{k}.$$

Let $H(k)$ be the $k^{\text{th}}$ harmonic number; then

$$\Pr[S_r > b] = e^{-\theta H(b)}\sum_{s=0}^{r-1}\frac{(\theta H(b))^s}{s!}.$$

Now,

$$
\begin{aligned}
\mathbb{E}[S_r^m] &= \sum_{k=1}^{\infty} k^m \Pr[S_r = k] \\
&= \sum_{k=1}^{\infty} k^m (\Pr[S_r > k-1] - \Pr[S_r > k]) \\
&= 1 + \sum_{k=1}^{\infty} [(k+1)^m - k^m] \Pr[S_r > k] \\
&\to 1 + \sum_{k=1}^{\infty} [(k+1)^m - k^m] e^{-\theta H(k)} \sum_{s=0}^{r-1} \frac{(\theta H(k))^s}{s!}.
\end{aligned}
$$

The summand goes to

$$
(m-1)k^{m-1}k^{-\theta}\frac{(\theta \log k)^{r-1}}{(r-1)!} \in \Theta(k^{m-\theta-1}).
$$

If $\theta > m$, then $m - \theta - 1 < -1$ so the sum converges. This means that $\mathbb{E}[S_r^m]$ converges to a constant (depending on $r$, $\theta$ and $m$).

## 4.4 Numerical Results

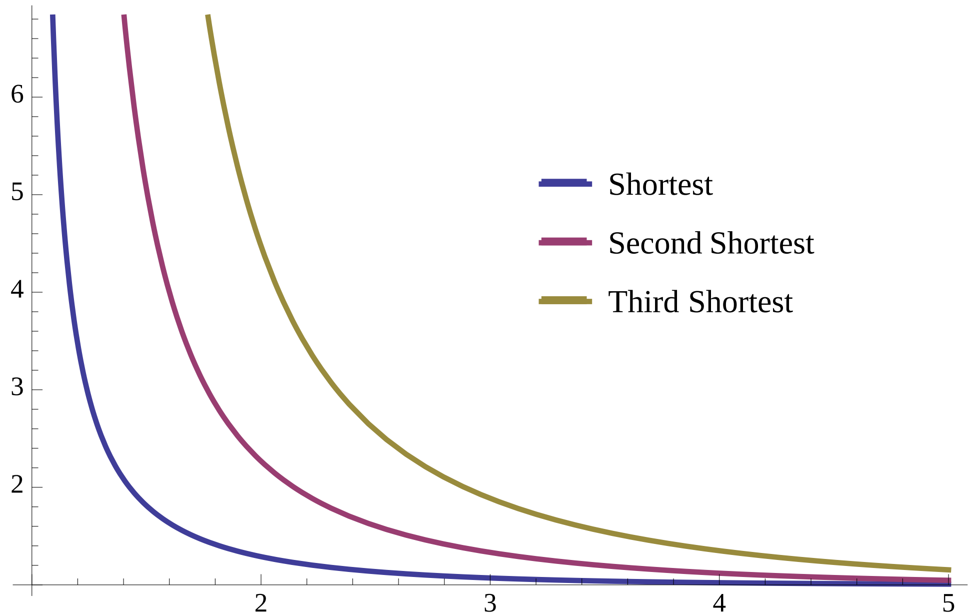
In this section we summarizes some numerical results based on the derivation above.

Notice that on average, the ratio of the $(r+1)^{\text{th}}$ cycle and the $r^{\text{th}}$ cycle length goes to $\dfrac{\theta}{\theta - 1}$.

**Table 4.1** The expected length of shortest cycles for $\theta = 2, 5, 10$.

| | 2 | 5 | 10 |
|---|---|---|---|
| 1 | 1.288959430 | 1.007446918 | 1.000045718 |
| 2 | 2.268155928 | 1.046817690 | 1.000504528 |
| 3 | 4.474205650 | 1.154242096 | 1.002811008 |
| 4 | 8.966061894 | 1.359368143 | 1.010562612 |
| 5 | 17.95570554 | 1.674805931 | 1.030184791 |
| 6 | 35.92668346 | 2.102451450 | 1.070180329 |
| 7 | 71.86249650 | 2.647148053 | 1.138787134 |
| 8 | 143.7307614 | 3.325670126 | 1.241163552 |
| 9 | 287.4653367 | 4.168363732 | 1.377795642 |
| 10 | 574.9332279 | 5.217553080 | 1.545089138 |
| 11 | 1149.868166 | 6.526791700 | 1.737698143 |
| 12 | 2299.737473 | 8.162286822 | 1.951195210 |
| 13 | 4599.475708 | 10.20608077 | 2.183787419 |
| 14 | 9198.951925 | 12.76037803 | 2.436613707 |
| 15 | 18397.90419 | 15.95284011 | 2.712993706 |
| 16 | 36795.80860 | 19.94304375 | 3.017348891 |
| 17 | 73591.61735 | 24.93047162 | 3.354371592 |
| 18 | 147183.2348 | 31.16447967 | 3.728658414 |
| 19 | 294366.4697 | 38.95675830 | 4.144729927 |
| 20 | 588732.9394 | 48.69691377 | 4.607242229 |

**Figure 4.1** The expected length of the shortest cycle, the second shortest cycle, and the third shortest cycle, for $1 < \theta < 5$




## 4.5 Joint Moments of Shortest Cycles: $m = m_1 + \cdots + m_r > \theta + r - 1$

Joint moments of longest cycles have been studied using the Poisson-Dirichlet distribution. However, the Poisson-Dirichlet distribution only characterizes the relative size of longest cycles in a large random permutation. In this section, we will derive the joint moments of shortest cycles in a special case when $m = m_1 + \cdots + m_r > \theta + r - 1$.

Since a Poisson process has independent increment, we get

$$\mathbb{E}[S_1^{m_1} \cdots S_r^{m_r}] = \sum_{j_r \geq \cdots \geq j_1 = 1}^{\infty} j_1^{m_1} \cdots j_r^{m_r} \int_{t_{j_1}(z)}^{t_{j_1+1}(z)} \cdots \int_{t_{j_r}(z)}^{t_{j_r+1}(z)} e^{-s_r} ds_r \cdots ds_1.$$

Using the same change of variable, we have

$$\mathbb{E}[S_1^{m_1} \cdots S_r^{m_r}] = (1-z)^{\theta} \sum_{j_r \geq \cdots \geq j_1 = 1}^{\infty} j_1^{m_1} \cdots j_r^{m_r} \int \frac{e^{\theta E(s_r) - s_1 - \cdots - s_r}}{s_1 \cdots s_r} \theta^r ds.$$

Assuming $m = m_1 + \cdots + m_r > \theta + r - 1$. Taking $z \to 1$, we have

$$(1-z)^m \mathbb{E}[S_1^{m_1} \cdots S_r^{m_r}] \sim \theta^r (1-z)^\theta \int x_1^{j_1-1} \cdots x_r^{j_r-1} e^{\theta E(s_r) - s_1 - \cdots - s_r} dx_1 \cdots dx_r.$$

Let

$$H_{\theta, m_1, \cdots, m_r} = \int x_1^{j_1-1} \cdots x_r^{j_r-1} e^{\theta E(s_r) - s_1 - \cdots - s_r} dx_1 \cdots dx_r.$$

Then

$$\theta^r H_{\theta, m_1, \cdots, m_r} = \sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} (1-z)^m z^n E_n[S_1^{m_1} \cdots S_r^{m_r}].$$

Tauberian theorems give us

$$\sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} E_n[S_1^{m_1} \cdots S_r^{m_r}] \sim \frac{H_{\theta, m_1, \cdots, m_r}}{m!} \theta^r n^m.$$

Thus,

$$\frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} E_n[S_1^{m_1} \cdots S_r^{m_r}] \sim \frac{H_{\theta, m_1, \cdots, m_r}}{(m-1)!} \theta^r n^{m-1}.$$

Applying

$$\lim_{n\to\infty} \frac{n! n^{\theta-1}}{\Gamma(n+\theta)} = 1,$$

we have

$$\frac{1}{\Gamma(\theta)} E_n[S_1^{m_1} \cdots S_r^{m_r}] \sim \theta^r \frac{H_{\theta, m_1, \cdots, m_r}}{(m-1)!} n^{m-\theta}.$$

We can rewrite this as

$$E_n\left[\left(\frac{S_1}{n}\right)^{m_1} \cdots \left(\frac{S_r}{n}\right)^{m_r}\right] \sim \frac{\theta^r \Gamma(\theta)}{\Gamma(m)} n^{-\theta} \int x_1^{j_1-1} \cdots x_r^{j_r-1} e^{\theta E(s_r) - s_1 - \cdots - s_r} dx_1 \cdots dx_r.$$

As mentioned above, this holds for $m = m_1 + \cdots + m_r > \theta + r - 1$.

# Chapter 5

# Correlations

We have already described the asymptotic behavior of longest and shortest cycles. In this chapter we study the correlations among them. For example, one question we may ask is, what's the correlation between the longest cycle and the second longest cycle in a random permutation? We will see that the answer is not as simple as we expect.

## 5.1 Correlations for Longest Cycles

We have already shown that all longest cycles grow as $\Theta(n)$. We will use $L_r$ to describe the normalized length of $r^{\text{th}}$ longest cycle.

In this section we will study the correlation coefficient between $L_{r_1}$ and $L_{r_2}$. First, we can use the same method as above to find an expression for $\mathbb{E}[L_{r_1}^{m_1} L_{r_2}^{m_2}]$.

$$\begin{aligned}\mathbb{E}[L_{r_1}^{m_1} L_{r_2}^{m_2}] =& \frac{\theta^{r_2}\Gamma(\theta)}{\Gamma(\theta+m)} \\ & \int x_1^{m_1-1} x_2^{m_2-1} \frac{E(x_1)^{r_1-1}[E(x_2)-E(x_1)]^{r_2-r_1}}{(r_1-1)!(r_2-1)!} e^{-E(x_2)-x_1-x_2} dx_1 dx_2.\end{aligned}$$

In this section we are interested in the case $m_1 = m_2 = 1$, so

$$\mathbb{E}[L_{r_1} L_{r_2}] = \frac{\theta^{r_2-1}}{(\theta+1)} \int \frac{E(x_1)^{r_1-1}[E(x_2)-E(x_1)]^{r_2-r_1}}{(r_1-1)!(r_2-1)!} e^{-E(x_2)-x_1-x_2} dx_1 dx_2.$$

We already know that

$$\mathbb{E}[L_r] = \theta^{r-1} \int_0^\infty \frac{E(x)^{r-1}}{(r-1)!} e^{-x-\theta E(x)}\, dx.$$

**Table 5.1** Correlation among Longest Cycles when $\theta = 1$

| $L_{r_1} \backslash L_{r_2}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | -0.76 | -0.78 | -0.68 | -0.58 | -0.5 | -0.42 | -0.36 | -0.31 | -0.27 |
| 2 | | 1 | 0.36 | 0.16 | 0.09 | 0.05 | 0.03 | 0.02 | 0.02 | 0.01 |
| 3 | | | 1 | 0.62 | 0.44 | 0.34 | 0.27 | 0.22 | 0.18 | 0.16 |
| 4 | | | | 1 | 0.72 | 0.56 | 0.45 | 0.37 | 0.31 | 0.27 |
| 5 | | | | | 1 | 0.78 | 0.63 | 0.52 | 0.43 | 0.37 |
| 6 | | | | | | 1 | 0.81 | 0.66 | 0.56 | 0.47 |
| 7 | | | | | | | 1 | 0.82 | 0.69 | 0.58 |
| 8 | | | | | | | | 1 | 0.84 | 0.71 |
| 9 | | | | | | | | | 1 | 0.84 |
| 10 | | | | | | | | | | 1 |

**Table 5.2** Correlation among Longest Cycles when $\theta = 0.5$

| $L_{r_1} \backslash L_{r_2}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | -0.89 | -0.75 | -0.57 | -0.43 | -0.32 | -0.24 | -0.18 | -0.14 | -0.1 |
| 2 | | 1 | 0.42 | 0.23 | 0.15 | 0.11 | 0.08 | 0.06 | 0.04 | 0.03 |
| 3 | | | 1 | 0.6 | 0.41 | 0.29 | 0.21 | 0.16 | 0.12 | 0.09 |
| 4 | | | | 1 | 0.67 | 0.48 | 0.35 | 0.26 | 0.19 | 0.14 |
| 5 | | | | | 1 | 0.71 | 0.51 | 0.38 | 0.28 | 0.21 |
| 6 | | | | | | 1 | 0.72 | 0.53 | 0.39 | 0.29 |
| 7 | | | | | | | 1 | 0.73 | 0.54 | 0.4 |
| 8 | | | | | | | | 1 | 0.74 | 0.55 |
| 9 | | | | | | | | | 1 | 0.74 |
| 10 | | | | | | | | | | 1 |

$$\mathbb{E}[L_r^2] = \frac{\theta^{r-1}}{\theta+1} \int_0^\infty \frac{E(x)^{r-1}}{(r-1)!} x e^{-x-\theta E(x)}\, dx.$$

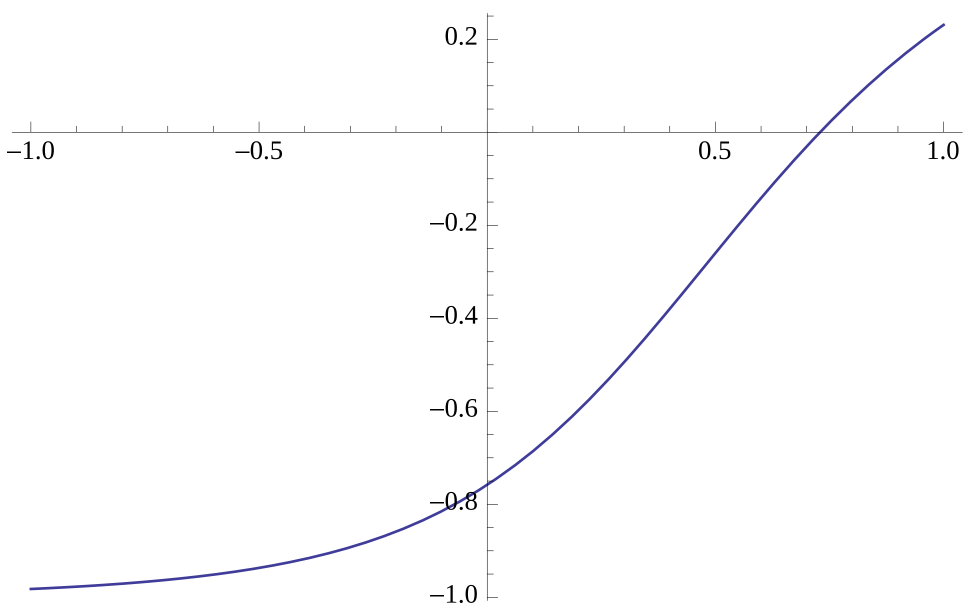
With these formulae, we can calculate the correlation coefficients between $L_{r_1}$ and $L_{r_2}$. The results are summarized below.

Notice that (especially when $\theta$ is small), the length of the longest cycle is negatively correlated with the cycle lengths of the rest. This is due to the fact that the longest cycle usually occupies a significant portion of the size of permutation. The following is a plot of the correlation between the longest cycle and the second longest cycle across different $\theta$'s.

**Table 5.3** Correlation among Longest Cycles when $\theta = 2$

| $L_{r_1} \backslash L_{r_2}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | -0.49 | -0.73 | -0.72 | -0.68 | -0.62 | -0.56 | -0.52 | -0.47 | -0.43 |
| 2 | | 1 | 0.25 | 0. | -0.09 | -0.13 | -0.15 | -0.15 | -0.14 | -0.14 |
| 3 | | | 1 | 0.57 | 0.37 | 0.26 | 0.19 | 0.15 | 0.12 | 0.1 |
| 4 | | | | 1 | 0.71 | 0.54 | 0.44 | 0.37 | 0.32 | 0.28 |
| 5 | | | | | 1 | 0.78 | 0.65 | 0.55 | 0.48 | 0.42 |
| 6 | | | | | | 1 | 0.83 | 0.71 | 0.62 | 0.55 |
| 7 | | | | | | | 1 | 0.86 | 0.75 | 0.67 |
| 8 | | | | | | | | 1 | 0.87 | 0.78 |
| 9 | | | | | | | | | 1 | 0.89 |
| 10 | | | | | | | | | | 1 |

**Figure 5.1** Correlation between Lengths of the Longest Cycle and the Second Longest Cycle, $\theta \in [0.1, 10]$. The horizontal axis is $10^{\theta}$. The vertical axis is the correlation coefficient.



## 5.2 Correlations for Shortest Cycles when $\theta < 1$

In this section we will study the correlation coefficient between $S_{r_1}$ and $S_{r_2}$, where $1 \leq r_1 < r_2$. First we will derive a formula for $\mathbb{E}[S_{r_1}^{m_1} S_{r_2}^{m_2}]$.

$$\mathbb{E}_z[S_{r_1}^{m_1} S_{r_2}^{m_2}] = (1-z)^\theta \sum_{1 \leq j_1 < j_2} j_1^{m_1} j_2^{m_2} \int_{0<x_1<x_2} \frac{[t_\infty - \theta E(x_1)]^{r_1-1}}{(r_1-1)!}$$
$$\frac{[\theta E(x_2) - \theta E(x_1)]^{r_2-r_1-1}}{(r_2-r_1-1)!} \frac{\theta^2 e^{\theta E(x_2)-x_1-x_2}}{x_1 x_2}\, dx_1 dx_2.$$

As $z \to 1$, we have

$$(1-z)^m \mathbb{E}_z[S_{r_1}^{m_1} S_{r_2}^{m_2}] = (1-z)^\theta \ln\left(\frac{1}{1-z}\right)^{r_1-1} \frac{\theta^{r_2}}{(r_1-1)!(r_2-r_1-1)!}$$
$$\int_{0<x_1<x_2} x_1^{m_1-1} x_2^{m_2-1} e^{\theta E(x_2)-x_1-x_2}\, dx_1 dx_2.$$

Using the relation between $\mathbb{E}_z$ and $\mathbb{E}_n$, we have

$$K = \frac{\theta^{r_2}}{(r_1-1)!(r_2-r_1-1)!} \int_{0<x_1<x_2} x_1^{m_1-1} x_2^{m_2-1} e^{\theta E(x_2)-x_1-x_2}\, dx_1 dx_2$$
$$= (1-z)^{m-\theta} \mathbb{E}_z[S_{r_1}^{m_1} S_{r_2}^{m_2}] \ln\left(\frac{1}{1-z}\right)^{-(r_1-1)}$$
$$= \sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} E_n[S_{r_1}^{m_1} S_{r_2}^{m_2}] \ln\left(\frac{1}{1-z}\right)^{-(r_1-1)} (1-z)^m z^n$$

Applying Tauberian theorems [deBrujin],

$$\sum_{n=0}^{\infty} \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} \mathbb{E}_n[S_{r_1}^{m_1} S_{r_2}^{m_2}] \sim \frac{K}{\Gamma(m+1)} n^m [\ln(n)]^{r_1-1}.$$

$$\lim_{n\to\infty} \frac{n^{\theta-1}}{\Gamma(\theta)} \mathbb{E}_n[S_{r_1}^{m_1} S_{r_2}^{m_2}] \sim \frac{K}{\Gamma(m)} n^{m-1} [\ln(n)]^{r_1-1}.$$

Therefore,

$$\mathbb{E}_n\left[\frac{S_{r_1}^{m_1} S_{r_2}^{m_2}}{n^m}\right] \sim \frac{K\Gamma(\theta)}{\Gamma(m)} n^{-\theta} [\ln(n)]^{r_1-1}$$
$$\sim \frac{\Gamma(\theta)}{\Gamma(m)} \frac{\theta^{r_2}}{(r_1-1)!(r_2-r_1-1)!}$$
$$\left(\int_{0<x_1<x_2} x_1^{m_1-1} x_2^{m_2-1} e^{\theta E(x_2)-x_1-x_2}\, dx_1 dx_2\right) n^{-\theta} [\ln(n)]^{r_1-1}.$$

Hence,

$$\mathbb{E}_n\left[\frac{S_{r_1}S_{r_2}}{n^2}\right] \sim \frac{\theta^{r_2}\Gamma(\theta)}{\Gamma(r_1)\Gamma(r_2-r_1)}\left(\int_{0<x_1<x_2} e^{\theta E(x_2)-x_1-x_2}\, dx_1 dx_2\right) n^{-\theta}[\ln(n)]^{r_1-1}.$$

Recall that

$$\mathbb{E}_n\left[\frac{S_r}{n}\right] \sim \frac{\Gamma(1+\theta)}{\Gamma(r)}\left(\int_0^\infty e^{\theta E(x)-x}\, dx\right) n^{-\theta}[\ln(n)]^{r-1}.$$

$$\mathbb{E}_n\left[\left(\frac{S_r}{n}\right)^2\right] \sim \frac{\Gamma(1+\theta)}{\Gamma(r)}\left(\int_0^\infty x e^{\theta E(x)-x}\, dx\right) n^{-\theta}[\ln(n)]^{r-1}.$$

From here we can see that the covariance of $S_{r_1}/n$ and $S_{r_2}/n$ grows as $\Theta(n^{-\theta}[\ln(n)]^{r_1-1})$, but the product of their variances is $\Theta(n^{-2\theta}[\ln(n)]^{r_1+r_2-2})$. Since $2(r_1-1) < r_1+r_2-2$, we conclude that the correlation coefficient between $S_{r_1}/n$ and $S_{r_2}/n$ goes to zero as $n \to \infty$.

# Chapter 6

# Limit Theorems for the Exponents

Looking back at the results we have presented so far, we see that the moments for the longest cycles in a $\theta$-biased random permutations are much simpler than the moments for the shortest cycles. In particular, the $m^{\text{th}}$ moment of the length of the $r^{\text{th}}$ longest cycle is always $\Theta(n^m)$, but the $m^{\text{th}}$ moment of the length of the $r^{\text{th}}$ shortest cycle has a complicated asymptotic expressions depending on $m$, $\theta$ and $r$. These results are unsatisfactory considering the simplicity of the problem. In this chapter, we change our perspective to look at the exponents of these cycle lengths, and the results are much simpler and elegant.

## 6.1 A Caveat

So far we have mostly focused on the expected length of ordered cycles in a large random permutation. We have seen that the expected length of the shortest cycle in a random permutation with $\theta = 1$ is $\Theta(\log n)$, which is much smaller than that in a random permutation with $\theta = 0.5$, which is $\Theta(\sqrt{n})$. However, it turns out that with probability one, the shortest cycle in a $\theta = 0.5$ random permutation will have length $o(n^\epsilon)$ for any $\epsilon > 0$. The reason why we have such a high expected value for the shortest cycle length is that the distribution has a heavy tail.

We can illustrate this through an example. Consider a random variable

$X$, such that

$$X = \begin{cases} n^{1/3} & \text{with probability } 1 - \dfrac{1}{n^{1/3}} \\ n & \text{with probability } \dfrac{1}{n^{1/3}} \end{cases}.$$

Then $X \sim n^{1/3}$ with probability 1 as $n \to \infty$. But $\mathbb{E}[X] = n^{1/3} - 1 + n^{2/3} \to n^{2/3}$. So the expectation is not a good measure of the tendency in this case. So instead we look at the exponent of $X$ defined as below:

**Definition 6.1** *Let $X_n$ be a function of $n$. Define*

$$X_n^{\#} = \frac{\log X_n}{\log n}$$

*and call it the exponent of $X_n$.*

For our example above, we have

$$X^{\#} = \begin{cases} 1/3 & \text{with probability } 1 - \dfrac{1}{n^{1/3}} \\ 1 & \text{with probability } \dfrac{1}{n^{1/3}} \end{cases}.$$

Then $X^{\#} \sim 1/3$ with probability 1 as $n \to \infty$, and $\mathbb{E}[X^{\#}]$ also goes to $1/3$. Here the expected value is a better indicator because $X^{\#}$ is now bounded.

Since $(L_1/n, L_2/n, \cdots)$ follows the Poisson-Dirichlet distribution, we immediately have the following result:

**Proposition 6.1** *In a $\theta$-biased random permutation, $L_r^{\#} \to 1$ with probability one as $n \to \infty$.*

## 6.2 Shortest Cycles

We first define $T_{bn}$ to be the total count of elements in $z$-cycles with size larger than $b$:

$$T_{bn} = \sum_{j=b+1}^{n} jZ_j, \qquad 0 \le b < n. \tag{6.1}$$

**Lemma 4.13(Arratia et al., 2003)** Suppose that $m = m_n \in \mathbb{Z}_+$ satisfies $m/n \to y \in (0, \infty)$ as $n \to \infty$, and that $b = b_n = o(n)$. Then

$$n \Pr[T_{bn} = m] \sim p_\theta(y), \quad n \to \infty.$$

Here, $p_\theta$ is the limit of the density of the random variable $n^{-1}T_{0n}$ as $n \to \infty$. We look for the probability that the shortest cycle is greater than $n^p$:

$$\begin{aligned}\Pr[S_1 > n^p] &= \Pr[C_1 = \cdots = C_{n^p} = 0] \\ &= \Pr[Z_1 = \cdots = Z_{n^p} = 0] \frac{\Pr[T_{(n^p+1)n} = n]}{\Pr[T_{0n} = n]}.\end{aligned} \tag{6.2}$$

Assume $p < 1$; then both the numerator and the denominator are asymptotic to $n^{-1}p_\theta(y)$. The first factor is

$$e^{-\theta \sum_{k=1}^{n^p} \frac{1}{k}} \sim e^{-\theta p \log n} = n^{-p\theta}.$$

So as long as $\theta > 0$ and $p > 0$, this probability goes to zero. Therefore, $S_1^\# = \frac{\log S_1}{\log n}$ is concentrated at zero, so we have

$$S_1^\# \to 0 \text{ with probability one as } n \to \infty.$$

Similarly,

$$\begin{aligned}
&\Pr[S_r > n^p] \\
&= \Pr[C_1 + \cdots + C_{n^p} < r] \\
&= \frac{\Pr[Z_1 + \cdots + Z_{n^p} < r, T_{(n^p+1)n} = n - T_{0n^p}]}{\Pr[T_{0n} = n]} \\
&= \frac{\sum_{j=0}^{(r-1)n^p} \Pr[Z_1 + \cdots + Z_{n^p} < r, T_{(n^p+1)n} = n - T_{0n^p} | T_{0n^p} = j] \Pr[T_{0n^p} = j]}{\Pr[T_{0n} = n]} \\
&= \frac{1}{\Pr[T_{0n} = n]} \times \\
&\quad \sum_{j=0}^{(r-1)n^p} \Pr[Z_1 + \cdots + Z_{n^p} < r | T_{0n^p} = j] \Pr[T_{(n^p+1)n} = n - T_{0n^p} | T_{0n^p} = j] \Pr[T_{0n^p} = j] \\
&= \sum_{j=0}^{(r-1)n^p} \Pr[Z_1 + \cdots + Z_{n^p} < r | T_{0n^p} = j] \Pr[T_{0n^p} = j] \\
&= \Pr[Z_1 + \cdots + Z_{n^p} < r].
\end{aligned}$$

This goes to

$$e^{-\lambda} \sum_{s=0}^{r-1} \frac{\lambda^s}{s!}, \qquad \lambda = \theta \sum_{k=1}^{n^p} \frac{1}{k} = \theta H(n^p).$$

The last term in the expansion dominates, so

$$\Pr[S_r > n^p] \to n^{-p\theta} \frac{(\theta H(n^p))^{r-1}}{(r-1)!} \to 0 \tag{6.3}$$

Therefore, we have the following proposition:

**Proposition 6.2** *In a $\theta$-biased random permutation, $S_r^\# \to 0$ with probability one as $n \to \infty$.*

## 6.3 Cycle Lengths at the $100q^{\text{th}}$ percentile

We look for the probability that the $100q$ percentile cycle is greater than $n^p$.

$$\begin{aligned}\Pr[M_q > n^p] &= \Pr[C_1 + \cdots + C_{n^p} < qK_{0n}] \\ &= \Pr[Z_1 + \cdots + Z_{n^p} < qK_{0n}] \frac{\Pr[T_{(n^p)n} = n - T_{0n^p}]}{\Pr[T_{0n} = n]}\end{aligned}$$

Following a similar derivation as in the previous section, the ratio goes to one because $0 < T_{0n^p} < qK_{0n}n^p = o(n)$ with probability 1. Therefore,

$$\Pr[M_q > n^p] = e^{-\lambda} \sum_{s=0}^{qK_{0n}} \frac{\lambda^s}{s!}, \qquad \lambda = \theta \sum_{k=1}^{n^p} \frac{1}{k}. \tag{6.4}$$

This approaches

$$e^{-\lambda} \sum_{s=0}^{q\theta \log n} \frac{\lambda^s}{s!}, \qquad \lambda = p\theta \log n.$$

Notice that the density of a Poisson distribution with mean $p\theta \log n$ has standard deviation $\sqrt{p\theta \log n}$ which is $o(\theta \log n)$, so the probability above goes to 1 if $q > p$ and goes to 0 if $q < p$.

Therefore, the log of the length of the $100p$ percentile cycle is concentrated at $n^p$.

**Proposition 6.3** *In a $\theta$-biased random permutation, $M_p^\# \to p$ with probability one as $n \to \infty$.*

# Chapter 7

# Unified Description of Ordered Cycle Lengths

After studying the exponents of ordered cycle lengths, we now come back to study the expected ordered cycle lengths and attempt to give a unified description for the asymptotic behavior of ordered cycle lengths.

## 7.1 Shortest Cycles

In Chapter 4 we have thoroughly studied the expected lengths of the shortest cycles for all possible $\theta$. Recall that

$$\mathbb{E}[S_r] = \begin{cases} \Theta(n^{1-\theta}(\ln n)^{r-1}) & 0 < \theta < 1 \\ \Theta((\ln n)^r) & \theta = 1 \\ \Theta(1) & \theta > 1. \end{cases} \tag{7.1}$$

We write it as

$$\mathbb{E}[S_r]^\# \to \begin{cases} 1-\theta & 0 < \theta < 1 \\ 0 & \theta \geq 1. \end{cases} \tag{7.2}$$

Notice that $\mathbb{E}[\bullet^\#] \leq \mathbb{E}[\bullet]^\#$ by Jensen's Inequality. We now explain why the formula is as it is. Recall Equation 6.3:

$$\Pr[S_r > n^p] \to n^{-p\theta}\frac{(\theta H(n^p))^{r-1}}{(r-1)!} \to e^{-\gamma}n^{-p\theta}\frac{(p\theta \ln n)^{r-1}}{(r-1)!}. \tag{7.3}$$

Define

$$F_{S_r^\#}(p) = \Pr[S_r \leq n^p]. \tag{7.4}$$

Then,

$$\mathbb{E}[S_r] = \int_0^1 f_{S_r^\#}(p) n^p \, dp + O(1). \tag{7.5}$$

We have

$$\begin{aligned} f_{S_r^\#}(p) &= \frac{d}{dp}\left(e^{-\gamma} n^{-p\theta} \frac{(p\theta \ln n)^{r-1}}{(r-1)!}\right) \\ &= e^{-\gamma} \frac{n^{-p\theta}\theta \ln n (p\theta \ln n)^{r-2}(r-1-p\theta \ln n)}{(r-1)!} \\ &\to e^{-\gamma} \frac{n^{-p\theta}(\theta p \ln n)^r}{p(r-1)!}. \end{aligned}$$

Hence,

$$\int_0^1 f_{S_r^\#}(p) n^p \, dp = \int_0^1 \frac{e^{-\gamma} n^{-p\theta}(\theta p \ln n)^r}{p(r-1)!} n^p \, dp. \tag{7.6}$$

**Lemma 7.1** *Let $f_n(p)$ be defined from $[0,1]$ to the set of functions on $n$. Then*

$$\left[\int_0^1 f_n(p) \, dp\right]^\# \to \max_{p\in[0,1]} f_n^\#(p).$$

For Equation 7.6, we have

$$f_n(p) = \frac{e^{-\gamma} n^{-p\theta}(\theta p \ln n)^r}{p(r-1)!} n^p.$$

Thus,

$$f_n^\#(p) \to p(1-\theta).$$

$$\mathbb{E}[S_r]^\# = \max_p f_n^\#(p) \to \begin{cases} 1-\theta & \theta < 1 \\ 0 & \theta \geq 1. \end{cases}$$

This agrees with Equation 7.2.

## 7.2 Cycle Lengths at the $100q^{\text{th}}$ percentile

Recall from Equation 6.4 that

$$\Pr[M_q > n^p] \to e^{-\lambda} \sum_{s=0}^{q \ln n} \frac{\lambda^s}{s!}, \qquad \lambda = \theta H(n^p).$$

As in the previous section, we define

$$F_{M_q^\#}(p) = \Pr[M_q \le n^p].$$

Then,

$$\mathbb{E}[M_q] = \int_0^1 f_{M_q^\#}(p) n^p \, dp + O(1).$$

We have

$$\begin{aligned} f_{M_q^\#}(p) &= -\frac{d}{dp}\left(e^{-\lambda}\sum_{s=0}^{qK_{0n}} \frac{\lambda^s}{s!}\right) \\ &= -e^{-\lambda}(-\theta \ln n) \sum_{s=0}^{q\theta \ln n} \frac{\lambda^s}{s!} - e^{-\lambda} \sum_{s=0}^{q\theta \ln n - 1} \frac{\lambda^s}{s!}(\theta \ln n) \\ &= e^{-\lambda}(\theta \ln n) \frac{\lambda^{q\theta \ln n}}{(q\theta \ln n)!}. \end{aligned}$$

We will use Lemma 7.1, and the Stirling approximation $\ln n! = n \ln n - n + O(\ln n)$,

$$\begin{aligned} f_{M_q^\#}^\#(p) &= \frac{1}{\log n}\left(-\lambda + \ln\theta + \ln\ln n + q\theta \ln n \ln \lambda - \ln\left(q\theta \ln n\right)!\right) \\ &\to \frac{-(p\theta \ln n + \theta\gamma) + q\theta \ln n \ln\left(\theta p \ln n + \theta\gamma\right)) - q\theta \ln n \ln(q\theta \ln n) - q\theta \ln n}{\ln n} \\ &= -p\theta + q\theta \ln\left(p\theta \ln n\right)) - q\theta \ln(q\theta \ln n) - q\theta \\ &= -p\theta + q\theta\left(1 + \ln\frac{p}{q}\right). \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}[M_q]^\# &\to \max_p f_{M_q^\#}^\#(p) n^p = \max_p p(1-\theta) + q\theta\left(1 + \ln\frac{p}{q}\right) \\ &= \begin{cases} (1-\theta) + q\theta(1 - \ln q), & \theta \le \dfrac{1}{1-q}, \text{ achieved at } p = 1 \\ q\theta \ln\dfrac{\theta}{\theta - 1}, & \theta > \dfrac{1}{1-q}, \text{ achieved at } p = \dfrac{q\theta}{\theta - 1}. \end{cases} \end{aligned} \tag{7.7}$$

Similarly, we can find the moments of these ordered cycle lengths:

$$\mathbb{E}[M_q^m]^\# \to \begin{cases} (m-\theta) + q\theta(1 - \ln q), & \theta \le \dfrac{1}{m-q}, \text{ achieved at } p = 1 \\ q\theta \ln\dfrac{\theta}{\theta - m}, & \theta > \dfrac{1}{m-q}, \text{ achieved at } p = \dfrac{q\theta}{\theta - m}. \end{cases} \tag{7.8}$$

## 7.3 Visualization

Based on the results in the previous section, we create the following graphs for visualization of the asymptotic behavior of ordered cycle lengths in a random permutation.

**Figure 7.1** The value of $\mathbb{E}^{\#}[M_q]$ for $\theta = 0.1, 0.5, 1, 2, 10$. The horizontal axis is $q$. The vertical axis is $\mathbb{E}^{\#}[M_q]$.

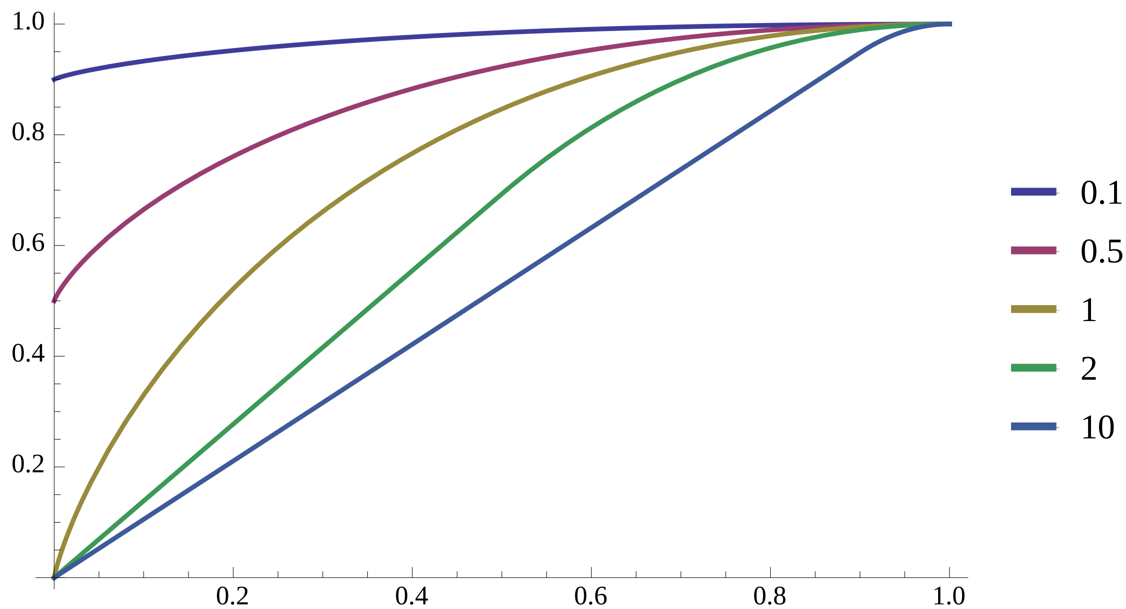


**Figure 7.2** The value of $\mathbb{E}[M_q]^\#$ for $0 < \theta < 5$. The horizontal axes represent $\theta$ and $q$. The vertical axis is $M_q^\#$.

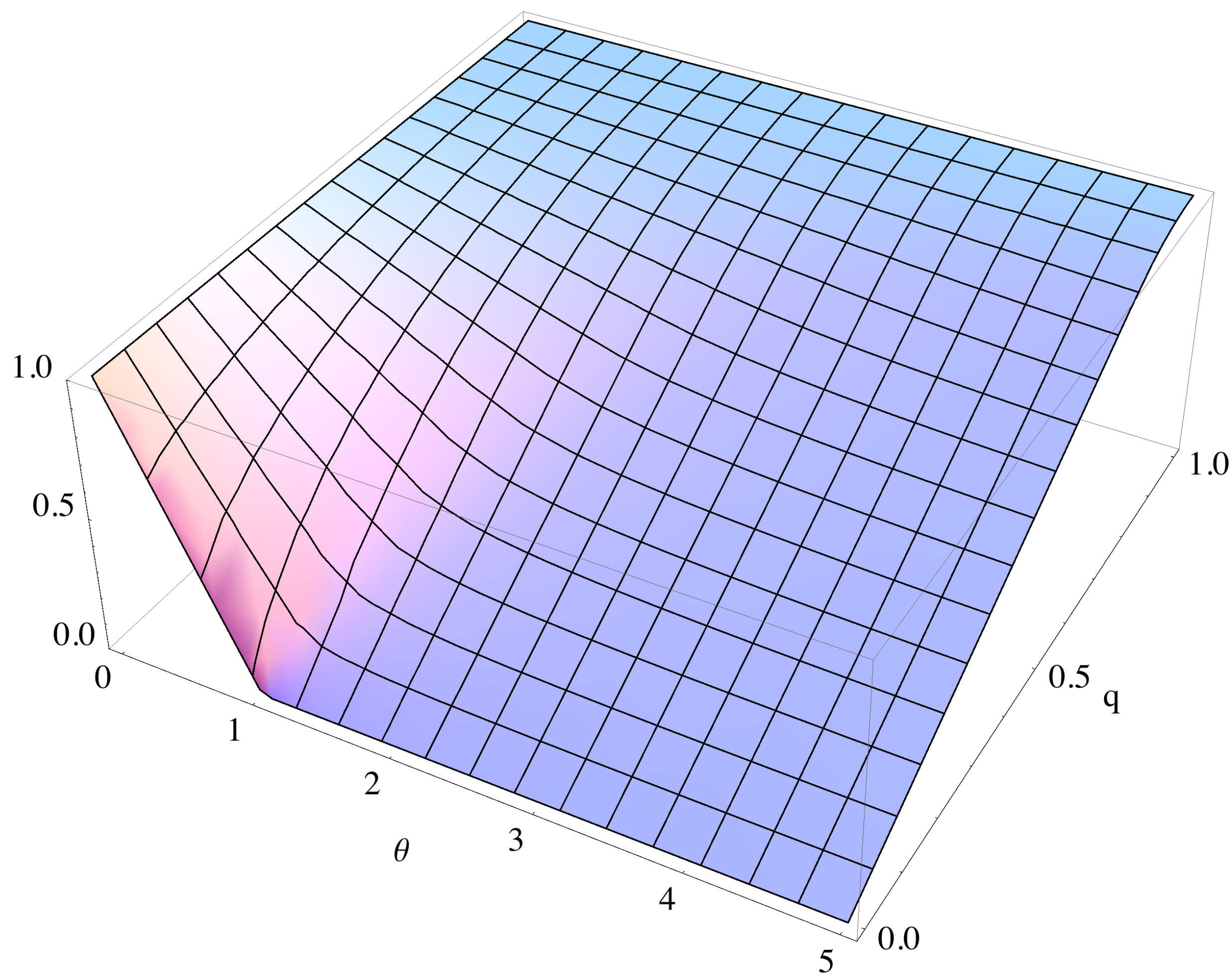


**Figure 7.3** The exponent of variance of cycle lengths Var$[M_q]^{\#}$ for $0 < \theta < 5$.

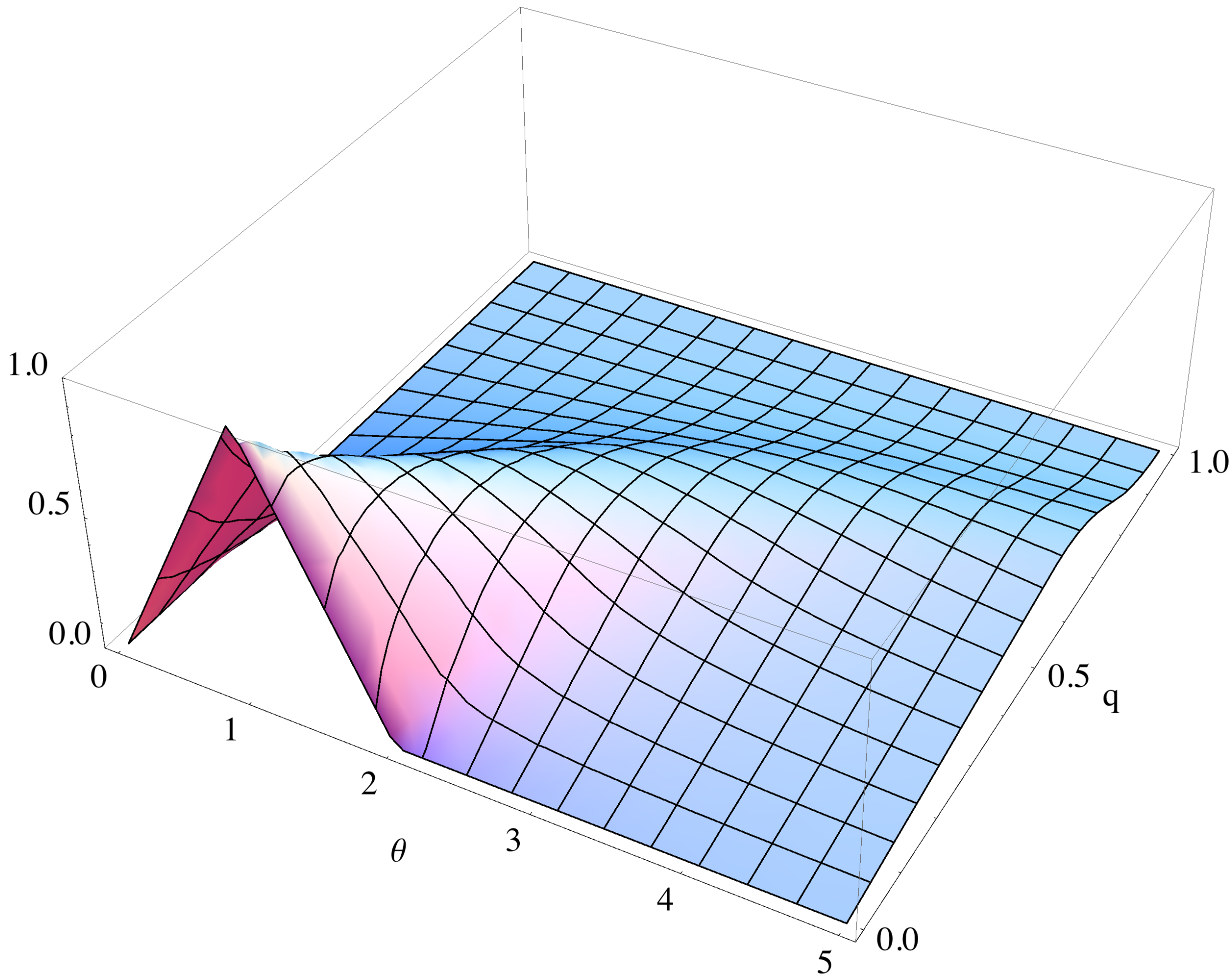
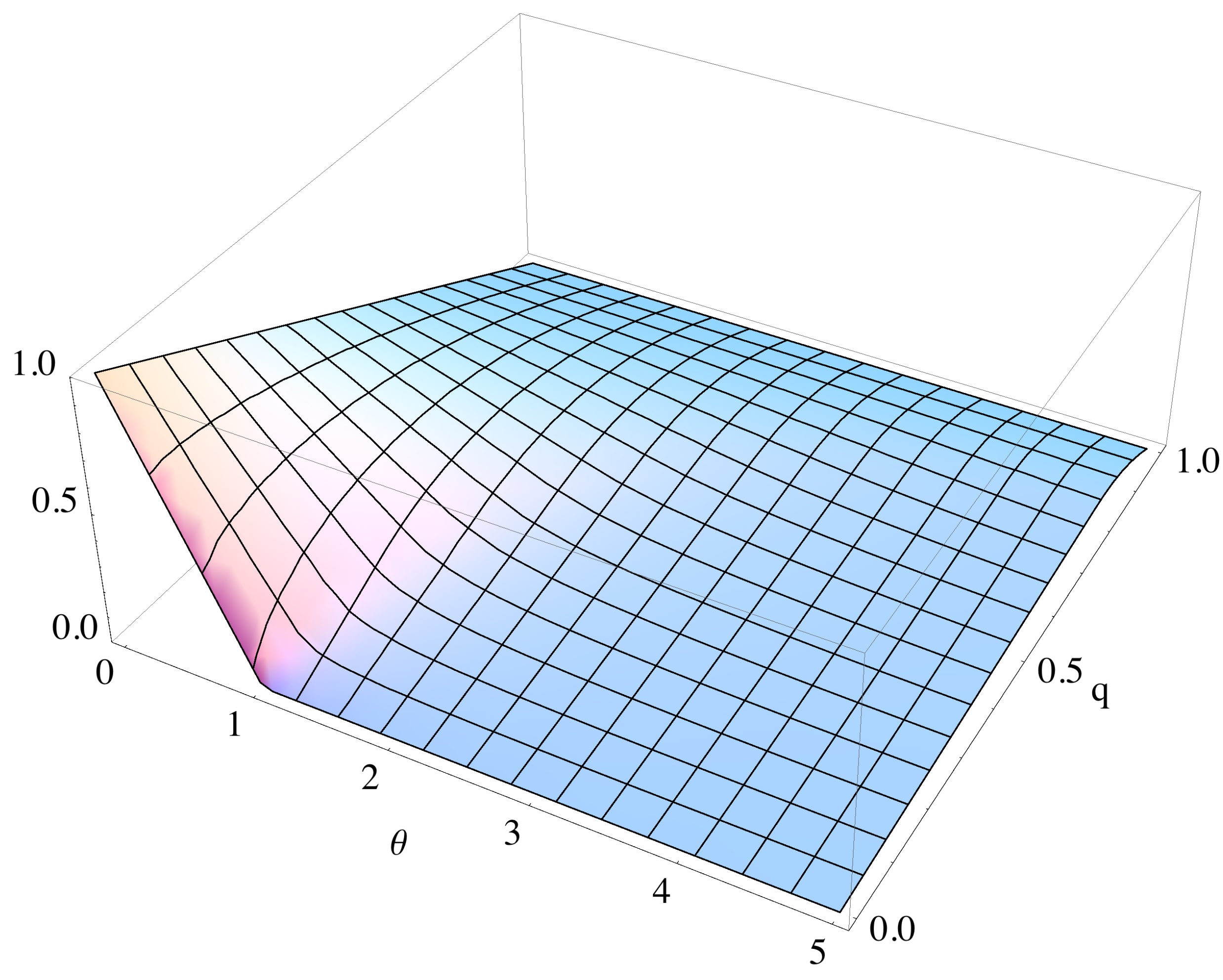


**Figure 7.4** The excess of $\mathbb{E}[M_q]^\#$ over $\mathbb{E}[M_q^\#]$ for $0 < \theta < 5$. It indicates the magnitude of the "heavy tail". The horizontal axes represent $\theta$ and $q$. The vertical axis is $M_q^\#$.



# Chapter 8

# Conclusions and Future Work

In this thesis, we have studied the asymptotic ordered cycle lengths of a $\theta$-biased random permutation. Previously, longest cycles have been well studied in the context of the Poisson-Dirichlet Distribution. Shepp and Lloyd started the study of the shortest cycles in an unbiased random permutation ($\theta = 1$). Following Shepp and Lloyd's paper, Pippenger first studied the first moment of the shortest cycle in a biased random permutation with $\theta = 0.5$, which opens a series of questions: "What's the $m^{\text{th}}$ moment of the $r^{\text{th}}$ shortest cycle in a $\theta$-biased random permutation with any $\theta$?"

This thesis answers the questions posed by Pippenger's paper, and contribute to the current literature in two other ways. First, the cycle lengths at any percentile are studied, which bridges the asymptotic behavior of longest cycles and that of shortest cycles. Second, we have identified that the discrepancies of cycle lengths across different $\theta$'s are due to the heavy tails, because the expected exponent of ordered cycle lengths is not a function of $\theta$. For smaller $\theta$ the tails are heavier, and the moments are affected more heavily.

There are several directions that the research can go from this thesis. The family of random permutations we have considered is controlled by a single parameter $\theta$ indicating the degree of scattering. One may generalize this family by adding additional parameters. Ideally, another stochastic process similar to the Chinese restaurant process can be used to draw samples from any distribution in this new family. Also ideally, a form of conditioning relation will hold for this family of structures.

The technique of exponent analysis used in this thesis is also worth deeper research. The exponent space is some kind of projection of the function space. Although some information is lost, such as the constant terms and the log terms, its ability to convert integration to maximization under some restrictions provides a useful tool and demonstrates a change of perspective.

Other directions of future research include applications in number theory, population genetics, machine learning, etc., finding the exact constants and the log terms in the case of cycle lengths at $100q^{\text{th}}$ percentile, and comparing the rate of convergence to asymptotic behaviors across different $\theta$'s or $q$'s. Random permutations are among the most natural mathematical structures, and they deserve a characterization from us that is just as elegant as themselves.

# Bibliography


Richard Arratia, Andrew D Barbour, and Simon Tavaré. *Logarithmic Combinatorial Strucutures: A Probabilistic Approach*. European Mathematical Society, 2003.

Nathanaël Berestycki. Recent progress in coalescent theory. *Ensaios Matematicos*, 16:1–193, 2009.

N.G. de Bruijn. *Asymptotic Methods in Analysis*. Interscience Publishers, 1958.

Warren J Ewens. The sampling theory of selectively neutral alleles. *Theoretical population biology*, 3(1):87–112, 1972.

RC Griffiths. On the distribution of allele frequencies in a diffusion model. *Theoretical Population Biology*, 15(1):140–158, 1979.

Nicholas Pippenger. Random cyclations. *The Electronic Journal of Combinatorics*, 20(4):9, 2013.

LA Shepp and SP Lloyd. Ordered cycle lengths in a random permutation. *Transactions of the American Mathematical Society*, 121(2):340–357, 1966.