COVERING PROBLEMS FOR RANDOM WALKS
ON SPHERES AND FINITE GROUPS

BY

PETER CLAVER MATTHEWS

TECHNICAL REPORT NO. 234
AUGUST 1985

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS80-24649

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
COVERING PROBLEMS FOR RANDOM WALKS
ON SPHERES AND FINITE GROUPS

By

Peter Claver Matthews

TECHNICAL REPORT NO. 234
August 1985

PREPARED UNDER THE AUSPICIES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS80-24649

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. INEQUALITIES FOR COUPON COLLECTOR TYPE PROBLEMS</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Notation and Preliminary Lemmas</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Inequalities for the Mean</td>
<td>14</td>
</tr>
<tr>
<td>2.4 Inequalities for the Moment Generating Function</td>
<td>15</td>
</tr>
<tr>
<td>2.5 Comments</td>
<td>18</td>
</tr>
<tr>
<td>3. PRELIMINARIES FOR FINITE GROUP PROBLEMS</td>
<td>23</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>23</td>
</tr>
<tr>
<td>3.2 Review of Finite Groups</td>
<td>23</td>
</tr>
<tr>
<td>3.3 Random Walks on Finite Groups as Markov Chains</td>
<td>32</td>
</tr>
<tr>
<td>3.4 Generating Functions</td>
<td>36</td>
</tr>
<tr>
<td>3.5 Range of Existence of Generating Functions</td>
<td>37</td>
</tr>
<tr>
<td>3.6 The Ratio of Gamma Functions</td>
<td>46</td>
</tr>
<tr>
<td>4. FINITE GROUP PROBLEMS</td>
<td>47</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>47</td>
</tr>
<tr>
<td>4.2 The Discrete Circle</td>
<td>48</td>
</tr>
<tr>
<td>4.3 The Discrete Cube</td>
<td>55</td>
</tr>
<tr>
<td>4.4 The Symmetric Group</td>
<td>69</td>
</tr>
<tr>
<td>5. PRELIMINARIES FOR SPHERE PROBLEMS</td>
<td>78</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>78</td>
</tr>
<tr>
<td>5.2 The Grand Tour</td>
<td>79</td>
</tr>
<tr>
<td>5.3 Covering and Packing a Sphere with Caps</td>
<td>82</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

My deepest thanks go to my advisor, Professor Persi Diaconis. Besides introducing me to an interesting new field, he provided me with insight, information, references, and encouragement throughout the writing of this thesis. I thank Professor David Siegmund for reading this thesis, as well as for several helpful discussions during its preparation. I also thank Professor Iain Johnstone for reading this thesis and for several enlightening suggestions.

Carolyn Knutsen did a wonderful job of typing from afar with all the telepathic insight it requires. Judi Davis did a great job of revising. I also thank Judi for her help getting through all the bureaucratic necessities of writing a thesis and of the rest of a graduate education.

I owe special thanks to Sherrie Emoto for her constant encouragement and consolation.

Finally, for help that I cannot begin to acknowledge sufficiently, I want to thank my parents, Lawrence and Mary Fran Matthews.
CHAPTER 1

INTRODUCTION

The classical Coupon Collector's Problem is a result from elementary probability that appears frequently, in one guise or another, in many statistical problems. As such, it and many generalizations have been studied extensively. This thesis considers another generalization of it into a useful but relatively unexplored area.

The classical Coupon Collector's Problem (Feller, 1968) deals with the following waiting time problem. Suppose that with each visit to a market a shopper is given a coupon, each time equally likely to be labelled 1, 2, ..., or N. Upon collecting a full set of N distinct coupons, the shopper receives a prize. Thus the coupon collector is interested in $C_N$, the number of visits to the market he will have to make to collect a complete set of coupons. Assuming the coupons received for different visits to be independently distributed, the distribution of $C_N$ is well known. A standard asymptotic result is

$$P\left(\frac{1}{N}(C_N - N \log N) \leq x\right) \to \exp(-e^{-x}) \text{ as } N \to \infty. \tag{1.1}$$

Suitably normalized $C_N$ converges in distribution to the extreme value (or Gumbel) distribution. In fact the moment generating function also converges. A great deal more is known about the Coupon Collector's Problem.
The book *Random Allocations* by V.F. Kolchin, B.A. Sevast'yanov and V.P. Chistyakov (1978) contains a variety of results on various aspects and variants of this problem.

To motivate the generalizations to be considered here, consider the Coupon Collector's Problem in a different way, as a Markov chain. Let the coupon labels $1, 2, \ldots, N$ be the state space of a Markov chain. Then the chain is in state $n$ at time $i$ if the coupon received for visit $i$ is labeled $n$. Clearly this is a Markov chain with $N \times N$ transition matrix

$$
\begin{pmatrix}
\frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\
\frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\
\frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N}
\end{pmatrix}.
$$

With this terminology the Coupon Collector's interest lies in the time taken by the Markov chain to visit every point in its state space.

The same problem can be posed for any finite Markov chain, though finding a solution may generally be quite difficult. There is a class of Markov chains where the problem is tractable with a solution similar to (1.1), namely rapidly mixing random walks on finite groups. Loosely, a Markov chain with $N$ states is rapidly mixing if it converges to its stationary distribution rapidly; if the time it takes to get close to its stationary distribution is small compared to $N$. Aldous (1983b) gives a discussion of rapidly mixing Markov chains. Let $C$ denote the
time taken by a random walk on a finite group $G$ to cover $G$, i.e., to visit every element of $G$. Aldous (1983a) shows that for many rapidly mixing examples, $EC$ is about $|G| \log |G|$, where $|G|$ is the number of elements in $G$. Aldous obtains his results from the classical Coupon Collector's Problem and the rates of convergence of the random walks to their stationary distribution.

Using a different technique, this thesis gives results like (1.1) for several random walks on finite groups. Consider a random walk on $\mathbb{Z}_N$, the discrete $N$-point circle, with step length uniformly distributed on $\{-aN,-aN+1, \ldots, aN\}$, where $\frac{1}{2} < a > 0$ and $\lfloor aN \rfloor$ denotes the largest integer less than or equal to $aN$. With $C_N$ the time taken by the random walk to visit every state, the moment generating function of $C_N$ satisfies

$$E \exp \left( \frac{\theta}{N} (C_N - N \log N) \right) \to \Gamma(1-s) \text{ as } N \to \infty$$

for $-\infty < s < 1$. $\Gamma(1-s)$ is the moment generating function of the extreme value distribution, so this problem has the same asymptotic properties as the Coupon Collector's Problem. Similarly, for certain sequences of random walks on the symmetric groups, $C_N$ the time taken by the random walk on Sym$(N)$, the symmetric group on $N$ letters, to visit every state,

$$E \exp \left[ \frac{\theta}{N!} (C_N - N! \log(N!)) \right] \to \Gamma(1-s) \text{ as } N \to \infty$$

for $-\infty < s < 1$. $N!$ is the order of Sym$(N)$, so again the problem is much like the Coupon Collector's Problem. Finally, consider $N$
distinguishable balls in two urns. One of the \( N \) balls is selected at random, located, and moved to the other urn. Let \( C_N \) be the number of such independent switches needed for all \( 2^N \) possible arrangements to occur. This is the classical Ehrenfest urn model (Kemperman, 1961). It can be considered as a random walk on \( \mathbb{Z}_2^N \). Here it is shown that

\[
E \exp \left[ \frac{S}{2N} (C_N - 2^N \log(2^{N+1})) \right] \to \Gamma(1-s) \quad \text{as} \quad N \to \infty
\]

for \(-1 < s < 1\). The \( \log(2^{N+1}) \) in the mean of \( C_N \) makes this result slightly different from the previous ones.

The technique used is to find upper and lower bounds on the moment generating functions of the quantities termed \( C_N \). The technique works whenever the random walk is sufficiently rapidly mixing, though no good general convergence theorem is yet available.

Another version of the Coupon Collector's Problem is the problem of covering a sphere with caps. Let \( S^p_{p-1} \) denote the unit sphere in \( \mathbb{R}^p \). Fix \( \varepsilon > 0 \) and choose points \( X_1, X_2, \ldots \) independently according to the uniform distribution on \( S^p_{p-1} \). The quantity analogous to \( C_N \) of the Coupon Collector's Problem is \( C_{\varepsilon} \), the smallest integer \( M \) such that the caps of geodesic radius \( \varepsilon \) about \( X_1, X_2, \ldots, X_M \) cover \( S^p_{p-1} \). If volume is normalized so that \( S^p_{p-1} \) has volume 1, then as \( \varepsilon \to 0 \) the volume of a cap of radius \( \varepsilon \) is asymptotic to \( \alpha(p, \varepsilon) \), where

\[
\alpha(p, \varepsilon) = \frac{\varepsilon^{p-1} \Gamma \left( \frac{p}{2} \right)}{(p-1) \Gamma \left( \frac{p-1}{2} \right) \Gamma \left( \frac{1}{2} \right)}.
\]
A fair amount is known about this covering problem. L. Flatto and D.J. Newman (1977) proved

\[ EC_\epsilon = \frac{1}{\alpha(p, \epsilon)} \left[ \log \left( \frac{1}{\alpha(p, \epsilon)} \right) + (p-1)\log\log \left( \frac{1}{\alpha(p, \epsilon)} \right) + o(1) \right] \]

as \( \epsilon \to 0 \). Recently, employing Poissonization, S. Janson (1984) showed that

\[ \alpha(p, \epsilon) C_\epsilon \log \left( \frac{1}{\alpha(p, \epsilon)} \right) - (p-1)\log\log \left( \frac{1}{\alpha(p, \epsilon)} \right) - c(p) \]

converges in distribution to the extreme value distribution as \( \epsilon \to 0 \), where

\[ c(p) = \log \left\{ \frac{1}{\Gamma(p)} \left[ \sqrt{\pi} \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \right]^{p-2} \right\}. \]

Both Flatto and Newman and Janson give their results in greater generality than for the sphere.

This problem has a generalization which is analogous to the generalization of the Coupon Collector's Problem to random walks on finite groups. Instead of choosing \( X_1, X_2, \ldots \) independently on \( S_{p-1} \), let \( X_1, X_2, \ldots \) be a symmetric random walk on \( S_{p-1} \) or let \( \{X_t\}_{t \geq 0} \) be a Brownian motion on \( S_{p-1} \). We let \( C_\epsilon \) be the smallest number such that caps of radius \( \epsilon \) about the points \( X_1, \ldots, X_{C_\epsilon} \) or the points \( \{X_t\}_{0 \leq t \leq C_\epsilon} \) cover \( S_{p-1} \). Equivalently, we can think of \( C_\epsilon \) as the
first time the random walk or Brownian motion has been within a distance \( \varepsilon \) of all points of \( S_{p-1} \).

The consideration of \( C_\varepsilon \) for random walks was motivated by the Grand Tour (Asimov, 1985). A Grand Tour is a procedure to examine \( p \)-dimensional data by eye by looking at a sequence of one or two dimensional projections of the data. One method of making a Grand Tour is to look at the sequence of projections of the data onto the one dimension subspaces of \( \mathbb{R}^p \) spanned by the sequence of positions of a random walk on \( S_{p-1} \). A question of interest is the expected value of the time taken by such a Grand Tour to be close to every possible projection. This motivates a slightly more complicated covering problem than that of a random walk on a sphere, since points of \( S_p \) must now be identified with their reflections in the origin. However, the techniques needed to solve the two problems are identical. The consideration of \( C_\varepsilon \) for Brownian motion follows naturally from letting the step size of a random walk on \( S_{p-1} \) shrink. The Brownian motion problem is much easier computationally than the random walk problem, and results for Brownian motion problems can be used to get information on random walk problems.

The results given here are not as sharp as (1.2); of course they are results for a different problem. For Brownian motion on \( S_{p-1} \) we have, for \( p \geq 4 \) as \( \varepsilon \to 0 \)

\[
E(C_\varepsilon) = 4\sqrt{\pi} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{\log\left(\frac{1}{\varepsilon}\right)}{\varepsilon^{p-1}} \quad \left[1 + O\left(\frac{\log\log\left(\frac{1}{\varepsilon}\right)}{\log\left(\frac{1}{\varepsilon}\right)}\right)\right]
\]

with a weaker result for \( p = 3 \). Again the technique used is finding
upper and lower bounds on $EC_\varepsilon$. For the Grand Tour problem, with points on the sphere identified with their reflections in the origin, all results are half as large as the corresponding results for the simple covering problem. For random walks, I have no good explicit bounds or asymptotic results. However, as the step size of a random walk is reduced and time is suitably renormalized, the coverage time eventually lies between the Brownian motion bounds. This will be made precise in Chapter 6.

The outline for the rest of this dissertation is as follows. Chapter 2 gives some results along the lines of the Coupon Collector's Problem that will be useful in the problems considered here. Chapters 3 and 4 and 5 and 6 are pairs; Chapters 3 and 4 are on finite group problems while Chapters 5 and 6 are on sphere problems. Chapters 3 and 5 give necessary background material for Chapters 4 and 6, which state and prove the results on covering problems. The final chapter, Chapter 7, discusses the results of the previous chapters and suggests possibilities for further results.

Throughout this thesis $C$, with various subscripts, will refer to a covering time as described above. $T$, again with various subscripts, will refer to a hitting time. Finally $f(s)$, with subscripts possibly, will refer to the moment generating function of a hitting time,

$$f(s) = E \exp(sT).$$
CHAPTER 2

INEQUALITIES FOR COUPON COLLECTOR TYPE PROBLEMS

2.1 INTRODUCTION

Consider a set $K$ and a finite collection $\{A_1, A_2, \ldots, A_N\}$ of subsets of $K$. Let $(X_t)$, where $t$ ranges over the non-negative integers or reals, be a time-homogeneous strong Markov process with state space $K$. This chapter investigates $C$, the first time the process has visited all of $A_1, A_2, \ldots, A_N$. The covering problem for a random walk on a finite group fits into this framework. The expected covering times for Brownian motion or random walks on spheres can be bounded above and below by expected covering times in problems of this sort. This chapter gives some inequalities for $C$, in particular, upper and lower bounds on its mean and moment generating function. These upper and lower bounds coincide for the classical Coupon Collector's Problem and become tight asymptotically for many of the processes considered here.

2.2 NOTATION AND PRELIMINARY LEMMAS

This section gives the notation to be used and some preliminary lemmas for bounding the means and moment generating functions of various covering times. The class of processes covered will be broad enough to include the examples mentioned in the introduction, but will be overly restrictive in some ways. It will be clear from the proofs of the lemmas in this section that they hold in somewhat greater generality.
Let $K$ be either a finite set or a metric space. Let $\{A_1, \ldots, A_N\}$ be a collection of $N$ distinct closed subsets of $K$. Let $(X_t(\omega))$ be a Markov process taking values in $K$, where $\omega \in \Omega$ indexes the particular sample path. It will be assumed that either time is discrete ($t \in \mathbb{N}_0$) or, if time is continuous ($t \in \mathbb{R}$), that $X_t(\omega)$ is a diffusion. In either case $(X_t(\omega))$ has the strong Markov property. Let $X_0 \in K$ be the initial position of $(X_t)$.

Next consider hitting times for $(X_t)$. For any $a \in K$ and $A \in \{A_1, \ldots, A_N\}$, let $T_{aA}$ denote the time taken by $(X_t)$ to hit $A$ from $a$. Let $\hat{A}_i = \{X_0\} \cup \bigcup_{j \neq i} A_j$.

Define

\begin{equation}
\mu^- = \min_{i=1,2,\ldots,N} \inf_{a \in \hat{A}_i} E T_{aA_i} \\
\mu^+ = \max_{i=1,2,\ldots,N} \sup_{a \in \hat{A}_i} E T_{aA_i}
\end{equation}

Thus $\mu^-$ and $\mu^+$ are minimal and maximal expected hitting times, minimized or maximized over the time to hit some $A_j$ from $X_0$ or from some $A_i$, for $i \neq j$, for $j = 1,2,\ldots,N$. Similarly, for $s \in \mathbb{R}$ define

\begin{equation}
f^-(s) = \min_{i=1,2,\ldots,N} \inf_{a \in \hat{A}_i} E[\exp(sT_{aA_i})] \\
f^+(s) = \max_{i=1,2,\ldots,N} \sup_{a \in \hat{A}_i} E[\exp(sT_{aA_i})].
\end{equation}
It will generally be necessary to make the following assumption about $X_0$ and $\{A_1, ..., A_N\}$.

\[(2.2.3)\quad A_i \cap A_j = \emptyset \quad \text{for } i \neq j\]

and

$$X_0 \notin \bigcup_{j=1,...,N} A_j.$$  

Finally, let $C$ be the time taken by $(X_t)$, starting at $X_0$, to visit all of $\{A_1, ..., A_N\}$. The most tractable route to bounds on $EC$ and $E \exp(sC)$ for $s \in \mathbb{R}$ is the introduction of auxiliary randomization. Heuristically, we can think of $C$ as the sum of $N$ independent randomized hitting times as follows. Choose $A_{\sigma_1}$ at random from $\{A_1, ..., A_N\}$. Let $R_1 = S_1$ be the time taken by $(X_t)$ to hit $A_{\sigma_1}$. Choose $A_{\sigma_2}$ at random from $\{A_1, ..., A_N\} \setminus \{A_{\sigma_1}\}$. Let $R_2 = S_2 - S_1$ be the additional time, if any, until $(X_t)$ has visited $A_{\sigma_2}$. Thus $R_2 = 0$ if $A_{\sigma_2}$ is visited before $A_{\sigma_1}$. Then $S_2 = R_1 + R_2$ is the time taken to hit both $A_{\sigma_1}$ and $A_{\sigma_2}$. In general, choose $A_{\sigma_i}$ at random from $\{A_1, ..., A_N\} \setminus \{A_{\sigma_1}, ..., A_{\sigma_{i-1}}\}$ and let $R_i = S_i - S_{i-1}$ be the additional time, if any, until $(X_t)$ has visited $A_{\sigma_i}$. Then $S_i$ is the first time $(X_t)$ has visited all of $\{A_{\sigma_1}, ..., A_{\sigma_i}\}$. In particular $S_N = C$. This partitioning of $C$ into $R_1 + T_2 + ... + R_N$ turns out to be very useful. This motivates the following construction.

Let $\hat{\pi}$ be the set of all $N!$ permutations of $\{1, 2, ..., N\}$. Let $\sigma = (\sigma_1, ..., \sigma_N)$ be a permutation chosen from the uniform distribution on $\hat{\pi}$, independently of the $(X_t)$ process. Recalling that the process $(X_t(\omega))$ is defined on $\Omega$, we can form the product space $\hat{\pi} \times \Omega$ and think
of both \((X_t(\omega))\) and \(\sigma\) as being defined on \(\mathbb{F} \times \Omega\). Let \(F_t\) be the \(\sigma\)-field on \(\mathbb{F} \times \Omega\) generated by the permutation \(\sigma\) and \(\{X_s, 0 \leq s \leq t\}\). As in the heuristic discussion, let \(S_1\) be the first time \((X_t)\) visits \(A_{\sigma_1}\), \(S_1 = \inf\{t \geq 0: X_t \in A_{\sigma_1}\}\). In general let \(S_i\) be the first time \((X_t)\) has visited all of \(A_{\sigma_1}, \ldots, A_{\sigma_i}\), \(S_i = \inf\{t \geq 0: X_{s_1} \in A_{\sigma_1} \text{ for some } s_1 \leq t, \ldots, X_{s_i} \in A_{\sigma_i} \text{ for some } s_i \leq t\}\). For notational convenience, let \(S_0 = 0\).

Some simple consequences of these definitions are given in the following lemmas. When assumption (2.2.3) is required it will be mentioned.

Lemma (2.2.4) \(S_N = C\). In words, \(S_N\) is the first time all of \(\{A_1, \ldots, A_N\}\) have been visited.

Proof. This is clear from the definition of \(S_N\), since \(\{A_{\sigma_1}, \ldots, A_{\sigma_N}\}\) is the same set as \(\{A_1, \ldots, A_N\}\).

Lemma (2.2.5) For \(1 \leq i \leq N\), for all \(t \geq 0\), \(\{S_i \leq t\} \in F_t\). In words, \(S_i\) is a stopping time.

Proof. Fix \(t \geq 0\) and \(i\) between 1 and \(N\). Let \(\pi = \{\pi_1, \ldots, \pi_i\}\) range over all \(\binom{N}{i}\) possible choices of \(i\) distinct elements from \(\{1, 2, \ldots, N\}\). Then

\[
\{S_i \leq t\} = \cup_{\pi} \{S_i \leq t, \{A_{\sigma_1}, \ldots, A_{\sigma_i}\} = \{A_{\pi_1}, \ldots, A_{\pi_i}\}\}
\]

\[
= \cup_{\pi} \{\{A_{\sigma_1}, \ldots, A_{\sigma_i}\} = \{A_{\pi_1}, \ldots, A_{\pi_i}\}\} \cap \{X_s\}\text{ has visited all of }\{A_{\pi_1}, \ldots, A_{\pi_i}\}\text{ by time }t\}
\]

11
Each of these events is in $F_t$, so the union of their intersections is as well.

So we can define $F_i$ to be the stopped $\sigma$-field on $\mathbb{P} \times \Omega$. $F_i$ is the smallest $\sigma$-field such that $\sigma$ and $(X_s), 0 \leq s \leq S_1$, are measurable with respect to it. To make the notation a bit briefer, define $R_i = S_i - S_{i-1}$ for $i=1, \ldots, N$. Thus $R_i$ is the additional time, if any, taken by $(X_t)$ to have visited $A_{\sigma_i}$ after having visited $\{A_{\sigma_1}, \ldots, A_{\sigma_{i-1}}\}$.

Lemma (2.2.5) \( \{R_i = 0\} \in F_{i-1} \) for $i=1,2, \ldots, N$.

Proof. \( \{R_i = 0\} = \bigcup_{j=1}^{N} \{A_{\sigma_i} = A_j\} \cap \{(X_t) \text{ visits } A_j \text{ before time } S_i\} \). Each of these events is in $F_{i-1}$.

Lemma (2.2.6) $P(R_i \neq 0) \leq \frac{1}{i}$ and under assumption (2.2.3), $P(R_i \neq 0) = \frac{1}{i}$.

Proof. First assume (2.2.3). $P(S_i \neq 0) = 1$ since $X_0 \notin A_1$. For $i > 1$, we proceed as follows. For $j=1, \ldots, N$, $A_j \in \{A_1, \ldots, A_N\}$, recall that $T_{X_0 A_j}$ is the first time $(X_t)$ visits $A_j$. Under (2.2.3) the $T_{X_0 A_j}$ are distinct with probability one. $\{T_{X_0 A_1}, \ldots, T_{X_0 A_N}\}$ have $N!$ possible orderings, though some may not have positive probability. Let $\pi$ index these orderings. This partitions $\mathbb{P} \times \Omega$ into $N!$ components, $((\mathbb{P} \times \Omega)_{\pi})$, as $\pi$ runs over all permutations of $\{1, \ldots, N\}$. For example, if $N=3$ and $\pi = (1,3,2)$ $(\mathbb{P} \times \Omega)_{\pi} = \{T_{X_0 A_1} < T_{X_0 A_3} < T_{X_0 A_2}\}$. Since the partitioning depends only on $(X_t)$ and not $\sigma$, by independence we have that if $P((\mathbb{P} \times \Omega)_{\pi}) > 0$ then $\sigma$ is uniformly distributed given $(\mathbb{P} \times \Omega)_{\pi}$.

On $(\mathbb{P} \times \Omega)_{\pi}$, the event $\{R_i \neq 0\}$ is simply the event that $\sigma_i$ appears farther to the right in $\pi$ than all of $\sigma_1, \ldots, \sigma_{i-1}$. By the uniformity
of σ, this has conditional probability 1/i. Thus P(Ri≠0) = 1/i on the whole space $\mathcal{F}_W$ as well.

Without assuming (2.2.3) ties among $T_{X_0 A^1}, \ldots, T_{X_0 A^N}$ are possible. But this partitions $\mathcal{F}_W$ independently of σ. On one of the subsets $(\mathcal{F}_W)_\pi$, \( \{R_i \neq 0\} \) is again the event that σ is farther to the right in π than $\sigma_1, \ldots, \sigma_{i-1}$. With the possibility of ties, this has conditional probability $\leq 1/i$. Thus, unconditionally $P(R_i \neq 0) \leq 1/i$.

Note that the part of the proof using (2.2.3) did not use it too strongly. All that was required was that $T_{X_0 A^1}, \ldots, T_{X_0 A^N}$ were distinct with probability 1.

Lemma (2.2.7) Under (2.2.3) $S_1$ and $I_{\{R_i = 0\}}$ are independent.

Proof. Intuitively this is because $S_1$ depends only on the set $\{\sigma_1, \ldots, \sigma_i\}$, not their particular order.

To make this rigorous, let $\pi = (\pi_1, \ldots, \pi_i)$ index all $i!$ ordered choices of $i$ indices from $\{1, 2, \ldots, N\}$.

$$P(S_1 \leq t, R_i \neq 0) =$$

$$\sum_{\pi} P(\text{every } T_{X_0 A^\pi_1} < T_{X_0 A^\pi_2} < \cdots < T_{X_0 A^\pi_i} \leq t, R_i \neq 0, \{\sigma_1, \ldots, \sigma_i\} = \{\pi_1, \ldots, \pi_i\}),$$

where $\{\sigma_1, \ldots, \sigma_i\} = \{\pi_1, \ldots, \pi_i\}$ means equality as sets; the two sets have the same members. This is

$$= \sum_{\pi} P(\text{every } T_{X_0 A^\pi_1} < \cdots < T_{X_0 A^\pi_i} \leq t, \sigma_i = \pi_i, \{\sigma_1, \ldots, \sigma_i\} = \{\pi_1, \ldots, \pi_i\})$$

13
\[
= \sum_{\pi} P(\sigma_i = \pi_i | T_{X_0\Lambda_{\pi_1}} < \cdots < t, \{\sigma_1, \ldots, \sigma_1\} = \{\pi_1, \ldots, \pi_1\}) 
\]

\[
P(T_{X_0\Lambda_{\pi_1}} < \cdots < t, \{\sigma_1, \ldots, \sigma_1\} = \{\pi_1, \ldots, \pi_1\})
\]

\[
= \frac{1}{\pi} \sum_{\pi} P(T_{X_0\Lambda_{\pi_1}} < \cdots < t, \{\sigma_1, \ldots, \sigma_1\} = \{\pi_1, \ldots, \pi_1\})
\]

\[
= P(R_i \neq 0) P(S_i \leq t),
\]

where we have used the fact that given \(\{\sigma_1, \ldots, \sigma_1\} = \{\pi_1, \ldots, \pi_1\}\), \(\sigma_1\) is equally likely to be any of \(\{\pi_1, \ldots, \pi_1\}\), even conditional on some event involving only \((X_c)\) as well. ■

2.3 INEQUALITIES FOR THE MEAN

In this section the following theorem is proven.

Theorem (2.3.1)

\[
EC \leq \mu^+ + \sum_{i=1}^{N} \frac{1}{i}
\]

and under (2.2.3)

\[
EC \geq \mu^- - \sum_{i=1}^{N} \frac{1}{i}.
\]

Proof. If either \(\mu^-\) or \(\mu^+\) are infinite, it is easy to see that the corresponding bounds hold. Therefore assume \(\mu^+\), and hence \(\mu^-\) as well, is finite. By Lemma (2.2.4) \(C = S_N = \sum_{i=1}^{N} R_i\). Hence
Let \( T_{X|A}^{(i)} \) be the additional time taken to hit \( A \) from \( x \) starting at time \( S_1 \). By time homogeneity and the Strong Markov Property, this has the same distribution as \( T_{X|A} \).

By Lemma (2.2.5) \( \{R_i \neq 0\} \in F_{i-1} \). So \( E(R_i | F_{i-1}) = I_{\{R_i \neq 0\}} E T_{X|A}^{(i-1)} \).

On the set \( \{R_i \neq 0\} \), \( X_{S_{i-1}} \) must be in \( \hat{A} \). Using this, time-homogeneity, the Strong Markov Property, and the definitions (2.2.1) of \( \mu^- \) and \( \mu^+ \),

\[
I_{\{R_i \neq 0\}} \mu^- \leq E(R_i | F_{i-1}) \leq I_{\{R_i \neq 0\}} \mu^+ .
\]

From Lemma (2.2.6) \( P(R_i \neq 0) \leq 1/i \) and, under (2.2.3) \( P(R_i \neq 0) = 1/i \).

The assertions of the theorem follow immediately.

### 2.4 Inequalities for the Moment Generating Function

In this section lower and upper bounds on \( E[\exp(sC)] \) are calculated.

Here, (2.2.3) must be assumed for both bounds. The result is

**Theorem (2.4.1)** Under assumption (2.2.3), for \( s \in \mathbb{R} \)

\[
\frac{\Gamma(N+1)f^-(s)}{\Gamma(N+1/f^-(s))} \leq E[\exp(sC)] \leq \frac{\Gamma(N+1)f^+(s)}{\Gamma(N+1/f^+(s))}
\]

where we interpret \( \Gamma(0) \) as \( +\infty \).

**Proof.** If \( f^-(s) \) is infinite, it is easy to see that \( E[\exp(sC)] \)
must be infinite as well. If \( f^+(s) \) is infinite, the right inequality is trivially true. Therefore from now on assume \( f^+(s) < \infty \). \( f^-(s) \)
will then be finite as well.
The plan is to get bounds on \( E[\exp(sS_i)] \) in terms of \( E(\exp(sS_{i-1})) \). As a starting point, note that

\[
(2.4.2) \quad \frac{1}{l^+} = f^+(s) \leq E[\exp(sS_i)] \leq f^+(s) = \frac{1}{l^-} \quad \frac{1}{l^-+1/f^-(s)}.
\]

For \( i \geq 2 \) write \( E \exp(sS_i) = E(E(\exp(sS_i) | F_{i-1})) \). Consider the inner conditional expectation. \( S_i = S_{i-1} + R_i \), and \( S_{i-1} \) is measurable with respect to \( F_{i-1} \). Therefore \( E(\exp(sS_i) | F_{i-1}) = \exp(sS_{i-1})E[\exp(sR_i) | F_{i-1}] \).

By Lemma (2.2.5), \( \{R_i = 0\} \in F_{i-1} \) and \( X_{S_{i-1}} \) and \( A_{\sigma_i} \) are known given \( F_{i-1} \). So the conditional expectation becomes

\[
E[\exp(sS_{i-1})[I_{\{R_i = 0\}}^+ I_{\{R_i \neq 0\}} E[\exp(sX_{S_{i-1}})^{(i-1)} A_{\sigma_i} | F_{i-1}]],
\]

On the set \( \{R_i \neq 0\} \), \( X_{S_{i-1}} \in \hat{A}_{\sigma_i} \). By the definitions of \( f^-(s) \) and \( f^+(s) \) (2.2.2), time-homogeneity, and the Strong Markov Property,

\[
f^-(s)I_{\{R_i \neq 0\}} \leq I_{\{R_i \neq 0\}} E[\exp(sX_{S_{i-1}})^{(i-1)} A_{\sigma_i} | F_{i-1}] \leq f^+(s)I_{\{R_i \neq 0\}}.
\]

Putting this all together,

\[
E[\exp(sS_{i-1})[I_{\{R_i = 0\}}^+ I_{\{R_i \neq 0\}} f^-(s)] \leq E[\exp(sS_i)] \leq \quad \leq E[\exp(sS_{i-1})[I_{\{R_i = 0\}}^+ I_{\{R_i \neq 0\}} f^+(s)]]
\]

16
For brevity, consider the right inequality only; the calculations on
the other are identical. Adding and subtracting \( I_{\{R_i = 0\}} f^+(s) \) on the
right, it becomes

\[
E[\exp(sS_i)] \leq E[\exp(sS_{i-1}) (1 - f^+(s)) + f^+(s)] .
\]

On \( \{R_i = 0\} \), \( S_{i-1} = S_i \). By Lemma (2.2.7), \( S_i \) and \( 1_{\{R_i = 0\}} \) are
independent. So

\[
E[\exp(sS_i)] \leq E[\exp(sS_i)] (1 - \frac{1}{i}) (1 - f^+(s)) + f^+(s) E[\exp(sS_{i-1})] .
\]

Rearranging,

\[
E[\exp(sS_i)] \leq \frac{1}{i - 1 + 1/f^+(s)} E[\exp(sS_{i-1})] .
\]

Similarly,

\[
E[\exp(sS_i)] \geq \frac{1}{i - 1 + 1/f^-(s)} E[\exp(sS_{i-1})] .
\]

Combining these facts for \( 2 \leq i \leq N \) with (2.4.2),

\[
\prod_{i=1}^{N} \frac{1}{i - 1 + 1/f^-(s)} \leq E[\exp(sS_N)] \leq \prod_{i=1}^{N} \frac{1}{i - 1 + 1/f^+(s)} .
\]

This is precisely the assertion of the theorem. \( \blacksquare \)
2.5 COMMENTS

1. These techniques can be used to give a somewhat roundabout solution to the Coupon Collector's Problem. In this framework it becomes a Markov chain with \( N+1 \) states \( \{ A_0, A_1, \ldots, A_N \} \). \( A_0 \) is the initial state of the chain; \( A_i \) is identified with receiving coupon \( i \), for \( i=1, 2, \ldots, N \). \( C_N \) is the time taken by the chain to visit all of \( \{ A_1, A_2, \ldots, A_N \} \). The one step transition matrix \( P \) has entries

\[
P_{ij} = \begin{cases} 
\frac{1}{N} & j \neq 0 \\
0 & j = 0 
\end{cases}
\]

\( T_{ij} \), the time to hit \( A_j \) from \( A_i \), for \( j \neq 0 \), has a geometric distribution with parameter \( \frac{1}{N} \). So

\[
E[\exp sT_{ij}] = \frac{1}{N} \frac{e^s}{1-(1-\frac{1}{N})e^s} \text{ for } -\infty < s < \log\left(\frac{N}{N-1}\right).
\]

From (2.2.2)

(2.5.1) \[ f^{-}(s) = f^{+}(s) = \frac{1}{N} \frac{e^s}{1-(1-\frac{1}{N})e^s} \text{ for } -\infty < s < \log\left(\frac{N}{N-1}\right). \]

Thus the bounds in Theorem (2.4.1) become equalities and

(2.5.2) \[ E[\exp(sC_N)] = \frac{\Gamma(N+1)\Gamma(1/f^{+}(s))}{\Gamma(N+1)/f^{+}(s)} \text{ for } -\infty < s < \log\left(\frac{N}{N-1}\right). \]

Next rescale \( s \) to get a nontrivial limit as \( N \to \infty \). Replace \( s \) by \( s/N \). Since \( 1/N < \log\left(\frac{N}{N-1}\right) \), \( f^{+}(\frac{s}{N}) \) exists for all \( N \), for
\( -\infty < s \leq 1 \). An easy calculation shows \( f^+(\frac{s}{N})^{-1} = 1 - s + o(\frac{1}{N}) \) as \( N \to \infty \).

So (2.5.2) becomes

\[
(2.5.3) \quad E[\exp(\frac{s}{N}C_N)] = \frac{\Gamma(N+1)\Gamma(1-s+0(\frac{1}{N}))}{\Gamma(N+1-s+0(\frac{1}{N}))}, \quad -\infty < s < 1.
\]

Stirling's formula gives

\[
\frac{\Gamma(N+1)}{\Gamma(N+1-s+0(\frac{1}{N}))} = e^{s\log N (1+o(1))} \quad \text{as} \quad N \to \infty.
\]

Moving this term to the other side of (2.5.3),

\[
(2.5.4) \quad E[\exp(\frac{s}{N}C_N - N \log N)] \to \Gamma(1-s) \quad \text{as} \quad N \to \infty
\]

for \( -\infty < s < 1 \).

\( \Gamma(1-s) \) is the moment generating function of the extreme value distribution. Convergence of moment generating functions in a neighborhood of 0 implies convergence in distribution as well as convergence of all moments. This gives the standard result (see, e.g., Kolchin, et. al., 1978) of convergence of \( C_N \) to the extreme value distribution.

Notice that in this problem, the auxiliary randomization introduced in Section 2.2 is unnecessary. Taking \( S_i \) to be the first time states \( \{A_1,\ldots,A_j\} \) have been visited, rather than \( \{A_{\sigma_1},\ldots,A_{\sigma_j}\} \) gives the same results. This is because the labels on the states are arbitrary to begin with for \( j \neq 0 \); the probabilities of transitions from \( A_i \) to \( A_j \) are all equal. For random walks on finite groups, as considered in Chapters 3 and 4, this randomization will be necessary.

19
2. In the Coupon Collector's Problem as above, for \( j \neq 0 \), \( ET_i = N \).
Thus \( \mu^- = \mu^+ = N \). Theorem (3.2.1) implies

\[
EC_N = N \sum_{i=1}^{N} \frac{1}{i}.
\]

The extreme value distribution has mean \( -\Gamma'(1) = 0.57721566... \), Euler's constant (Gradshteyn and Ryzhik, 1980, 8.367). Differentiating (2.5.4),

\[
\frac{1}{N} (EC_N - \log N) = -\Gamma'(1) + o(1) \text{ as } N \to \infty.
\]

Hence

\[
\lim_{N \to \infty} \left( \sum_{i=1}^{N} \frac{1}{i} - \log N \right) = -\Gamma'(1) \text{ (ibid.)}
\]

3. In the Coupon's Collector's Problem, (2.5.1) stated

\[
f^+(s) = \frac{\frac{1}{N} e^s}{1 - (1 - \frac{1}{N}) e^s}.
\]

Thus \( f^+(\frac{s}{N}) = \frac{1}{1-s} + O(\frac{1}{N}) \) for \( -\infty < s < 1 \). \( \frac{1}{1-s} \) is the moment generating function of the exponential distribution. Hitting times, suitably normalized, have an asymptotic exponential distribution in this example. This was necessary for the extreme value character of the final results, since it gave \( \Gamma(1/f^+(\frac{s}{N})) \to \Gamma(1-s) \). It appears that such asymptotic exponentiality is necessary for obtaining the extreme value distribution as a limiting result in Theorem (2.4.1). The consideration of \( f^+(s) \) and \( f^-(s) \) requires something even stronger of hitting times, namely, in order to get a limiting extreme value distribution, the hitting times
must be in some sense uniformly asymptotically exponentially distributed.

Some work (Aldous (1983b), Flatto, Odlyzko and Wales (1983)), has been done showing that hitting times are asymptotically exponentially distributed for rapidly mixing Markov chains and random walks on finite groups. The results of Flatto, Odlyzko and Wales will be used in Section 4.4 on the permutation group.

4. Sometimes the bounds in Theorem (2.4.1) are not asymptotically tight using \( f^- (s) \) and \( f^+ (s) \), but are improvable. In Section 4.3 on the discrete cube, naive use of \( f^- (s) \) and \( f^+ (s) \) is not good enough. It will be necessary to consider the distance between \( X_{S_{i-1}} \) and \( A_{o_1} \). The mechanics of this will be considered in Section 4.3.

5. Naturally, the examples in this thesis will be problems where the bounds are given in this chapter give good results. However, the techniques work well only in a fairly narrow range of problems, loosely, those that are sufficiently rapidly mixing. In other examples, these bounds can be quite bad. As a single bad example, consider a deterministic random walk around on \( N \) point circle, the random walk moving one step to the right at a time. This process will take \( C_N = N-1 \) steps to visit all \( N-1 \) points of the circle, excepting its starting point. Clearly \( \mu^- = 1, \mu^+ = N-1 \), so Theorem (2.3.1) gives

\[
\sum_{i=1}^{N-1} \frac{1}{i} \leq EC_N \leq (N-1) \sum_{i=1}^{N-1} \frac{1}{i}.
\]
On the good side, there is an abundant supply of rapidly mixing processes. Processes that are intrinsically high dimensional, like a random walk on a sphere, or low dimensional processes that spread out rapidly tend to be rapidly mixing. Thus, the results of this chapter are applicable to a limited, but interesting and fairly wide, class of problems.
CHAPTER 3
PRELIMINARIES FOR FINITE GROUP PROBLEMS

3.1  INTRODUCTION

The aim of Chapters 3 and 4 is to prove results analogous to those of the Coupon Collector's Problem for random walks on finite groups. Together they investigate the time taken by certain random walks on finite groups to visit every element. The present chapter gives some necessary general machinery that will be applied to several examples in Chapter 4. This chapter first reviews some concepts from the representation theory of finite groups. Then the idea of a random walk on a finite group is discussed, along with its place in Markov chain theory. From this, generating functions for hitting times are developed. Finally, the subset of $\mathbb{R}$ where these generating functions are finite is investigated.

3.2  REVIEW OF FINITE GROUPS

Here some definitions and facts from the representation theory of finite groups are given. This section is not comprehensive enough to serve as a proper introduction for a person unfamiliar with the subject. For a proper introduction, the standard reference is Serre (1977). Naimark and Stern (1982) give a more detailed treatment. Diaconis (1982) gives uses of group theory in statistics. The aim of this section is to introduce the terminology to be used later and to list
some theorems for easy reference. All the results given here are standard; the material until Fourier transforms is all in Naimark and Stern (1982). The material on Fourier transforms can be found in Diaconis (1982).

THE BASICS

A group \((G, \cdot)\) is a nonempty set \(G\) with a product \(\cdot\) mapping \(G \times G\) into \(G\) that satisfies

1) **Associativity:** \((g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)\) for all \(g_1, g_2, g_3 \in G\).

2) **Identity:** There exists a unique element \(e\) in \(G\) such that \(g \cdot e = e \cdot g = g\) for all \(g \in G\).

3) **Inverses:** For each \(g \in G\), there exists a unique inverse to \(g\), denoted \(g^{-1}\), satisfying \(g^{-1} \cdot g = g \cdot g^{-1} = e\).

As usual \(g_1 \cdot g_2\) will often be written \(g_1g_2\) and \((G, \cdot)\) will be denoted simply by \(G\) when no misunderstanding is likely.

\((G, \cdot)\) is Abelian, or commutative, if \(g_1 \cdot g_2 = g_2 \cdot g_1\) for all \(g_1, g_2 \in G\).

The order of \((G, \cdot)\), written \(|G|\), is the number of elements in the set \(G\). \(G\) is finite if \(|G| < \infty\).

For two groups \((G_1, \cdot_1)\) and \((G_2, \cdot_2)\) a map \(\phi\) from \(G_1\) into \(G_2\) is a homomorphism if it agrees with the product structures on \(G_1\) and \(G_2\); if \(\phi(g_3 \cdot_1 g_4) = \phi(g_3) \cdot_2 \phi(g_4)\) for all \(g_3, g_4 \in G_1\).

CONJUGACY CLASSES

Two elements \(g_1\) and \(g_2\) in \(G\) are said to be conjugate if there is a \(g_3 \in G\) such that \(g_3^{-1} g_1 g_3 = g_2\). This is an equivalence relation
and thus partitions $G$ into equivalence classes, called conjugacy classes.

Examples

1) If $(G, \cdot)$ is Abelian, the notion of conjugacy classes is trivial, since every element of $G$ is in a conjugacy class by itself $g_3^{-1}g_1g_3 = g_3g_3^{-1}g_1 = g_1$, so $g_3^{-1}g_1g_3 = g_2$ implies $g_1 = g_2$.

2) The identity is the only member of its conjugacy class in any group, by a similar argument.

3) Two square matrices $A$ and $B$ are similar if there is a nonsingular square matrix $C$ such that $C^{-1}AC = B$. The notion of conjugacy is much like that of similarity for matrices.

REPRESENTATIONS

Let $(G, \cdot)$ be a finite group. Let $V$ be a vector space over some field $H$. For our purposes, $H$ will always be the complex numbers $C$ and $V$ will always have finite dimension. Let $GL(V)$ denote the group of invertible linear mappings of $V$ into itself. A representation $(\rho, V)$ of $(G, \cdot)$ is a homomorphism $\rho$ of $(G, \cdot)$ into $GL(V)$. In English, we represent the elements of $G$ by invertible linear operators (matrices, essentially) on $V$ that multiply in the same fashion as the elements of $G$.

The simplest example is $\rho(g)v = v$ for all $g \in G$, $v \in V$. Each $g$ is represented by the identity map from $V$ to $V$. This is easily seen to be a representation and is called a trivial representation.
A subspace \( W \) of \( V \) is invariant under \( G \) if \( \rho(g)w \in W \) for all \( g \in G \) and \( w \in W \). For example, in a trivial representation \( (\rho, V) \), every subspace is invariant. A representation \( (\rho, V) \) is irreducible if the only invariant subspaces of \( V \) are \( \{0\} \) and \( V \) itself. A trivial representation is irreducible only if \( V \) has dimension 1. The dimension \( d_\rho \) of an irreducible representation \( (\rho, V) \) is defined to be the dimension of \( V \).

Two irreducible representations \( (\rho, V_1) \) and \( (\rho_2, V_2) \) are equivalent if there is a linear isomorphism \( A \) mapping \( V_1 \) into \( V_2 \) such that

\[
A\rho_1(g)v = \rho_2(g)Av \quad \text{for all } g \in G, v \in V_1.
\]

Any two irreducible trivial representations are easily seen to be equivalent. This notion of equivalence is an equivalence relation and partitions all irreducible representations into equivalence classes. The number of irreducible representations of \( G \) can now be defined to be this number of equivalence classes. Three standard results can now be stated.

(3.2.1) The number of irreducible representations of a finite group \( G \) is equal to the number of conjugacy classes of \( G \). In particular, \( G \) has only finitely many irreducible representations.
(3.2.2) All equivalent irreducible representations of \( G \) have the same dimension \( d_\rho \). If \( \rho_i \) is any representative of the \( i^{th} \) equivalence class, then summing over all equivalence classes gives \( \sum_i (d_{\rho_i})^2 = |G| \). Thus the dimensions of irreducible representations are bounded.

(3.2.3) If \( G \) is Abelian, each irreducible representation of \( G \) is one-dimensional. It follows that there must be exactly \( |G| \) inequivalent irreducible representations of such \( G \).

Each equivalence class has some particular nice members. \( V \) can be chosen to be a Euclidean space with the standard inner product. Further \( (\rho, V) \) can be chosen such that all the operators \( \rho(g) \) are unitary. Choosing an orthonormal basis for \( V \), the operators \( \rho(g) \) become unitary \( d_\rho \times d_\rho \) matrices. The resulting matrices are not unique; a change of basis in \( V \) could change the matrices. However all that is necessary is to choose, from each equivalence class of irreducible representations, one such unitary representation \( (\rho, V) \) along with an orthonormal basis for \( V \). The matrices corresponding to these representations in these bases will from here on be referred to as the irreducible representations of \( G \). If \( G \) has \( K \) equivalence classes of irreducible representations, the matrices corresponding to the chosen representations will be denoted \( \rho_1, \rho_2, \ldots, \rho_K \). \( \rho_1 \) will be the trivial representation of \( G \), \( \rho_1(g) = 1 \) for all \( g \in G \).

\( d_1, d_2, \ldots, d_K \) will denote the dimensions of \( \rho_1, \ldots, \rho_K \). Thus \( d_1 = 1 \), since \( \rho_1(g) \) is always the \( 1 \times 1 \) identity matrix.
The character $\chi_k$ of a representation $\rho_k$ is $\chi_k(g) = \text{Trace}(\rho_k(g))$, the sum of the diagonal elements of $\rho_k(g)$. Several more standard results are, for $k = 1, \ldots, K$,

(3.2.4) $\rho_k(g^{-1}) = [\rho_k(g)]^*$, $A^*$ denoting the complex conjugate of the transpose of a matrix $A$, i.e.,

$$ (A^*)_{ij} = \overline{A_{ji}}. $$

(3.2.5) $\overline{\chi_k(g)} = \chi_k(g^{-1})$ — denoting complex conjugate.

(3.2.6) $|\chi_k(g)| \leq d_k$ for all $g \in G$.

(3.2.7) $\chi_k(g_1^{-1}g_2g_1) = \chi_k(g_2)$, for all $g_1, g_2$ in $G$.

In words, the characters of the irreducible representations are constant on conjugacy classes.

FUNCTIONS AND MEASURES ON FINITE GROUPS

Haar measure, or the uniform distribution on $G$ is the measure that assigns mass $|G|^{-1}$ to each element of $G$. There are many other probability measures $P$ on $G$; they need only satisfy $P(g) \geq 0$ for all $g \in G$ and $\sum_{g \in G} P(g) = 1$. A probability measure $P$ is constant on conjugacy classes if it is, i.e., if $P(g_1^{-1}g_2g_1) = P(g_2)$ for all $g_1$ and $g_2$ in $G$. 

28
More generally, let $L^2(G)$ denote the vector space of all complex-valued functions on $G$. In the obvious way, probability measures on $G$ can be considered elements of $L^2(G)$. The class functions are the elements of $L^2(G)$ which are constant on conjugacy classes. $L^2(G)$ has a natural inner product,

$$<f_1 | f_2> = \frac{1}{|G|} \sum_{g \in G} f_1(g) \overline{f_2(g)},$$

justifying the name $L^2(G)$. Convolution in $L^2(G)$ is defined by

$$(f_1 \ast f_2)(g_1) = \sum_{g \in G} f_1(g_1 g^{-1}) f_2(g)$$

for all $g_1 \in G$, $f_1, f_2 \in L^2(G)$. If $G$ is not Abelian, convolution is not commutative.

Group representation theory enters naturally into analysis on $L^2(G)$. The $K$ inequivalent unitary irreducible representations $\rho_1, \ldots, \rho_K$ selected above have $\sum_{k=1}^{K} d_k^2 = |G|$ matrix elements by (3.2.2). Each matrix element $\rho_k(g)_{ij}$ can be considered as an element of $L^2(G)$. The following is often called the Peter-Weyl Theorem.

Theorem (3.2.8)

$$<\rho_k(\cdot)_{ij}, \rho_{k'}(\cdot)_{i'j'}> = \begin{cases} \frac{1}{d_k} & \text{if } i=i', j=j', k=k' \\ 0 & \text{otherwise} \end{cases}$$

Thus $\{\sqrt{d_k} \rho_k(\cdot)_{ij} \}_{k=1,2,\ldots,K, \ i,j=1,\ldots,d_k}$ form an orthonormal system.
in $L^2(G)$. Since the dimension of $L^2(G)$ is $|G|$, \( \{ \sqrt{d_k} \rho_k(\cdot) \}_{ij} \)
form an orthonormal basis of $L^2(G)$. Also, the characters $\chi_1, \ldots, \chi_K$
form an orthonormal basis for the $K$ dimensional subspace of $L^2(G)$
consisting of class functions. Another interesting property of characters is

\[
(3.2.9) \quad \sum_{k=1}^K d_k \chi_k(g_1) \overline{\chi_k(g_2)} = \begin{cases} |G| & \text{if } g_1 = g_2 \\ 0 & \text{otherwise} \end{cases}.
\]

There are many orthonormal bases of $L^2(G)$, but this choice of
\( \{ \rho_k(\cdot) \}_{ij} \) is most useful because it interacts nicely with the convolution structure of $L^2(G)$. The irreducible representations \( \{ \rho_k \} \) behave very much like the exponentials in standard Fourier transform.

Next, a Fourier transform for elements of $L^2(G)$ must be defined. There is really only one possibility, though different normalizations are possible. The version in Diaconis (1982), rather than the one of Naimark-Stern (1982), will be used here. For $f \in L^2(G)$, \( \rho_k \) one of the chosen unitary irreducible representations of $G$, let \( \rho_k(f) \), often called
\( \hat{f}(\rho_k) \) be

\[
(3.2.10) \quad \rho_k(f) = \sum_{g \in G} f(g) c_k(g).
\]

\( \rho_k(f) \) is called the Fourier transform of $f$ at $\rho$. For example, for the trivial representation $\rho_1$, \( \rho_1(f) = \sum_{g \in G} f(g) \cdot 1 = \sum_{g \in G} f(g) \). For any probability measure $\mu$ on $G$, \( \rho_1(\mu) = 1 \). Also, \( \chi_k(f) \) can be defined by \( \chi_k(f) = \text{Trace}(\rho_k(f)) = \sum_{g \in G} f(g) \chi_k(g) \). An important fact is
Theorem (3.2.11) For $f$ a class function, $d_k$ the dimension of $ho_k$, of the irreducible representations, $I_k$ the $d_k$ by $d_k$ identity matrix

$$
\rho_k(f) = \frac{\chi_k(f)}{d_k} I_k .
$$

$\rho_k(f)$ is diagonal with identical diagonal entries. The Fourier transform acts on convolutions in the usual way.

(3.2.12) $\rho_k(f_1 \ast f_2) = \rho_k(f_1) \rho_k(f_2)$ for $f_1, f_2 \in L^2(G)$.

There is also an inversion formula.

(3.2.13) For $f \in L^2(G)$, $g \in G$, $\rho_1, \ldots, \rho_K$ the irreducible representations of $G$,

$$
f(g) = \frac{1}{|G|} \sum_{k=1}^{K} d_k \text{Trace}[\rho_k(f) \rho_k(g^{-1})].
$$

Finally, there is the Plancherel Theorem.

(3.2.14) For $f \in L^2(G)$, $\rho_1, \ldots, \rho_K$ the irreducible representations of $G$,

$$
\sum_{g \in G} |f(g)|^2 = \frac{1}{|G|} \sum_{k=1}^{K} \text{Trace}[\rho_k(f) \rho_k(f)^*].
$$
Recall that the irreducible representations \( \rho_1, \ldots, \rho_K \) used above are not unique, but are unitary choices from equivalence classes. This non-uniqueness is not a problem; the above theorems are true for any unitary choices \( \rho_1, \ldots, \rho_K \). It turns out that the characters of the irreducible representations are unique.

3.3 RANDOM WALKS ON FINITE GROUPS AS MARKOV CHAINS

This section describes the notion of a random walk on a finite group and its properties as a Markov chain. The nicest property is that the theory of group representations leads to either an eigenvalue eigenvector decomposition of its transition matrix or something very similar.

Let \((G, \cdot)\) be a finite group with elements \(\{g_1, \ldots, g_N\}\), \(N = |G|\). Suppose \(\mu\) is a probability measure on \(G\). From \(\mu\), construct an \(N \times N\) one step transition matrix \(P\) by setting \(P_{ij} = \mu(g_j g_i^{-1})\). From here on, assume all states \(g_1, \ldots, g_N\) communicate, or equivalently, that the support of \(\mu\) generates \(G\). If this is not the case, then the support of \(\mu\) generates a subgroup \(H\) of \(G\) and a random walk on \(H\) could be considered instead. It is easily verified that \(P\) is a proper, in fact doubly stochastic, transition matrix. The Markov chain with this transition matrix will be called the random walk on \(G\) generated by \(\mu\).

The interpretation of \(P_{ij} = \mu(g_j g_i^{-1})\) is as follows. Fix \(X_0 \in G\) and consider \(X_1, X_2, \ldots\) independent, all with the same distribution \(\mu\) on \(G\). Let \(S_0 = X_0, S_1 = X_1 \cdot S_0, S_2 = X_2 \cdot S_1\) and in general \(S_n = X_n \cdot S_{n-1}\). If \(G\) were \(\mathbb{R}\) and \(\cdot\) were addition, this would be an ordinary random walk. The above process on \(G\) is a natural generalization of the notion
of random walks to more general groups. Suppose $S_{n-1} = g_i$. Since $S_n = X_n S_{n-1}$, $P(S_n = g_j \mid S_{n-1} = g_i) = P(X_n = g_j g_i^{-1}) = \mu(g_j g_i^{-1})$. Thus the Markov chain formulation in the previous paragraph describes the transition mechanism of this process.

Much is known about finite Markov chains, so an advantageous method of studying random walks on finite groups is by studying them as Markov chains or, in other words, by studying the matrix $P$. A primary object of interest is the decomposition of $P$ into eigenvectors and eigenvalues. This cannot be done in general, but for random walks on finite groups, group representation theory is a great help. If the representation theory of the group of interest has been adequately worked out, then such a decomposition, or something very similar, is immediately available. In fact, a rearrangement of the Fourier inversion formula gives this decomposition, a coordinate version of results in Diaconis and Shashahani (1981).

Recall that $P_{ij} = \mu(g_j g_i^{-1}) = \sum_{k=1}^{K} \frac{d_k}{|G|} \text{Trace}(\rho_k(\mu) \rho_k[(g_j g_i^{-1})^{-1}])$

(3.3.1)

by Fourier inversion (3.2.13). Here $\rho_1, \ldots, \rho_K$ are the irreducible representations of $G$ chosen in Section 3.2 and $d_1, \ldots, d_K$ their dimensions. $\rho_k[(g_j g_i^{-1})^{-1}] = \rho_k(g_i) \rho_k(g_j^{-1})$ by the definition of representation. The trace of a product is invariant under cyclic permutations, so (3.3.1) is

$\sum_{k=1}^{K} \frac{d_k}{|G|} \text{Trace}(\rho_k(g_j^{-1}) \rho_k(\mu) \rho_k(g_i))$.

Writing $\text{Trace}$ as the sum of diagonal elements this is
\begin{equation}
(3.3.2) \quad \frac{d_k}{|G|} \sum_{k=1}^K \sum_{j=1}^{d_k} \left[ \rho_k(g_j^{-1}) g_1^t, \ldots, \rho_k(g_j^{-1}) g_{d_k}^t \right] \rho_k(\mu) \left[ \rho_k(g_i_1), \ldots, \rho_k(g_i_{d_k}) \right]^t,
\end{equation}

\text{t denoting transpose. Let } \tilde{\rho}_k(\mu) \text{ be the } d_k^2 \text{ by } d_k^2 \text{ block diagonal matrix}

\begin{bmatrix}
\rho_k(\mu) & 0 \\
0 & \rho_k(\mu) \\
\rho_k(\mu) & \ddots \\
0 & \ddots & \ddots & 0 \\
\rho_k(\mu) & \ddots & \ddots & \ddots & \ddots
\end{bmatrix}

\text{with } d_k \text{ diagonal blocks, each a copy of } \rho_k(\mu) \text{ and zeros elsewhere.}

Let } \psi_k(g) \text{ be a column vector of length } d_k^2 \text{ created by stringing together the columns of } \rho_k(g).

\begin{equation}
\psi_k(g) = (\rho_k(g)_{11}, \rho_k(g)_{21}, \ldots, \rho_k(g)_{d_k1}, \rho_k(g)_{12}, \ldots, \rho_k(g)_{d_k2}, \ldots, \rho_k(g)_{d_k(d_k-1)}, \rho_k(g)_{d_k(d_k-2)}, \ldots, \rho_k(g)_{d_k1})^t.
\end{equation}

By (3.2.4) \( \rho_k(g)_{ab} = \rho_k(g^{-1})_{ba} \), so the row vector \( \psi_k(g)^* \) is the vector created by stringing together the rows of \( \rho_k(g^{-1}) \). (3.3.2) becomes

\begin{equation}
P_{ij} = \frac{d_k}{|G|} \sum_{k=1}^K \psi_k(g_j)^* \tilde{\rho}_k(\mu) \psi_k(g_i).
\end{equation}

Let } \phi_k(g) = \sqrt{d_k/|G|} \psi_k(g). \text{ Then}

\begin{equation}
(3.3.3) \quad P_{ij} = \frac{1}{\sqrt{d_k/|G|}} \sum_{k=1}^K \phi_k(g_j)^* \tilde{\rho}_k(\mu) \phi_k(g_i).
\end{equation}

34
(3.3.4) Let $\Lambda(\mu)$ be the $|G|$ by $|G|$ block diagonal matrix with diagonal blocks $\hat{\beta}_1(\mu), \ldots, \hat{\beta}_k(\mu)$ and zeros elsewhere. Let $\phi(g)$ be the column vector of length $|G|$ obtained by stringing together $\phi_1(g), \ldots, \phi_k(g)$.

\[ \phi(g) = (\phi_1(g)^t, \phi_2(g)^t, \ldots, \phi_k(g)^t)^t. \]

Then $P_{ij}$ is simply $\phi(g_j)^* \Lambda(\mu) \phi(g_i)$.

Letting $\phi$ be the $|G|$ by $|G|$ matrix $(\phi(g_1), \ldots, \phi(g_N))$, the matrix $P$ is

(3.3.5) \[ [\phi^* \Lambda(\mu) \phi]^t. \]

By the Peter-Weyl Theorem (3.2.8), $\{\sqrt{d_k} \rho_k(g)_{ab}\}, k=1, \ldots, K, a, b=1, \ldots, d_k$ are orthonormal functions on $G$. The rows of $\phi$ are just these functions divided by $\sqrt{|G|}$ and are thus orthonormal in the usual sense, as vectors in $C|G|$. So $\phi^* \phi = I |G|$, since $P$ is real, $P^* = P^t$, hence $P = \phi^* \Lambda^* (\mu) \phi$ is a decomposition of $P$ much like the traditional eigenvector-eigenvalue decomposition. In fact, if $\Lambda(\mu)$ is diagonal, $\phi^* \Lambda^* (\mu) \phi$ is precisely the decomposition of $P$ into eigenvectors and eigenvalues. $\Lambda(\mu)$ will certainly be diagonal if $G$ is Abelian or if $\mu$ is constant on conjugacy classes, since then all the $\rho_k(\mu)$ will be either one dimensional or diagonal, respectively. In other cases $\phi^* \Lambda^* (\mu) \phi$ may have $\Lambda^*(\mu)$ block diagonal, so it may not
give the complete decomposition of $P$. The applications considered in
Chapter 4 will have $\Lambda^*(\mu)$ diagonal. Nonetheless, the decomposition
(3.3.5) can be useful in cases with non-diagonal $\Lambda^*(\mu)$.

3.4 GENERATING FUNCTIONS

Using the decomposition (3.3.3) of $P$ obtained in Section 3.3,
a standard result in the theory of finite Markov chains gives the
generating function of the time taken by a random walk to hit one state
from another.

Let $z$ be a complex number, $|z| < 1$. Then $P(z) = \sum_{n=0}^{\infty} (zP)^n$
is well defined and equal to

\[
(3.4.1) \quad \sum_{n=0}^{\infty} [\phi^*(z\Lambda)^n \phi]^t = (\phi^*(I-z\Lambda)^{-1}\phi)^t.
\]

Let $F_{ij}^{(n)} = P(S_n = g_j, S_{n-1}, \ldots, S_1, S_0 = g_i)$ be the probability
that $g_j$ is first hit at time $n$ starting at $g_i$. Let $F_{ij}(z) = \sum_{n=1}^{\infty} F_{ij}^{(n)} z^n$
be the probability generating function of this hitting time.

A standard result (Kemperman, 1961, p. 18-19) is that for $i \neq j$,

$F_{ij}(z) = \frac{P_{ij}(z)}{P_{jj}(z)}$, where $P_{ij}(z)$ and $P_{jj}(z)$ are the entries of
the matrix (3.4.1). The derivation of (3.3.3) in reverse gives for $i \neq j$,

\[
(3.4.2) \quad F_{ij}(z) = \frac{\sum_{k=1}^{K} d_k \text{Trace}[\rho_k (g_j g_j^{-1})(I-z\rho_k(\mu))^{-1}]}{\sum_{k=1}^{K} d_k \text{Trace}[(I-z\rho_k(\mu))^{-1}]}, \quad |z| < 1.
\]

This formula has appeared a number of times. I.J. Good (1951) gives it
for random walks on finite Abelian groups. Flatto, Odlyzko, and Wales

36
(1985) derive it in a different manner.

If \( \mu \) is constant on conjugacy classes (3.4.2) simplifies considerably, as pointed out by Flatto, Odlyzko and Wales. Using (3.2.11), for \( i \neq j \),

\[
F_{ij}(z) = \sum_{k=1}^{K} \frac{d_k}{d_k} \left( \frac{1-zx_k(\mu)/d_k}{1-zx_k(\mu)/d_k} \right) \chi_k(g_jg_i^{-1}) |z| < 1.
\]

In the language of Chapter 2, let \( T_{ij} \) be the time to hit \( g_j \) from \( g_i \). Then \( F_{ij}(z) = E(z^{T_{ij}}) \) and \( f_{ij}(s) = E[\exp(sT_{ij})] = F_{ij}(e^s) \).

3.5 RANGE OF EXISTENCE OF GENERATING FUNCTIONS

In Section 3.4 the probability generating function \( F_{ij}(z) \) of the time taken by a random walk on \( G \) to hit \( g_j \) from \( g_i \) was derived, but only for \( |z| < 1 \). Much more powerful results will be made available by knowing the existence and value of \( F_{ij}(z) \) for all \( |z| < 1+\varepsilon \) for some \( \varepsilon > 0 \). Let \( F_{ij}^*(z) \) be the rational function defined by (3.4.2) for all \( z \). \( F_{ij}^*(z) \) may have poles; let \( z_0 \) be the pole of \( F_{ij}^*(z) \) of smallest magnitude. In this section, it will be shown that \( F_{ij}(z) \) exists and is equal to \( F_{ij}^*(z) \) for \( |z| < |z_0| \). The value of \( |z_0| \) will also be investigated.

The range of existence is a simple application of analytic continuation. Recall a standard result of complex analysis (Ahlfors, 1970, p. 120).

Lemma (3.5.1). Suppose \( F(z) \) is analytic in the ball \( \{z:|z|<y\} \).

For \( 0 < x < y \)
\[ \frac{d^n}{dz^n} F(z) \bigg|_{z=0} = \frac{n!}{2\pi i} \oint_{|z|=x} \frac{F(z)}{z^{n+1}} \, dz. \]

Further, \( F \) is bounded on \( \{|z|=x\} \), hence

\[ \left| \frac{d^n}{dz^n} F(z) \bigg|_{z=0} \right| \leq \frac{n!}{x^{n+1}} \sup_{\{|z|=x\}} |F(z)|. \]

The following proposition is an easy consequence of this lemma.

Proposition (3.5.2). Let \( z_0 \) be as above, the smallest pole, in magnitude, of \( F^*_i_j(z) \). Then for \( 1 < |z| < |z_0| \), \( F^*_i_j(z) = \sum_{n=1}^\infty z^n F^*_i_j(n) \) exists and is given by (3.4.2).

Proof.

\[ \frac{d^n}{dz^n} F^*_i_j(z) \bigg|_{z=0} = n! F^*_i_j(n). \]

By the lemma, taking \( x = (|z|+|z_0|)/2 \),

\[ F^*_i_j(n) \leq \left( \frac{2}{|z|+|z_0|} \right)^{n+1} \sup_{|z|=|z_0|/2} |F^*_i_j(x)|. \]

Therefore \( \sum_{n=1}^\infty z^n F^*_i_j(n) \) converges absolutely. So \( F^*_i_j(z) \) is analytic for \( |z| < |z_0| \) and, by the uniqueness of analytic continuations, \( F^*_i_j(z) = F^*_i_j(z) \) for \( |z| < |z_0| \).

\( |z_0| \) now becomes the object of interest. After some notation is given, theorem (3.5.5) gives a lower bound on \( |z_0| \).

Let \( |rI-\Lambda(\mu)| = \text{determinant} \ (rI-\Lambda(\mu)) \) be the characteristic polynomial of \( \Lambda(\mu) \), \( \Lambda(\mu) \) defined by (3.3.4).
For a square matrix $A$, let $\text{Cof}(A)$ denote the matrix of cofactors of $A$. If $A^{(i,j)}$ is the matrix $A$ with the $i$\textsuperscript{th} row and $j$\textsuperscript{th} column deleted, then $[\text{Cof}(A)]_{ij} = (-1)^{i+j} |A^{(i,j)}|$. Cramer's rule (Gantmacher, 1959) is

\[(3.5.3) \quad \text{If } |A| \neq 0 \text{ then } A \text{ is invertible and } A^{-1} = [\text{Cof}(A)]^t / |A|, \text{ } t \text{ denoting transpose.}\]

Next, the Jordan Normal Form must be introduced. Let $I_n$ denote the $n \times n$ identity matrix and $J_n$ denote the $n \times n$ matrix with ones on the first diagonal above the main diagonal and zeros elsewhere. A square matrix $A$ is in Jordan Normal Form if $A$ is block diagonal with diagonal blocks $A_1, \ldots, A_k$, for some $k$, with each diagonal block $A_i$ satisfying $A_i = \lambda_i I_{n_i} + J_{n_i}$ for some complex $\lambda_i$ and integer $n_i$. Another standard result (Gantmacher, 1959) is

\[(3.5.4) \quad \text{For any } n \times n \text{ matrix } B \text{ there are } n \times n \text{ matrices } A \text{ and } U, \text{ with } A \text{ in Jordan Normal Form and } U \text{ unitary, such that } B = U A U^*. \text{ In other words, every square matrix is unitarily similar to a matrix in Jordan Normal Form.}\]

If $\rho(\cdot)$ is an irreducible representation and $U$ is unitary, then $U^* \rho(\cdot) U$ is an equivalent irreducible representation. From (3.5.4), with proper choices of members of equivalence classes of irreducible representations, $\rho_1(\mu), \ldots, \rho_k(\mu)$ can be taken to be in Jordan Normal
Form. In turn, each \( \hat{\phi}_k(\mu) \) and hence \( \Lambda(\mu) \) of (3.3.4) will be in Jordan Normal Form as well.

Finally, let \( Q(r) \) be the minimal polynomial of \( \Lambda(\mu) \); the smallest polynomial such that \( Q(\Lambda(\mu)) = 0 \). \( Q(r) \) divides \( |rI-\Lambda(\mu)| \) (Gantmacher, 1959). With \( \Lambda(\mu) \) in Jordan Normal Form it is easy to determine \( Q(r) \).

Suppose \( |rI-\Lambda(\mu)| = \prod_{i=1}^{P} (r-\lambda_i)^{m_i} \); \( \Lambda(\mu) \) has eigenvalues \( \lambda_1, \ldots, \lambda_P \) with multiplicities \( m_1, \ldots, m_P \). Let \( n_i \) be the dimension of the largest block of the form \( \lambda_i I_{n_i} + J_{n_i} \) occurring in \( \Lambda(\mu) \). Then clearly \( n_i \leq m_i \), and

\[
Q(r) = \prod_{i=1}^{P} (r-\lambda_i)^{n_i}. \quad \text{Note that if } \Lambda(\mu) \text{ is diagonal, } Q(r) = \prod_{i=1}^{P} (r-\lambda_i). \]

Theorem (3.5.5). Let \( |rI-\Lambda(\mu)| \) be the characteristic polynomial for a random walk \( \mu \) on a finite group \( G \). Let \( Q(r) \) be the minimal polynomial of \( \Lambda(\mu) \). Let \( |r_0| \) be the magnitude of the largest zero of \( Q(r) \frac{d}{dr} \log |rI-\Lambda(\mu)| \). Then \( |z_0| \leq 1/|r_0| \), or, \( F_{ij}(z) \) exists, for all \( g_i \neq g_j \in G \), for all \( z \) with \( |z| < 1/|r_0| \).

Proof. Recall from (3.3.4) that \( \Lambda(\mu) \) is block diagonal with blocks \( \hat{\phi}_1(\mu), \ldots, \hat{\phi}_K(\mu) \). For \( g \in G \), similarly define \( \Lambda(g) \) to be block diagonal with blocks \( \hat{\phi}_1(g), \ldots, \hat{\phi}_K(g) \). (3.4.2) becomes

\[
(3.5.6) \quad F_{ij}(z) = \frac{\text{Trace}[\Lambda(g_i^{-1}g_j^{-1})(rI-\Lambda(\mu))^{-1}]}{\text{Trace}[(rI-\Lambda(\mu))^{-1}]} \quad \text{for all } r,
\]

where \( r = 1/z \).

Note an easy fact about matrices.

Proposition (3.5.7). Suppose \( rI-B \) is invertible. Then

\[
\frac{d}{dr} |rI-B| = |rI-B| \text{Trace}[(rI-B)^{-1}].
\]
Proof of Proposition. \(|rI-B| \) and \( \text{Trace}\ (rI-B)^{-1} \) are invariant under similarity transformations, so it suffices to consider \( B \) in Jordan Normal Form. Then \( |rI-B| = \prod_{i=1}^{N}(r-b_i) \), \( b_i \) the \( N \) diagonal elements of \( B \). If \( B \) is diagonal the result is obvious. If not, it is sufficient to note that the diagonal elements of \( (rI-B)^{-1} \) are still \( (r-b_i)^{-1} \).

Using (3.5.7) and the simple fact that

\[
\text{Trace}(AB^{-1}) = |B|^{-1} \text{Trace}[A(\text{Cof}(B))^t],
\]

(3.5.6) becomes

\[
F_{ij}^{*}(z) = \frac{\text{Trace}[\Lambda(g_j g_i^{-1})\text{Cof}(rI-\Lambda(\mu))^t]}{\frac{d}{dr}|rI-\Lambda(\mu)|}, \quad \text{where} \quad r = \frac{1}{z}.
\]

This will lead to the assertion of the theorem. \( F_{ij}^{*}(z) \) is represented as a ratio of polynomials in \( r = 1/z \). Let \( r_1 \) be the zero of the denominator of (3.5.8) of largest magnitude. \(|r_1| \leq 1\) since the eigenvalues of \( \Lambda(\mu) \), as eigenvalues of a transition matrix, lie in the unit ball of \( C \) and from the fact (see Marden, 1966) that the zeros of the derivative of a polynomial lie in the convex hull of the zeros of the polynomial. \( F_{ij}^{*} \) cannot have any poles for \(|r_1| < r \leq 1\). \( F_{ij}^{*}(z) \) is already known to be analytic for \(|z| < 1\). The above argument shows \( F_{ij}^{*}(z) \) is analytic for \(|z| < 1/|r_1|\).

In the same manner, the theorem can be proved once it is known that \( Q(r)^{-1}|rI-\Lambda(\mu)| \) divides both the numerator and denominator of (3.5.8). Dividing both by \( Q(r)^{-1}|rI-\Lambda(\mu)| \), the numerator will still be a polynomial in \( r \), while the denominator will become \( Q(r) \frac{d}{dr} \log|rI-\Lambda(\mu)| \).
With \( r_0 \) the largest zero of \( Q(r) \frac{d}{dr} \log|rI-\Lambda(\mu)| \), the theorem will be proven as for \( r_1 \) above.

Again let \( \lambda_1, \ldots, \lambda_p \) be the eigenvalues \( \Lambda(\mu) \), with multiplicities \( m_1, \ldots, m_p \). \( |rI-\Lambda(\mu)| = \prod_{i=1}^{p} (r-\lambda_i)^{m_i} \), so \( \frac{d}{dr} |rI-\Lambda(\mu)| \) is divisible by \( \prod_{i=1}^{p} (r-\lambda_i)^{m_i-1} \). From the definition of \( Q(r) \), \( Q(r) \) divides \( |rI-\Lambda(\mu)| \) and is divisible by \( \prod_{i=1}^{p} (r-\lambda_i) \). Thus \( Q(r)^{-1} |rI-\Lambda(\mu)| \) divides \( \prod_{i=1}^{p} (r-\lambda_i)^{m_i-1} \) and hence divides \( \frac{d}{dr} |rI-\Lambda(\mu)| \) as well. So \( Q(r) \frac{d}{dr} \log|rI-\Lambda(\mu)| \) is a polynomial.

Next consider the numerator of (3.5.8), which is

\[
\text{(3.5.9)} \quad \text{Trace}[\Lambda^{-1}(g_i g_j^t) \text{Cof}(rI-\Lambda(\mu))^t].
\]

It is sufficient to show that every element of \( \text{Cof}(rI-\Lambda(\mu)) \) is divisible by \( Q(r)^{-1} |rI-\Lambda(\mu)| \), for then (3.5.9) will be as well. This is easiest to show visually. \( rI-\Lambda(\mu) \) is block diagonal, with blocks of the form

\[
\text{(3.5.10)} \quad \begin{bmatrix}
  r-\lambda & -1 & 0 \\
  \vdots & \ddots & \vdots \\
  \vdots & & -1 \\
  0 & & r-\lambda
\end{bmatrix}
\]

An entry of \( \text{Cof}[rI-\Lambda(\mu)] \) is obtained by deleting a row and a column from \( rI-\Lambda(\mu) \) and taking the determinant of the remainder. If a deletion removes a row from one block and a column from a different block, the determinant of the remaining matrix is zero. For example, if any column of (3.5.10) is deleted, the rows corresponding to (3.5.10) are linearly dependent. So for an entry of \( \text{Cof}[rI-\Lambda(\mu)] \) to be nonzero, the corresponding deletions
must affect the same block. If the block is \( n \times n \) with diagonal entry \( \lambda \), then deleting a row and column of the block makes the resulting determinant smaller than \( |rI-\Lambda(\mu)| \) by a multiple of \( (r-\lambda)^n \), possibly as large a multiple as \( (r-\lambda)^n \). But if \( \Lambda(\mu) \) contains this \( n \times n \) block, then \( (r-\lambda)^n \) divides \( Q(r) \). So \( Q(r)^{-1}|rI-\Lambda(\mu)| \) divides this entry of \( \text{Cof}(rI-\Lambda(\mu)) \). Thus \( Q(r)^{-1}|rI-\Lambda(\mu)| \) divides (3.5.9), the numerator of (3.5.8). \( F^*_ij(s) \) is thus a rational function with denominator \( Q(r) \frac{d}{dr} \log|rI-\Lambda(\mu)| \). An argument identical to the one for (3.5.8) completes the proof.

This theorem, along with proposition (3.5.2) shows that \( F_{ij}(z) \) is given by (3.4.2) for \( |z| < 1/|r_0| \) for all \( i \neq j \). An assumption used implicitly here is that all states of the Markov chain communicate. This implies that \( T_{ij}^* \), the time to hit \( g_j \) from \( g_i \), is finite with probability one, and hence that

\[
T_{ij}^* = \sum_{n=1}^{\infty} F_{ij}(n) z^n \quad \text{for} \quad |z| < 1/|r_0| .
\]

It is possible that \( F_{ij}(z) \) is given by (3.4.2) for some \( z \) with \( |z| \geq 1/|r_0| \), in particular if there is further cancellation in the numerator and denominator of (3.4.2). However, this seems to be the best result that is true in general.

Theorem (3.5.5) gives \( r_0 \) as the largest zero of a polynomial. Finding \( r_0 \) is practice seems to be quite difficult; it involves finding the zeros of the derivative of a polynomial. Much of Marden (1966) is devoted to exact solutions and bounds on solutions to this
problem. In one special case the problem is much easier. This is the case of \( \mu \) symmetric, which is considered next.

A probability measure \( \mu \) on \( G \) is symmetric if \( \mu(g) = \mu(g^{-1}) \) for all \( g \in G \). Then the transition matrix \( P \) is symmetric as well. \( P \) is then diagonalizable and has real eigenvalues. Let \( \lambda_1 < \lambda_2 < \cdots < \lambda_p \) be the distinct eigenvalues of \( P \), \( \lambda_1 \) occurring with multiplicity \( m_1 \).

From basic Markov Chain Theory \( |\lambda_i| \leq 1 \) for \( i=1, \ldots, p \) and \( \lambda_p = 1 \), and \( m_p = 1 \) since all states of the chain communicate. Since \( P \) is diagonalizable, \( Q(r) = (r-\lambda_1)(r-\lambda_2)\cdots(r-\lambda_p) \) is the minimal polynomial of \( P \).

Similarly \( |rI-A(\mu)| \), the characteristic polynomial of \( P \) is

\[
\prod_{i=1}^{p} (r-\lambda_i)^{m_i}.
\]

Thus

\[
Q(r) \frac{d}{dr} \log|rI-A(\mu)| = \prod_{i=1}^{p} (r-\lambda_i) \sum_{j=1}^{p} \frac{m_i}{r-\lambda_j}.
\]

(3.5.11)

The zeros of (3.5.11) are easy to characterize. (3.5.11) is, up to a factor of \( \prod_{i=1}^{p} (r-\lambda_i)^{m_i-1} \), equal to \( \frac{d}{dr} |rI-A(\mu)| \). \( \frac{d}{dr} |rI-A(\mu)| \), as the derivative of a polynomial with real zeros, has real zeros. In fact, their locations are known (Marden, 1966). \( \frac{d}{dr} |rI-A(\mu)| \) has a zero of multiplicity \( m_i-1 \) at \( \lambda_i \) and one of its remaining \( p-1 \) zeros in each of the intervals \( (\lambda_i, \lambda_{i+1}) \), \( i=1, \ldots, p-1 \). (3.5.11) is just \( \frac{d}{dr} |rI-A(\mu)| \) divided by \( \prod_{i=1}^{p} (r-\lambda_i)^{m_i-1} \) and hence has one of its \( p-1 \) zeros in each of the intervals \( (\lambda_i, \lambda_{i+1}) \), \( i=1, \ldots, p-1 \). Thus \( r_0 \), its zero of largest magnitude, lies in either \( (\lambda_1, \lambda_2) \) or \( (\lambda_{p-1}, 1) \).

This gives an easy criterion to check, for given \( z \), whether \( |z| < 1/|r_0| \), or whether \( F_{ij}(z) \) exists for all \( i \neq j \). This is the content of the following theorem.

44
Theorem (3.5.12). Given $z \in \mathbb{C}$, $|z| \geq 1$, $\lambda_1, \ldots, \lambda_{p-1}$ the ordered distinct eigenvalues of $p$ (3.3.5), for $P_{ij}(z)$ to exist for all $i \neq j$, it is sufficient that the following conditions be satisfied.

i) $1/|z| > \lambda_{p-1}$.

ii) $-1/|z| < \lambda_2$.

iii) $1+(1-|z|) \prod_{i=1}^{p-1} \frac{m_i}{1-|z|\lambda_i} > 0$.

iv) $-1/|z| < \lambda_1$ or $m_1 + (1+\lambda_1|z|) \prod_{i=2}^{p} \frac{m_i}{1+|z|\lambda_i} > 0$.

Proof. (3.5.11) has a zero in each of the intervals $(\lambda_1, \lambda_2)$, $(\lambda_{p-1}, 1)$. Conditions i) and ii) guarantee that for any other zero $r'$ of (3.5.11), $|z| < 1/|r'|$. Consider the zero $r_1$ in the interval $(\lambda_{p-1}, 1)$. iii) should imply that $|z| < 1/|r_1|$. The polynomial $1 + \prod_{i=1}^{p} (r-\lambda_i) \Sigma_{i=1}^{p} \frac{m_i}{r-\lambda_i}$ has one zero, hence one change of sign, in the interval $(\lambda_{p-1}, 1)$. $\prod_{i=1}^{p-1} (r-\lambda_i)$ has a constant sign throughout $(\lambda_{p-1}, 1)$. Hence the problem is reduced to finding where

$$
(3.5.13) \quad (r-1) \prod_{i=1}^{p} \frac{m_i}{r-\lambda_i} = 1 + (r-1) \prod_{i=1}^{p-1} \frac{m_i}{r-\lambda_i}
$$

changes sign. As $r \to \infty$ (3.5.13) is clearly positive. Since (3.5.13) changes sign at $r_1$, if (3.5.13) is positive for some $r$ in $(\lambda_{p-1}, 1)$, then $r > r_1$. Thus if (3.5.13) is positive at $r = 1/|z|$, then $|z| < 1/|r_1|$. Condition iii) is a restatement of this positivity, hence iii) implies $|z| < 1/|r_1|$. Analogously, for $r_2$ the zero of (3.5.11) in $(\lambda_1, \lambda_2)$, if $-1/|z| < \lambda_1$, then $|z| < 1/|r_2|$ is trivially true. Otherwise,

45
\( \lambda_1 < 1/|z| < \lambda_2 \), and iv) implies \( |z| < 1/|r_2| \).

Thus if \( z \) satisfies i)-iv) then \( |z| < 1/|r_0|, |r_0| \) the magnitude of the largest zero of (3.5.11). Theorem (3.5.5) completes the proof.

3.6 THE RATIO OF GAMMA FUNCTIONS

Chapter Three concludes with a simple application of Stirlings formula.

Proposition (3.6.1). Suppose \( Y_N = O(1) \) as \( N \to \infty \) then

\[
\frac{Y_N}{\Gamma(N+Y_N)} = 1 \text{ as } N \to \infty.
\]

Proof. Write

\[
\log \Gamma(N) = N \log N - N - \frac{1}{2} \log 2\pi + O\left( \frac{1}{N} \right)
\]

\[
\log \Gamma(N+Y_N) = (N+Y_N) \log (N+Y_N) - (N+Y_N) - \frac{1}{2} \log (N+Y_N)
\]

\[
+ \log \sqrt{2\pi} + O\left( \frac{1}{N} \right)
\]

expand

\[
\log (N+Y_N) \text{ as } \log N + \frac{Y_N}{N} + O\left( \frac{1}{N^2} \right)
\]

and the result follows.
CHAPTER 4
FINITE GROUP PROBLEMS

4.1 INTRODUCTION

This chapter applies the results of Chapters 2 and 3 to random walks on families of finite groups. For certain interesting examples bounds on the moment generating function of \( C \), the time taken by the random walk to cover the group, are derived. In these examples the bounds are asymptotically tight (as the order of the group grows) and give the extreme value distribution as the limiting distribution of \( C \). All the examples considered are rapidly mixing.

The setup of each problem is as follows. \( G \) is a finite group, the random walk starts at some \( g_0 \in G \). \( C \) is the time taken by the random walk to visit the remaining \( |G| - 1 \) elements of \( G \). The whole problem is right invariant under \( G \), so the distribution of \( C \) does not depend on \( g_0 \). Also by right invariance, with \( T_{g_i g_j} \) the time to hit \( g_i \) from \( g_j \), \( T_{g_i g_j} \) has the same distribution as \( T_{(g_i g_j)^{-1} e} \) the identity. Thus (2.2.2) can be taken as

\[
f^+(s) = \max_{g \neq e} e^{s T_{g e}}
\]

and

\[
(4.1.1) \quad f^-(s) = \inf_{g \neq e} e^{s T_{g e}}.
\]
These hitting times will be calculated using the formulas developed in Section 3.4. Their existence will be guaranteed by the results of Section 3.5. Theorem (2.4.1) will transform $f^+(s)$ and $f^-(s)$ into bounds on the moment generating function of $C$.

Typically $f^+(s)$ and $f^-(s)$ will be very complicated. To streamline calculations, they will often be calculated only to $O(\text{something})$, where the $O(\text{something})$ approximation is as $|C|$ grows. This will be sufficient for asymptotically tight bounds and will greatly reduce the amount of calculations needed.

In the rest of this chapter, three families of examples are considered. First an easy example on $\mathbb{Z}_N$, the discrete circle, is given. Then examples on $\mathbb{Z}_2^N$, the discrete cube, and $\text{Sym}(N)$, the symmetric or permutation group, are considered. Chapter 7 includes a discussion of other possible results.

### 4.2 THE DISCRETE CIRCLE

The discrete circle $\mathbb{Z}_N$ is the familiar Abelian group of the integers \{0,1,...,N-1\} with addition modulo $N$ as "multiplication". There are $N$ irreducible representations of $\mathbb{Z}_N$, all one dimensional. They will be denoted by $\chi_0,...,\chi_{N-1}$ since they coincide with their characters. $k = 0,1,...,N-1$ will be used to index these characters while $j = 0,1,...,N-1$ will index the elements of $\mathbb{Z}_N$. It turns out that

$$x_k(j) = \exp\left(\frac{2\pi i jk}{N}\right); \quad i = \sqrt{-1} \text{ here.}$$

Note that $x_k$ depends implicitly on $N$; it might be preferable to
call it \( \chi_k^N \) because of this. \( \chi_0(j) \equiv 1 \) is (the character of) the trivial representation of \( \mathbb{Z}_N \).

This section considers a family \( \{\mu_N\} \) of measures on \( \{\mathbb{Z}_N\} \). As \( N \to \infty \), the time taken by the random walk on \( \mathbb{Z}_N \) induced by \( \mu_N \) to cover \( \mathbb{Z}_N \) is investigated. Fix \( b^* \), \( 0 < b^* < 1/2 \), and let \( b_N \) satisfy \( N b_N = \lfloor b^* N \rfloor \), where \( \lfloor X \rfloor \) denotes the greatest integer less than or equal to \( X \). Assume \( N \) is large enough that \( b_N \neq 0 \). For ease of notation, \( b_N \) will be denoted simply by \( b \), its dependence on \( N \) understood. Thus in what follows, \( b \) will vary with \( N \), but will always be close to \( b^* \). Let \( \mu_N \) put mass \( (2bN+1)^{-1} \) on each of \( \{N-bN,N-bN+1,\ldots,N-1,0,1,\ldots,bN\} \) (\( N \) is zero mod \( N \)). \( \mu_N \) is symmetric about \( 0 \) and uniformly distributed on the \( 2bN+1 \) points of \( \mathbb{Z}_N \) nearest \( 0 \). \( \mu_N \) induces a symmetric random walk on \( \mathbb{Z}_N \), the lengths of the steps having distribution \( \mu_N \).

Let \( C_N = C_N(b) \) be the time taken by \( \mu_N \)'s random walk to cover \( \mathbb{Z}_N \). To the precision calculated here, the asymptotic distribution of \( C_N \) will not depend on \( B \). To find this asymptotic distribution, maximal and minimal hitting times must be calculated. These, in turn, require the Fourier transform of \( \mu_N \).

\[
\chi_k(\mu_N) = \frac{1}{2bN+1} \sum_{j=-bN}^{bN} \exp\left(\frac{2\pi i j k}{N}\right) = \frac{\cos(2\pi k b)}{2bN+1} + \left(\frac{\sin(2\pi k b)}{2\pi k b}\right)\left(\frac{bN}{2bN+1}\right) \left[\frac{2\pi k \sin\left(\frac{2\pi k}{N}\right)}{N} \left(1-\cos\left(\frac{2\pi k}{N}\right)\right)\right].
\]

A simple bound on \( |\chi_k(\mu_N)| \) is available. First assume \( 1 \leq k \leq N/2 \). Then \( 0 < \frac{2\pi k}{N} \leq \pi \). On \((0,\pi]\) the function \( g(x) = x \sin(x)/(1-\cos(x)) \)
is nonnegative and decreasing. So for \( x \in (0, \pi) \) \( g(x) \leq \lim_{y \to 0} g(y) = 2 \).

So for \( 1 \leq k \leq N/2 \),

\[
|\chi_k(u_N)| \leq \frac{1}{2bN+1} + \left| \frac{\sin(2\pi kb)}{2\pi kb} \right| \frac{2bN}{2bN+1}.
\]

Similarly for \( N/2 < k \leq N-1 \)

\[
|\chi_k(u_N)| \leq \frac{1}{2bN+1} + \left| \frac{\sin(2\pi(N-k)b)}{2\pi(N-k)b} \right| \frac{2bN}{2bN+1}.
\]

Let \( T_{j0} \) be the time taken by \( \mu_N \)'s random walk to hit 0 from \( j \), for \( j \neq 0 \). \( F_{j0}(z) = \mathbb{E} f_{j0}(s) = E[\exp(st) \mathbf{j}] \).

\( Z_N \) is Abelian so each conjugacy class of \( Z_N \) is a single element. Thus \( \mu_N \) is trivially constant on conjugacy classes. Each irreducible representation of \( Z_N \) is one dimensional, so by (3.4.3), subject to verification of its existence, for \( s \in \mathbb{R} \),

\[
(4.2.2) \quad f_j(s) = \sum_{k=0}^{N-1} \frac{\exp(\frac{2\pi i j k}{N})}{\sum_{k=0}^{N-1} \frac{1}{1-e^{-s} \chi_k(u_N)}} = \sum_{k=0}^{N-1} \frac{\chi_k(u_N)}{1-e^{-s} \chi_k(u_N)}.
\]

Next the existence conditions for \( f_j(s) \) given by Theorem (3.5.12) are verified.

Proposition (4.2.3). For any fixed \( s, -\infty < s < 1 \), for \( N \) large enough \( f_j(s/N) \) exists for all \( j \in (1, \ldots, N-1) \).

Proof. The existence theorem (3.5.12) will involve only the denominator of \( f_j(s/N) \), which does not involve \( j \), and hence show the existence of all \( f_j(s/N) \) simultaneously. The eigenvalues \( \lambda_1, \ldots, \lambda_p \) occurring
in Theorem (3.5.12) are simply the characters $\chi_0(\mu_N), \ldots, \chi_{N-1}(\mu_N)$, ignoring multiplicities for the moment.

For $-\infty < s \leq 0$, existence is known, so fix $0 < s < 1$. Then as $N \to \infty$ $z = e^{s/N} \to 1$. For $k \neq 0$, (4.2.1) shows that $|\chi_k(\mu_N)|$ are bounded away from 1, uniformly in $k$ and $N$. Since $e^{s/N} \to 1$, (i), (ii) and (iv) of Theorem (3.5.12) will be satisfied for $N$ large enough. All that remains is (iii). The quantity of interest is

$$l+(l-e^{s/N}) \sum_{k=1}^{N-1} \frac{1}{1-e^{s/N} \chi_k(\mu_N)}.$$  

Expanding $e^{s/N} = 1 + \frac{s}{N} + O(N^{-2})$ and using the fact that the $\chi_k(\mu_N)$ are bounded away from 1, this is

$$l+\left(-\frac{s}{N} + O(N^{-2})\right) \sum_{k=1}^{N-1} \left(\frac{1}{1-\chi_k(\mu_N)} + O(1/N)\right) =

1 - \frac{s}{N} \sum_{k=1}^{N-1} \frac{1}{1-\chi_k(\mu_N)} + O(1/N).$$

Write

$$\sum_{k=1}^{N-1} \frac{1}{1-\chi_k(\mu_N)} = \sum_{k=1}^{N-1} \chi_k(\mu_N) + \frac{\chi_k^2(\mu_N)}{1-\chi_k(\mu_N)} \leq

N - 1 + \sum_{k=1}^{N-1} \chi_k(\mu_N) + \delta \sum_{k=1}^{N-1} |\chi_k^2(\mu_N)|, \text{ for some } \delta > 0.$$  

$$\sum_{k=1}^{N-1} \chi_k(\mu_N) = \sum_{k=1}^{N-1} \mu(j) \chi_k(j) = \sum_j \mu(j) \sum_{k=1}^{N-1} \chi_k(j) \chi_k(0),$$

where 0 is the identity, $\chi_k(0) = 1.$
By (3.2.9)

\[ \sum_{k=0}^{N-1} x_k(j)x_k(0) = \begin{cases} 
N & j=0 \\
0 & \text{otherwise}.
\end{cases} \]

Thus

\[ \sum_{k=1}^{N-1} x_k(j)x_k(0) = \begin{cases} 
N-1 & j=0 \\
-1 & \text{otherwise}.
\end{cases} \]

\[ \sum_{k=1}^{N-1} x_k(\mu_N) = (N-1)\mu(0)+(-1)(1-\mu(0)) = 0(1), \]

since

\[ \mu(0) = (2bN+1)^{-1}. \]

By the Plancherel Theorem (3.2.14)

\[ \frac{1}{N} \sum_{k=0}^{N-1} |x_k(\mu_N)|^2 = \sum_{j=0}^{N-1} |\mu(j)|^2 = \frac{(2bN+1)}{(2bN+1)^2} = 0\left(\frac{1}{N}\right). \]

Thus

\[ \sum_{k=1}^{N-1} |x_k(\mu_N)|^2 = 0(1) \]

as well and (4.2.4) becomes

52
\[ 1 - \frac{s}{N}(N+0(1)) = 1-s+0(\frac{1}{N}) . \]

For any \( s < 1 \), for \( N \) large enough, this will be positive, verifying iii).

Returning to the hitting time (4.2.2),

\[
f_j(\frac{s}{N}) = \frac{(1-e^{-s/N})}{1-s+0(\frac{1}{N})} \sum_{k=0}^{N-1} \frac{\exp\left(\frac{2\pi ik}{N}\right)}{\chi_k(\mu_N)}. \]

For \( s < 1 \), for \( N \) large enough, this moment generating function exists.

An expansion of the numerator of \( f_j(s/N) \), analogous to that of the denominator yields

\[
(4.2.5) \quad (1-e^{-s/N}) \sum_{k=0}^{N-1} \frac{\exp\left(\frac{2\pi ik}{N}\right)}{1-e^{-s/N}\chi_k(\mu_N)} =
\]

\[
1 - \frac{s}{N} \sum_{k=1}^{N-1} \exp\left(\frac{2\pi ik}{N}\right)(1 + \chi_k(\mu_N) + \frac{\chi_k^2(\mu_N)}{1-\chi_k(\mu_N)}) + 0(\frac{1}{N}).
\]

\[
\sum_{k=0}^{N-1} \exp\left(\frac{2\pi ik}{N}\right) = 0 , \quad \sum_{k=1}^{N-1} \exp\left(\frac{2\pi ik}{N}\right)\chi_k(\mu_N)
\]

and

\[
\sum_{k=1}^{N-1} \exp\left(\frac{2\pi ik}{N}\right)\frac{\chi_k^2(\mu_N)}{1-\chi_k(\mu_N)} \text{ are } 0(1), \text{ as before.}
\]

So (4.2.5) is \( 1+0(\frac{1}{N}) \), hence \( f_j(\frac{s}{N}) = \frac{1+0(\frac{1}{N})}{1-s+0(\frac{1}{N})} \). The \( 0(\frac{1}{N}) \) terms in \( f_j(\frac{s}{N}) \) are easily seen to be uniform in \( j \), so both \( f^+(\frac{s}{N}) \) and
\[ f^{-\left(\frac{S}{N}\right)} = \frac{1+O\left(\frac{1}{N}\right)}{1-s+O\left(\frac{1}{N}\right)} \quad \text{for } s < 1, \ N \text{ large enough. Theorem (2.4.1)} \]

shows that

\[ (4.2.6) \quad E \exp\left(\frac{sC_N}{N}\right) = \frac{\Gamma(N+1)\Gamma\left(\frac{1-s+O\left(\frac{1}{N}\right)}{1+O\left(\frac{1}{N}\right)}\right)}{\Gamma(N + \frac{1-s+O\left(\frac{1}{N}\right)}{1+O\left(\frac{1}{N}\right)})}. \]

Proposition (3.6.1) shows that

\[ \frac{\Gamma(N+1)}{\Gamma(N + \frac{1-s+O\left(\frac{1}{N}\right)}{1+O\left(\frac{1}{N}\right)})} = s+O\left(\frac{1}{N}\right) = e^{s \log N} (1+o(1)) . \]

Moving this term to the other side of (4.2.6),

\[ E[\exp\left(\frac{sC_N-N\log N}{N}\right)] = \Gamma\left(\frac{1-s+O\left(\frac{1}{N}\right)}{1+O\left(\frac{1}{N}\right)}\right)(1+o(1)) = \Gamma(1-s)(1+o(1)) \quad \text{for } s < 1 . \]

For these simple families of measures on \( \mathbf{Z}_N \), as \( N \to \infty \) the moment generating function of \( C_N \), properly normalized, converges to \( \Gamma(1-s) \), the moment generating function of the extreme value distribution. This implies that \( \frac{1}{N}(C_N-N\log N) \) converges in distribution to the extreme value distribution and that all moments of \( \frac{1}{N}(C_N-N\log N) \) converge to the corresponding moments of the extreme value distribution. In particular,

\[ EC_N = N\log N + N \Gamma'(1) + o(N), \text{ where } \Gamma'(1) \text{ is Euler's constant,} \]

\[ 0.57721566 \ldots \]. Also \( \text{Var}(C_N) \sim \pi^2 N^2/6(1+o(1)) \) since \( \pi^2/6 \) is the variance of the extreme value distribution. (Gradshteyn and Ryzhik, 1980, 8.366).
4.3 THE DISCRETE CUBE

The discrete cube $\mathbb{Z}_2^N$ is the direct product of $N$ copies of $\mathbb{Z}_2$. It can be thought of as the set of all ordered n-tuples $(a_1, \ldots, a_N)$, where $a_i = 0$ or 1. The group multiplication is componentwise addition modulo two. $\mathbb{Z}_2^N$ has an interpretation as the state space of the microscopic Ehrenfest chain (Kemperman, 1961). In this, $N$ particles can be in either of two states, 0 or 1. A random walk on $\mathbb{Z}_2^N$ then describes some chance mechanism by which these particles change states.

$\mathbb{Z}_2^N$ is Abelian; $|\mathbb{Z}_2^N| = 2^N$. Again $j$ denotes a generic element of $\mathbb{Z}_2^N$. Now $j$ stands for an n-tuple of zeros and ones. Sums indexed by $j$ will be sums over all of $\mathbb{Z}_2^N$. $\mathbb{Z}_2^N$ has $2^N$ irreducible representations, equivalent to their characters since $\mathbb{Z}_2^N$ is Abelian. $k$ will again index the characters; here $k$ stands for an n-tuple of zeros and ones. The character $\chi_k(j)$ is $(-1)^{j \cdot k}$, where $j \cdot k$ is the usual inner product of $j$ and $k$, or more simply, the number of components in which $j$ and $k$ are both one. Sums over $k$ will vary over all $2^N$ irreducible characters; over all $2^N$ possible n-tuples of zeros and ones.

Let $\mu_N^2$ be a measure on $\mathbb{Z}_2^N$, for each $N$, that puts mass $P N$ on $(0, \ldots, 0)$ and mass $(1-P N)/N$ on each of $(1,0,\ldots,0),(0,1,0,\ldots,0),\ldots,(0,\ldots,0,1)$. $P N$ may vary with $N$, but to avoid complications in asymptotics as $N \to \infty$, assume that $\sup P N < 1$. For each step the random walk on $\mathbb{Z}_2^N$ corresponding to $\mu_N$ does not move with probability $P N$, otherwise it

55
changes exactly one coordinate, with each coordinate equally likely to be changed. \( \mu_N \) is invariant under permutations of the coordinates of \( \mathbb{Z}_2^N \). This will simplify computations considerably. Calculations similar to those to be given here could be performed for other invariant measures, though they might be messier. However, \( \mu_N \) is, in some intuitive sense, the most slowly mixing invariant measure that generates \( \mathbb{Z}_2^N \). Hence if \( \mu_N \)'s covering time has an asymptotic extreme value distribution, then covering times for other invariant measures should have this property as well. This will be discussed in Chapter 7.

Again \( C_N \) is the time taken by a random walk generated by \( \mu_N \) to cover \( \mathbb{Z}_2^N \). The first step in investigating \( C_N \) is to find the Fourier transform of \( \mu_N \). Recall that for a character \( \chi_k \), \( k \) stands for an \( n \)-tuple of zeros and ones. Let \( d(k) \) be the number of ones in this \( n \)-tuple. Then

\[
(4.3.1) \quad \chi_k(\mu_N) = \sum_j (-1)^j \cdot k \mu_N(j) = p_N + (1-p_N)(1 - \frac{2d(k)}{N}).
\]

Note that \( \binom{N}{d} \) of the \( 2^N \) characters \( \chi_k \) have \( d(k) = d \), for \( d = 0, 1, \ldots, N \).

As explained in the introduction, only the times to hit the identity \((0, \ldots, 0)\), from all \( j \in \mathbb{Z}_2^N \) need to be considered. Since \( \mu_N \) is invariant under coordinate permutations, the distribution of \( T_j \), the time to hit the identity from \( j \), depends on \( j \) only through \( d(j) \). So, for \( d = 1, 2, \ldots, N \), let \( T_d \) be a random variable whose distribution is the time taken to hit the identity from some \( j \in \mathbb{Z}_2^N \) with \( d(j) = d \). Let \( f_d(s) = E[\exp(s T_d)] \).

To use Theorem (2.4.1) \( f^{-}(s) = \min_d f_d(s) \) and \( f^{+}(s) = \max_d f_d(s) \) must...
be found. This can be done already. If \( d_1 > d_2 \) then \( T_{d_1} \) is stochastically larger than \( T_{d_2} \), because to hit the identity from some state \( j \) with \( d(j) = d_1 \), the random walk must pass through some state \( j' \) with \( d(j') = d_2 \). This is because the random walk changes at most one coordinate at each step. Thus for \( s > 0 \), \( f^-(s) = f_1(s) \) and \( f^+(s) = f_N(s) \).

Similarly for \( s < 0 \), \( f^-(s) = f_N(s) \) and \( f^+(s) = f_1(s) \). To use Theorem (2.4.1) only \( f_1(s) \) and \( f_N(s) \) need to be calculated. It will turn out that Theorem (2.4.1) is slightly too crude for this problem and that a more refined approach, requiring \( f_2(s) \), is needed.

Using (4.3.1) and (3.4.3), \( f_1(s) \), \( f_2(s) \) and \( f_N(s) \) all have the same denominator, namely

\[
(4.3.2) \quad \sum_k (1-e^sX_k(\mu_N))^{-1} = \sum_{d=0}^{N} \left( \begin{array}{c} N \\ d \end{array} \right) [1-e^s(P_N)+(1-P_N)(1-2d/N)]^{-1}.
\]

The numerator of \( f_1(s) \) is

\[
(4.3.3) \quad \sum_{d=0}^{N} \sum_{k:d(k)=d} (-1)^{(k',1,0,\ldots,0)} [1-e^s(P_N)+(1-P_N)(1-2d/N)]^{-1} = \sum_{d=0}^{N} [2^{N-1}_d-(N-d)] [1-e^s(P_N)+(1-P_N)(1-2d/N)]^{-1}.
\]

This is because, of the \( \left( \begin{array}{c} N \\ d \end{array} \right) \) characters with \( d(k) = d \), \( \left( \begin{array}{c} N-1 \\ d \end{array} \right) \) of them have the first coordinate of \( k \) equal to zero. For these, \( (-1)^{(k',1,0,\ldots,0)} = 1 \). For the remaining \( \left( \begin{array}{c} N \\ d \end{array} \right) - \left( \begin{array}{c} N-1 \\ d \end{array} \right) \) characters, the first coordinate of \( k \) is zero, hence \( (-1)^{(k',1,0,\ldots,0)} = -1 \). Similarly the numerators of \( f_2(s) \) and \( f_N(s) \) are, respectively,
\begin{align}
\sum_{d=0}^{N} \left[2^{(N-2)_{d}}+2^{(N-2)_{d-2}}-(N)\right] [1-e^{s}\left[P_{N}+(1-P_{N})(1-\frac{2d}{N})\right]]^{-1}
\end{align}

and

\begin{align}
\sum_{d=0}^{N} [-(1)^{d}(N)_{d} [1-e^{s}\left[P_{N}+(1-P_{N})(1-\frac{2d}{N})\right]]^{-1}.
\end{align}

Next Theorem (3.5.12) is used to give the range of existence of $f_{d}(s)$. Recall $\{P_{N}\}$ are assumed to be bounded away from 1.

Proposition (4.3.6). Given $0 < s < 1$, for all $N$ sufficiently large, $f_{d}(s(l-P_{N})/2^{N})$ exists for all $d = 1, 2, \ldots, N$.

Proof. Let $\chi_{d} = P_{N}+(1-P_{N})(1-\frac{2d}{N})$. If $d(k) = d$, then $\chi_{k}(\mu_{N}) = \chi_{d}$. Thus $\chi_{d}$ is an eigenvalue of multiplicity $(N)_{d}$ of the transition matrix of the random walk under consideration. $\chi_{0} = 1$ and $\chi_{N} = 2P_{N}-1$. $\chi_{N}$ is bounded away from 1 since $P_{N}$ is. For $0 < d < N$, $1-\chi_{d} = 2(1-P_{N})d/N$. Thus $|1-\chi_{d}| > \frac{\epsilon}{N}$, for $0 < d < N$, for $\epsilon = 2 \inf_{N} |1-P_{N}|$. Similarly $|1+\chi_{d}| > \frac{\epsilon}{N}$, for $0 < d < N$.

Take $z$ of Theorem (3.5.12) to be $\exp(s(l-P_{N})2^{-N})$. $|1-1/z| = 0(\frac{1}{2^{N}})$, hence for $N$ large enough, for all $0 < d < N$, $\frac{1}{z} > |\chi_{d}|$. This gives conditions i) and ii). For condition iii), the quantity of interest is $(1-z)$ times (4.3.2), or

\begin{align}
1+[1-\exp(s(l-P_{N})2^{-N})] \sum_{d=1}^{N} \left(N\right)_{d} [1-\exp(s(l-P_{N})2^{-N})\chi_{d}]^{-1}.
\end{align}

Expand $\exp(s(l-P_{N})2^{-N})$ to $1+\frac{s(l-P_{N})}{2^{N}} + O(2^{-2N})$. (4.3.7) becomes

58
\[ 1 - \frac{s(1-P_N)}{2N} + O\left(\frac{N^2}{2N}\right) \sum_{d=1}^{N} \binom{N}{d} \frac{1}{1-\chi_d} + O\left(\frac{N^2}{2N}\right), \]

since \( |1-\chi_d|^{-2} = O(N^2) \) for \( d=1,\ldots,N \). \( 1-\chi_d = (1-P_N)(1-[1-2d]\frac{1}{N}) \), so this becomes

\[ (4.3.8) \quad 1 - s\left[ \sum_{d=1}^{N} \binom{N}{d} \frac{1}{2N} \right] \frac{1}{1-\frac{2d}{N}} + O\left(\frac{N^2}{2N}\right). \]

Note \( \frac{1}{1-\frac{2d}{N}} = 1 + [1-\frac{2d}{N}] + \cdots + [1-\frac{2d}{N}]^5 + \frac{N}{2d} \left[1-\frac{2d}{N}\right]^6. \)

Plugging this into (4.3.8) and using the moments of the Binomial distribution, (4.3.8) becomes

\[ 1-s\left[1-2^{-N} + 1-2^{-N} - \frac{N}{2} \cdot \frac{2}{N} + 4N^{-2}(\frac{N}{4} - \frac{N^2}{2^{N+2}}) + 8N^{-3}(-\frac{N^3}{2^{N+3}}) + 16N^{-4}\left(\frac{3N^2}{16} - \frac{N^3}{8}\right) + 32N^{-5}\left(-\frac{N^5}{2^{N+5}}\right) + 64N^{-6}\cdot\frac{N}{2} \cdot 0(N^{-3}) \right] + O\left(\frac{N^2}{2N}\right). \]

This is

\[ (4.3.9) \quad 1-s\left[1 + \frac{1}{N} + O(N^{-2})\right] \]

which, given \( s < 1 \), is positive for \( N \) large enough, verifying iii). iv) is verified similarly. 

(4.3.9) is \( [1-\exp(s(1-P_N)2^{-N})] \) times the common denominator of \( f_1(s(1-P_N)2^{-N}), \ldots, f_N(s(1-P_N)2^{-N}) \). Thus to find the numerator of \( f_1, f_2, \) and \( f_N \), (4.3.3), (4.3.4), and (4.3.5) must be multiplied by \( 1-\exp(s(1-P_N)2^{-N}) \). For \( f_1 \) this is
\[ (4.3.10) \quad 1 + (1 - \exp(s(1 - P_N)2^{-N})) \sum_{d=1}^{N} \frac{2^{(N-1)} - (N-1)}{d} \frac{1}{1 - \exp(s(1 - P_N)2^{-N})} = \]

\[ 1 - s \left[ 2 \sum_{d=1}^{N-1} \frac{(N-1)2^{-N}}{1 - \left[ 1 - \frac{2d}{N} \right]} - \left( 1 + \frac{1}{N} + 0(N^{-2}) \right) \right], \]

using (4.3.9) and the techniques of its derivation. Expanding

\[ (1 - [1 - \frac{2d}{N}])^{-1} = 1 + (1 - \frac{2d}{N}) + \ldots + (1 - \frac{2d}{N})^5 + \frac{N}{2d} (1 - \frac{2d}{N})^6 \]

again and noting that

\[ \sum_{d=1}^{N-1} \frac{(N-1)2^{-N}}{1 - \frac{2d}{N}} = \frac{2}{N} \sum_{d=1}^{N-1} \frac{N-1}{d} \sum_{n=0}^{P} \left( \frac{N-1}{2} - d \right) \left( \frac{1}{2} \right)^{m-n} \]

(4.3.10) becomes

\[ 1 - s \left( \frac{1}{N} + 0(N^{-2}) \right). \]

A similar computation for (4.3.4) and (4.3.5) yields

\[ (4.3.11) \quad f_1(s(1 - P_N)2^{-N}) = \frac{1 - \frac{s}{N} + 0(N^{-2})}{1 - s(1 + \frac{1}{N}) + 0(N^{-2})} \]

\[ f_2(s(1 - P_N)2^{-N}) = \frac{1 + 0(N^{-2})}{1 - s(1 + \frac{1}{N}) + 0(N^{-2})} \]

\[ f_N(s(1 - P_N)2^{-N}) = \frac{1 + 0(N^{-2})}{1 - s(1 + \frac{1}{N}) + 0(N^{-2})}. \]
Again \( C_N \) is the time taken by the random walk generated by \( \mu_N \) to cover \( \mathbb{Z}_N^2 \). As was mentioned earlier Theorem (2.4.1) is slightly too crude to give an asymptotic distribution for \( C_N \). Attempting to use it results in the following.

For \( s > 0 \) \( f^-(s) = f_1(s) \) and \( f^+(s) = f_N(s) \). Assume \( N \) is large enough that \( f^+ \) and \( f^- \) exist at \( s(1-P_N)2^{-N} \) for some given \( s \), \( 0 < s < 1 \). Theorem (2.4.1) yields

\[
\frac{\Gamma(2^N)\Gamma(1-s+0(1/N))}{\Gamma(2^N-1+\frac{1-s(1+1/N)}{1-s/N}) + O(N^-2)} \leq E[\exp\left(\frac{s(1-P_N)C_N}{2N}\right)]
\]

\[
< \frac{\Gamma(2^N)\Gamma(1-s+0(1/N))}{\Gamma(2^N-1+1+1/N) + O(N^-2)} \cdot \frac{1-s(1+1/N)}{1-s/N} = 1-s^2/N + O(N^-2),
\]

so the upper and lower bounds differ asymptotically by a factor of

\[
\frac{\Gamma(2^N-s-s/N+O(N^-2))}{\Gamma(2^N-s^2/N+O(N^-2))} = 2^{-s}(1+o(1)),
\]

so the bounds will not be tight asymptotically.

It turns out the lower bound, involving \( f_1(s) \) fails for \( s > 0 \). For \( s < 0 \), the upper bound, again involving \( f_1(s) \), fails. To obtain better bounds, return briefly to the notation of Section 2.4. Recall

\[
(4.3.12) \quad E[\exp(sR_i) | F_{i-1}] = 1_{\{R_1 = 0\}} + 1_{\{R_1 \neq 0\}} E[\exp(sT_{X_{R_i} \to s^A_{R_{i-1}}})].
\]

61
In the present context \( A_{\sigma_1} \) is a single element of \( \mathbb{Z}_2^N \), the next one due to be hit in the random order. \( X_{S_{i-1}} \) is the position of the random walk at the first time \( A_{\sigma_1}, \ldots, A_{\sigma_{i-1}} \) have all been visited. \( R_i = 0 \) if and only if \( A_{\sigma_1} \) has already been visited at time \( S_{i-1} \).

Let \( D \in F_{i-1} \) be the event that \( A_{\sigma_1} \) and \( X_{S_{i-1}} \) differ in exactly one coordinate. \( \overline{D} \) is the event that \( A_{\sigma_1} \) and \( X_{S_{i-1}} \) differ in two or more coordinates. \( A_{\sigma_1} \) cannot equal \( X_{S_{i-1}} \), so \( D \cup \overline{D} \) is the whole probability space. Clearly \( D \cap \overline{D} = \emptyset \). (4.3.12) can now be rewritten

\[
E[\exp(sR_i)|F_{i-1}] = 1_{\{R_i = 0\}} + 1_{\{R_i \neq 0\}} E[(1_D + 1_{\overline{D}}) \exp(sX_{S_{i-1}}|A_{\sigma_1})].
\]

Since on \( \overline{D} \) \( X_{S_{i-1}} \) and \( A_{\sigma_1} \) differ in two or more coordinates, on \( \overline{D} \)

\[
E[\exp(sX_{S_{i-1}}|A_{\sigma_1})] \geq f_1(s), \quad \text{for } s > 0. \quad \text{Let } s^* = s(1-P_N)2^{-N}.
\]

Then for \( 1 > s > 0 \),

\[
E[\exp(s^*R_i)|F_{i-1}] \geq 1_{\{R_i = 0\}} + 1_{\{R_i \neq 0\}} [1_D f_1(s^*) + 1_{\overline{D}} f_2(s^*)].
\]

From (4.3.11), for \( 0 < s < 1 \), for \( N \) large enough there is a constant \( b \) such that \( f_1(s^*) > f_2(s^*) - \frac{b}{N} \). Thus

\[
E[\exp(s^*R_i)|F_{i-1}] \geq 1_{\{R_i = 0\}}(1-f_2(s^*)+f_2(s^*)-1_{\{R_i \neq 0\}}(1-D N b).
\]

Recall that

\[
E[\exp(s^*S_{i-1})] = E[\exp(s^*S_{i-1})E[\exp(s^*R_i)|F_{i-1}]],
\]

62
so multiplying the above by \( \exp(s^* S_{i-1}) \) and taking expectations,

\[
E[\exp(s^* S_i)] = E[\exp(s^* S_i)](1-\frac{1}{i}) + f_2(s^*)E[\exp(s^* S_{i-1})] - \frac{b}{N} E[\exp(s^* S_{i-1})1_{\{R_i \neq 0\}}] .
\]

Rearranging, and using the fact that \( f_2(s^*)(1-\frac{1}{i}) + \frac{1}{i} \geq 1 \) for \( s^* > 0 \), for the same constant \( b \) above, for all \( i \) and for \( N \) large enough,

\[
(4.3.13) \quad E[\exp(s^* S_i)] = \frac{1}{i-1+\frac{1}{i}} E[\exp(s^* S_{i-1})] - \frac{b}{N} E[\exp(s^* S_{i-1})1_{\{R_i \neq 0\}}] .
\]

The last term must be shown to be negligible. \( P(R_i \neq 0) = \frac{1}{i} \), so conditioning on this the last term is

\[
\frac{b}{Ni} E[\exp(s^* S_{i-1})] \quad 1_{D|R_i \neq 0} .
\]

Now apply Hölders inequality with \( p = \frac{1+s}{2s} \), \( q = \frac{1+s}{1-s} \) to bound the above by

\[
(4.3.14) \quad \frac{b}{Ni} E[\exp((\frac{s+1}{2}) S_{i-1})1_{R_i \neq 0}] \quad \frac{2s}{1+s} P(D|R_i \neq 0) \quad \frac{1-s}{1+s}
\]

where \( (\frac{s+1}{2}) = (\frac{s+1}{2})(1-P_N)^{-N} \). Since \( \frac{s+1}{2} < 1 \), for \( N \) large enough this will exist by Proposition (4.3.6). Next consider the following proposition.
Proposition (4.3.15). For \( N \) and \( s > 0 \) such that the quantities under consideration exist and under (2.2.3),

\[
E[\exp(sS_{i-1})|R_{i} \neq 0] \leq E[\exp(sS_{i})].
\]

Proof. By Lemma (2.2.7)

\[
E[\exp(sS_{i})] = E[\exp(sS_{i})|R_{i} \neq 0].
\]

\( S_{i-1} \leq S_{i} \), so

\[
E[\exp(sS_{i-1})|R_{i} \neq 0] \leq E[\exp(sS_{i})|R_{i} \neq 0].
\]

(4.3.15) follows.

So (4.3.14) is less than

\[
(4.3.15) \quad \frac{b}{N_{1}} E[\exp((\frac{s+1}{2}) S_{i})]^{\frac{2s}{1+s}} P(D|R_{i} \neq 0)^{\frac{1-s}{1+s}}.
\]

As in the proof of Theorem (2.4.1),

\[
E[\exp((\frac{s+1}{2}) S_{i})]^{\frac{2s}{1+s}} \leq \left[ \frac{\Gamma(i+1)(1/f^{+}(\frac{s+1}{2}))^{\frac{2s}{1+s}}}{\Gamma(i+1)(\frac{s+1}{2})} \right].
\]

64
\[
\frac{1}{f_N(s^*)} = \frac{1}{\Gamma^+(s^*)} = 1 - \frac{s+1}{2} + O(\frac{1}{N}) \quad \text{by (4.3.11). Applying Proposition (3.6.1), this is}
\]
\[
\frac{s+1}{i^2 \Gamma(1 - \frac{s+1}{2})(1 + o(1))}^{1+s} \leq d_i^s
\]
for some constant \(d\), independent of \(i\), again for \(N\) large enough. But similarly \(E[\exp(s^*S_i)] \geq d' i^s\) for some \(d'\), independent of \(i\), for \(N\) large enough. Thus, changing the constant \(b\) if necessary, (4.3.15) is less than
\[
E[\exp(s^*S_i)] P(D | R_i \neq 0) \left(1 - \frac{s}{1+s}\right).
\]

Finally consider \(P(D | R_i \neq 0) = P(D \cap \{R_i \neq 0\}) / P(R_i \neq 0) \leq P(D) / P(R_i \neq 0)\).\(D\) is the event that \(X_{S_{i-1}}^c\) and \(A_{\sigma_i}\) differ in only one coordinate. \(A_{\sigma_i}\) is equally likely to be any of the \(2^{N-1}-(i-1)\) states not chosen. At most \(N\) of them differ from \(X_{S_{i-1}}^c\) in only one coordinate, so \(P(D) < N/(2^{N-1})\). \(P(R_i \neq 0) = \frac{1}{i!}\), so (4.3.16) is less than
\[
\frac{b}{N_i} E[\exp(s^*S_i)] \left(\frac{N_i}{2^{N-1}}\right)^{1+s}.
\]
Putting this into (4.3.13) and rearranging,
\[
E[\exp(s^*S_i)] \geq \frac{1}{i^2 \Gamma^+(s^*)} E[\exp(s^*S_{i-1})] \left[1 + \frac{\frac{b}{N_i}(\frac{N_i}{2^{N-1}})}{1+s}\right]^{-1}.
\]
This bound is good for \(i\) much smaller than \(2^N\), but bad for \(i\) near \(2^N\). For \(i\) large, it is enough bound \(P(D | R_i \neq 0)\) by \(1\) in (4.3.16).
and obtain

\[(4.3.19) \quad E[\exp(s^* S_i^1)] \geq \frac{1}{i-1 + \frac{1}{f_2(s^*)}} E[\exp(s^* S_{i-1})](1 + \frac{b}{N_l})^{-1} . \]

It turns out that using (4.3.18) for \( i < 2^N(1-1/\log N) \) and (4.3.19) for \( i > 2^N(1-1/\log N) \) works. Combining these,

\[(4.3.20) \quad E[\exp(s^* C_N)] \geq \frac{\Gamma(2^N)\Gamma(\frac{1}{f_2(s^*)})}{\Gamma(2^N-1+\frac{1}{f_2(s^*)})} . \]

\[
2^N(1-1/\log N) \prod_{i=1}^{2^N} \left[ 1 + \frac{b}{Ni} \left( \frac{Ni}{2^N-i} \right)^{1+s} \right]^{-1} \prod_{i=2^N(1-1/\log N)}^{2^N} \left[ 1 + \frac{b}{Ni} \left( \frac{Ni}{2^N-i} \right)^{1+s} \right]^{-1} . \]

The last product is

\[
\exp \left[ - \sum_{i=2^N(1-1/\log N)}^{2^N} \log \left( 1 + \frac{b}{Ni} \right) \right] = \exp \left[ - \sum_{i=1}^{2^N} \left( \frac{b}{iN} + O \left( \frac{1}{i^2} \right) \right) \right] = \exp \left[ - \frac{b}{N} \left( \log 2^N - \log 2^N(1-1/\log N) + o(1) \right) \right] ,
\]

which converges to 1 as \( N \to \infty \). For the middle product, choose any \( \epsilon > 0 \).

\[
\left( \frac{Ni}{2^N-i} \right)^{1+s} < \epsilon \quad \text{if} \quad i < \frac{2^N \epsilon^{1-s}}{1+s} .
\]
But for \( N \) large enough,

\[
\frac{1+s}{2^N e^{1-s} \left( \frac{1+s}{N+e} \right)^{1-s}} > 2^{N(1-1/\log N)}.
\]

For such \( N \), the second product in (4.3.20) is bigger than

\[
\frac{2^N}{\prod_{i=1}^{N} \left( 1 + \frac{b \epsilon}{N_i} \right)}.
\]

This is

\[
\exp\left[ - \sum_{i=1}^{2^N} \left( \frac{b \epsilon}{N_i} + 0\left( \frac{1}{N_i^2} \right) \right) \right] = \exp(-b \epsilon \log 2 + o(1)).
\]

Since \( b \) is fixed and \( \epsilon \) is arbitrarily small, this product converges to \( 1 \) as \( N \to \infty \).

Combining (4.3.20), the upper bound from Theorem (2.4.1) obtained previously, and the generating functions (4.3.11),

\[
E[\exp\left( -\frac{s(1-P_N)}{2^N C_N} \right)] = \frac{\Gamma(2^N) \Gamma(1-s+0(\frac{1}{N}))}{\Gamma(2^N -1 +1-s(1 + \frac{1}{N}) + 0(\frac{1}{N^2}))} (1+o(1))
\]

for \( 0 < s < 1 \), for \( N \) large enough. By Proposition (3.6.1), as \( N \to \infty \) the right hand side is

\[
\Gamma(1-s)(2^N)^{s(1+1/N)+0(1/N^2)} (1+o(1)) = \Gamma(1-s)2^{(N+1)s} (1+o(1)).
\]

Dividing by \( 2^{(N+1)s} \),
\[ (4.3.21) \quad \mathbb{E}[\exp\left(\frac{s(1-P_N)}{2^N} \left(C_N - 2N \log 2^{N+1}\right)\right)] = \Gamma(1-s)(1+o(1)) \]

as \( N \to \infty \) for \( 0 < s < 1 \).

For \( s < 0 \), the upper bound, again involving \( f_1(s) \), fails in Theorem (2.4.1). Proposition (4.3.15) is not true for \( s < 0 \) so a slightly different argument is needed. This new argument works only for \( -1 < s < 0 \). With \( s^* \) again equal to \( s(1-P_N)2^{-N} \), an argument just like the one leading to (4.3.13) yields

\[
\mathbb{E}[\exp(s^*(S_1))(f_2(s^*)(1-\frac{1}{i}) + \frac{1}{i})] \leq f_2(s^*)E[\exp(s^*S_{i-1})] \\
+ \frac{b}{N} E[\exp(s^*S_{i-1})1_{\{R_i \neq 0\}}1_D]
\]

for \( s < 0 \), some constant \( b > 0 \), and \( N \) large enough. For fixed \( s \), (4.3.11) guarantees that \( f_2(s^*) \) is bounded from below as \( N \to \infty \), thus

\[ (4.3.22) \quad \mathbb{E}[\exp(s^*S_1)] \leq \frac{1}{i-1-\frac{1}{f_2(s^*)}} \mathbb{E}[\exp(s^*S_{i-1})] \\
+ \frac{b}{N} E[\exp(s^*S_{i-1})1_{\{R_i \neq 0\}}1_D]
\]

for some new constant \( b \). Again, the last term must be shown to be negligible.

For \( -1 < s < 0 \), let \( q = \frac{s-1}{2s} \), \( p = r = 2(1-s)/(1+s) \). \( \frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1 \), so

writing the last term of (4.3.22) as

\[
\frac{b}{N} E[\exp(s^*S_1)\exp(-s^*R_i)1_D1_{R_i \neq 0}] ,
\]

68
Hölder's inequality bounds this term by

\[ (4.3.23) \quad \frac{b}{N} \left[ E[\exp\left(\frac{2s(1-s)}{1+s}\right)^*_{S} L_{1(\{R_1 \neq 0\})}] \right]^{\frac{1+s}{2(1-s)}} \cdot \]

\[ E[\exp\left(-\frac{(s-1)^*}{2} R_1 \right)^*_{1(\{R_1 \neq 0\})}]^{\frac{2s}{s-1}} P(D_0(\{R_1 \neq 0\})^{\frac{1+s}{2(1-s)}}] \]

For \( s > -1 \), the middle term exists for \( N \) large enough, since \( -(s-1)/2 < 1 \). Conditioning on \( \{R_1 \neq 0\} \), (4.3.11) guarantees the middle term is \( 0(1^{2s}) \). Just as in the case of \( s > 0 \) (4.3.23) is seen to be

\[ \leq \frac{b}{N} E[\exp(s^* S_i) P(D|R_1 \neq 0)]^{\frac{1+s}{2(1-s)}} \]

and the remainder of the argument is identical.

Thus, it has been shown that for \(-1 < s < 1\), as \( N \to \infty \)

\[ E[\exp(s(1-P_{N})_{N} N N(2^{N} \log N^{2N+1}))] \to \Gamma(1-s) \cdot \]

As for the discrete circle, this convergence of moment generating functions in an interval containing the origin implies convergence in distribution and of all moments to the corresponding quantities of the extreme value distribution.

4.4 THE SYMMETRIC GROUP

Let \( \text{Sym}(N) \) denote the symmetric, or permutation, group on \( N \) letters. \( \text{Sym}(N) \) has order \( N! \) and is not Abelian for \( N > 2 \). Multiplication in \( \text{Sym}(N) \) is simply composition of elements.
The conjugacy classes of \( \text{Sym}(N) \) are best described through cycles. Suppose \( \text{Sym}(N) \) acts on \( \{1, \ldots, N\} \) and \( a_1, \ldots, a_i \) are \( i \) distinct elements of \( \{1, \ldots, N\} \) for some \( i, \ 1 < i < n \). The permutation \((a_1, \ldots, a_i)\) that maps \( a_1 \rightarrow a_2, \ a_2 \rightarrow a_3, \ldots, a_{i-1} \rightarrow a_i, \ a_i \rightarrow a_1 \) and leaves everything else unchanged is called an \( i \)-cycle. Every permutation can be written uniquely as a product of disjoint cycles (having no elements in common). Being disjoint, these cycles commute. Thus the uniqueness is only in terms of the cycles in the product, not their order. For example \( \text{Sym}(3) \) is the set \( \{(1)(2)(3), (12)(3), (13)(2), (23)(1), (123), (132)\} \). Often \( 1 \)-cycles are not written, thus \( \text{Sym}(3) = \{e, (12), (13), (23), (123), (132)\} \). Multiplication is composition, for example in \( \text{Sym}(3) \) \((12)(23) = (123)\) since

\[
\begin{pmatrix}
a \\
b \\
c
\end{pmatrix}
= \begin{pmatrix}
a \\
c \\
b
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
a \\
c \\
b
\end{pmatrix}
= \begin{pmatrix}
c \\
a \\
b
\end{pmatrix} .
\]

For any permutation \( \pi \) in \( \text{Sym}(N) \), its decomposition into disjoint cycles has \( d_1(\pi) \) \( 1 \)-cycles, \( d_2(\pi) \) \( 2 \)-cycles, \ldots, \( d_N(\pi) \) \( N \)-cycles, where

\[
\sum_{i=1}^{N} i d_i(\pi) = N .
\]

\( \pi \) and \( \sigma \) in \( \text{Sym}(N) \) are conjugate if and only if they have the same cycle structure, i.e., if \( d_i(\pi) = d_i(\sigma) \) for \( i = 1, 2, \ldots, N \). From this it follows that \( \pi \) and \( \pi^{-1} \) are conjugate. This gives a convenient enumeration of the conjugacy classes of \( \text{Sym}(N) \).

A permutation can also be written as a product of \( 2 \)-cycles. For example, the \( 3 \)-cycle \((123)\) is \((12)(23)\). This decomposition is not unique,
but one facet of it is. For any permutation $\pi$, decompositions of $\pi$ into 2-cycles will either always have an even number of terms or always have an odd number of terms. Call $\pi$ even if it can be written as a product of an even number of two cycles, otherwise call $\pi$ odd. A conjugacy class will contain either all even or all odd permutations, so conjugacy classes can be called even or odd in the same manner.

Consider probability measures on $\text{Sym}(N)$ of the following special form. Fix $N_0$ and let $L_1, \ldots, L_M$ be conjugacy classes of $\text{Sym}(N_0)$. Choose $m_1, \ldots, m_M$, $m_i > 0$, $\sum_{i=1}^M m_i = 1$. Let the measure $\mu_{N_0}$ put total mass $m_i$ on class $L_i$, uniformly distributed among the members of $L_i$. $\mu_{N_0}$ is a measure on $\text{Sym}(N_0)$ that is constant on conjugacy classes. Since $\pi \in \text{Sym}(N_0)$ and $\pi^{-1}$ are conjugate, $\mu_{N_0}$ is symmetric as well.

For $N > N_0$, there is a measure on $\text{Sym}(N)$ much like $\mu_{N_0}$. Suppose $L_i$, a conjugacy class of $\text{Sym}(N_0)$, has $d_2(L_i)$ 2-cycles, $d_3(L_i)$ 3-cycles, ..., $d_{N_0}(L_i)$ $N_0$-cycles. Let $L_i^*$ be the conjugacy class of $\text{Sym}(N)$ that also has $d_2(L_i)$ 2-cycles, ..., $d_{N_0}(L_i)$ $N_0$-cycles and no longer cycles. Let $\nu_N$ be the measure on $\text{Sym}(N)$ that puts mass $m_i$ on $L_i^*$, uniformly distributed among the members of $L_i^*$. $\nu_N$ is also constant on conjugacy classes and symmetric. Thus $\{\nu_N\}$, for $N \geq N_0$, form a family of measures on $\{\text{Sym}(N)\}, N \geq N_0$ that are constant on "the same" conjugacy classes. For example if $N_0 = 2$, $M = 1$, $L_1 = \{(1,2)\}$, then $\nu_N$ is the measure on $\text{Sym}(N)$ putting mass $\frac{N}{2}$ on each 2-cycle. Assume further that $\mu_{N_0}$ puts positive mass $q > 0$ on the odd conjugacy classes. This will assure that the random walk generated by each $\nu_N$ can reach every state of $\text{Sym}(N)$. 

71
Let $C_N$ be the time taken by the random walk associated with $\mu_N$ to cover $\text{Sym}(N)$. Asymptotically tight bounds on the moment generating function of $C_N$ will be given.

Much of the analysis necessary for this problem has already been done. In Random Shuffles and Group Representations (1985), Flatto, Odlyzko and Wales study hitting times on $\text{Sym}(N)$. They derive (3.4.2) and (3.4.3) in a different manner and use them to give asymptotics for the Laplace transforms of hitting times of random walks whose step distributions are constant on the conjugacy classes of $\text{Sym}(N)$. Their Laplace transforms are the moment generating functions of interest here for $s < 0$. The generating functions for $0 < s < 1$ are easily obtained from their results and Section 3.5. Much of their paper depends on the asymptotic estimates of Wasserman (1981) for the characters of the symmetric group.

Distinguish two representations of $\text{Sym}(N)$, the trivial and alternating representations, denoted $\text{Triv}$ and $\text{Alt}.\text{Triv}(\pi) = 1$ for all $\pi \in \text{Sym}(N)$. $\text{Alt}(\pi) = 1$, if $\pi$ is an even permutation and $-1$ if $\pi$ is odd. This is easily seen to be a representation. For any other irreducible representations $\rho$ of $\text{Sym}(N)$, let $\chi_\rho$ be its character. For any conjugacy class $L$ of $\text{Sym}(N)$, $\chi_\rho(L)$ is the value of $\chi_\rho$ on $L$. With $d_\rho$ the dimension of $\rho$, define $r_\rho(L) = \chi_\rho(L)/d_\rho$. Flatto, Odlyzko and Wales prove the following proposition.

Proposition (4.4.1). If $L \neq \{e\}$, then as $N \to \infty$

$$\sum_{\rho \neq \text{Triv,Alt}} \chi_\rho^2(L) \frac{|r_\rho(L)|}{1-|r_\rho(L)|} = O\left(\frac{N!}{|L|}\right),$$
where \(|L|\) is the number of elements of \(L\). As in the construction of
the measures \(\mu^*_N\), \(L\) is considered to be a conjugacy class of all
\(\text{Sym}(N)\), for \(N\) bigger than some \(N_0\).

Next, consider how large the conjugacy classes of \(\text{Sym}(N)\) are.
The class \(\{e\}\) always contains exactly one member. Any other class \(L\)
contains at least one \(d\)-cycle, for some \(d\) between 2 and \(N\), in its
presentation as a product of disjoint cycles. If \(L\) contains \(k\) \(d\)-cycles
in this presentation, by counting conjugates, \(L\) has at least
\[
\frac{N![(d-1)!]^k}{(d!)^k(N-kd)!}
\]
elements. This is minimized by taking \(k=1\) and \(d=2\).
Thus any conjugacy class of \(\text{Sym}(N)\), except \(\{e\}\), has at least \(N(N-1)/2\)
elements.

Now suppose \(\{\mu^*_N\}\) is a family of measures on \(\{\text{Sym}(N)\}\), as described
earlier, except that \(\mu^*_N(e) = 0\) for all \(N\). Thus \(\mu^*_N\) is constant on
conjugacy classes and concentrated on a fixed set of conjugacy classes.
Flatto, Odlyzko and Wales assert, and it follows easily from their proofs,
that Proposition (4.4.1) extends to measures \(\mu^*_N\) as described above.
Precisely,

Proposition (4.4.2). For \(\mu^*_N\) as above,

\[
\sum_{\rho \neq \text{Triv, Alt}} 2 \chi_\rho(\mu^*_N) \frac{|r_\rho(\mu^*_N)|}{1-|r_\rho(\mu^*_N)|} = O\left(\frac{N!}{N^2}\right) \quad \text{as} \ N \to \infty.
\]

\(\chi_\rho(\mu^*_N)\) has been defined; \(r_\rho(\mu^*_N)\) is \(\chi_\rho(\mu^*_N)/d_\rho\). The \(O(N!N^{-2})\) term
follows from the fact that the conjugacy classes under consideration have
at least \(O(N^2)\) elements.
Flatto, Odlyzko and Wales also prove the following.

**Proposition (4.4.3).** If \( L \neq \{e\} \) is a conjugacy class, \( \rho \neq \text{Triv,Alt} \) an irreducible representation of \( \text{Sym}(N) \), then there exists \( \theta = \theta(L) > 0 \) such that for all \( N \geq 5 \)

\[
|r_\rho(L)| \leq 1 - \frac{\theta}{N}.
\]

Note that the left side of this equality depends on \( N \), though this dependence is not explicit. This easily extends to measures of the form \( \mu_N^* \).

**Proposition (4.4.4).** If \( \mu_N^* \) is as described above, then there exist constants \( \theta_1 > 0 \), \( \theta_2 > 0 \) such that for \( \rho \neq \text{Triv,Alt} \),

\[
|r_\rho(\mu_N^*)| \leq 1 - \frac{\theta_1}{N^2}.
\]

Now suppose \( \mu_N \) is as originally described. \( \mu_N \) can be represented \( \mu_N = p^\delta_e + (1-p)\mu_N^* \), where \( 0 \leq p < 1 \), \( \delta_e \) is a point mass at the identity, and \( \mu_N^* \) is as recently discussed. Recall that \( \mu_N \) is required to put mass \( q > 0 \) on the odd conjugacy classes of \( \text{Sym}(N) \). Thus \( \text{ALT}(\mu_N) = 1-2q \) is the alternating representation, or character, of \( \mu_N \). For any irreducible representation \( \rho \), \( r_\rho(\mu_N) = p+(1-p)r_\rho(\mu_N^*) \).

Let \( T_{ge} \) be the time taken by a random walk generated by \( \mu_N \) to hit the identity \( e \) from \( g \in \text{Sym}(N) \), \( g \neq e \). Subject to verification of its existence, from (3.4.3), \( T_{ge} \) has moment generating function
(4.4.5) \[ f_g(s) = \frac{\sum_{\rho} d_\rho (1 - e^{s \rho}(\mu_N))^{-1}}{\sum_{\rho} d_\rho^2 (1 - e^{s \rho}(\mu_N))^{-1}} \chi_\rho(g). \]

Note that this depends on \( g \) only through the \( \chi_\rho(g) \), hence \( f_g(s) \) depends on \( g \) only through its conjugacy class. Theorem (3.5.12) now gives the range of existence of \( f_g(s) \).

Proposition (4.4.6). Given \( s < 1, p = \mu_N(e), s^* = s(1-p(N!))^{-1} \), for all \( g \neq e \) and all sufficiently large \( N \), \( f_g(s^*) \) exists.

Proof. \( \mu_N \) is constant on conjugacy classes, hence symmetric, so Theorem (3.5.12) applies. The eigenvalues of interest are as follows. TRIV(\( \mu_N \))=1. ALT(\( \mu_N \))=1-2q, which is bounded away from 1 but may be -1. The remaining eigenvalues are the \( r_\rho(\mu_N) \) for \( \rho \neq \text{TRIV,ALT} \), each occurring with multiplicity \( d_\rho^2 \). Proposition (4.4.4) shows that these \( N!-2 \) eigenvalues cannot approach \( \pm 1 \) very rapidly. It implies that if \( \rho \neq \text{TRIV,ALT} \), then there are constants \( \theta > 0, \theta_2 > 0 \), independent of \( \rho \), such that

\[ |1 - |r_\rho(\mu_N)|| > \frac{\theta}{\theta_2 N^2}. \]

Thus these remaining eigenvalues can approach \( \pm 1 \) at a rate that is uniformly at most a power of \( N^{-1} \). Thus \( z \) of Theorem (3.5.12) is \( \exp(-s(1-p)(N!)^{-1}) \) which for any \( s < 1 \) approaches one at a much faster rate than any power of \( N^{-1} \). Thus for \( N \) large enough, conditions i) and ii) will be satisfied.

Condition iii) involves showing
\[ 1 + (1 - e^{s^*}) \sum_{\rho \neq \text{Triv}} d_\rho^2 (1 - e^{s^*} r_\rho (\mu_N))^{-1} > 0 \]

for \( N \) large enough. Expanding \( e^{s^*} \) and \( 1 - e^{s^*} r_\rho (\mu_N) \), this is equivalent to

\[ (4.4.7) \quad 1 - \frac{s (1 - p)}{N!} \sum_{\rho \neq \text{Triv}} d_\rho^2 \frac{1}{1 - [p+(1-p)r_\rho (\mu_N^*)]} + O \left( \frac{N^2}{N!} \right) > 0. \]

As before,

\[ \mu_N^*(g) = \begin{cases} \mu_N(g)/(1-p) & g \neq e \\ 0 & g = e \end{cases}. \]

The \( 1-p \) term can be removed from the sum and cancelled. Then

\[ [1 - r_\rho (\mu_N^*)]^{-1} \]

can be expanded to \( 1 + r_\rho (\mu_N^*) + r_\rho^2 (\mu_N^*) + r_\rho^3 (\mu_N^*) (1 - r_\rho (\mu_N^*))^{-1} \), as in the paper of Flatto, Odlyzko and Wales. Proposition (4.4.2) can be used on the only difficult term of this expansion to show that (4.4.7) is \( 1-s+O(N^{-2}) \). This verifies condition iii). If \( q \neq -1 \) condition iv) is trivial; otherwise it can be verified like condition iii).

Proposition (4.4.6) in effect follows from a small calculation and the results of Flatto, Odlyzko and Wales. In a similar manner, the following is an easy consequence of their results:

Proposition (4.4.8). With \( \mu_N, p, f_\rho \) as described earlier, for any \( s < 1 \), uniformly in \( g \in \text{Sym}(N) \setminus \{e\} \),
\[ f_g(s(l-p)(N!)^{-1}) = \frac{1+o(N^{-2})}{1-s+o(N^{-2})}. \]

Recall \( C_N \) is the time taken by the random walk generated by \( \mu_N \) to cover \( \text{Sym}(N) \). In Theorem (2.4.1), \( f^+(s(l-p)(N!)^{-1}) \) and \( f^-(s(l-p)(N!)^{-1}) \) are both \( (1+o(N^{-2}))/ (1-s+o(N^{-2})) \), hence

\[
E[\exp\left(\frac{s(l-p)C_N}{N!}\right)] = \frac{\Gamma(N!)\Gamma(1-s+o(N^{-2}))}{\Gamma(N!-s+o(N^{-2}))}.
\]

\( \log(N!) = O(N\log N) \), so Proposition (3.6.1) shows the right side is

\[
\Gamma(1-s(\exp[(\log N!)(-s+o(N^{-2}))])(1+o(1)) = \\
\Gamma(1-s)\exp(-s\log N!)(1+o(1)).
\]

Thus, rearranging terms, as \( N \to \infty \), for \( s < 1 \)

\[
E[\exp\left(\frac{s(l-p)}{N!}(C_N-N!\log(N!))\right)] \to \Gamma(1-s).
\]

So, as in the other cases, the moments and distribution of \( C_N \) converge to those of the extreme value distribution.
CHAPTER 5
PRELIMINARIES FOR SPHERE PROBLEMS

5.1 INTRODUCTION

This chapter and the next consider bounds on the mean time taken by a random walk or Brownian motion on $S_{p-1}$ (the unit sphere in $\mathbb{R}^p$) to come within a distance $\varepsilon$ of all points of $S_{p-1}$. For a random walk, this can be thought of as placing spherical caps of radius $\varepsilon$ onto the points of $S_{p-1}$ visited by the random walk and considering the time taken by this process of caps to cover $S_{p-1}$. There is a similar reformulation for Brownian motion. Sometimes a point of $S_{p-1}$ will be identified with its opposite point, its reflection in the origin. For this problem, two caps will be placed on $S_{p-1}$ for each step of a random walk. One cap will be centered at the point visited, the other at its opposite. Again, interest is in the time taken to cover $S_{p-1}$. $S_{p-1}$, with opposite points thus identified, is isomorphic to the space of one dimensional subspaces of $\mathbb{R}^p$. Section 5.2 on the Grand Tour will motivate the consideration of this problem.

The remainder of this chapter will be as follows. Section 5.3 will mention some published results on packing caps onto spheres. Unfortunately these do not yield what is needed here, so a crude construction is also given. Section 5.4 formally introduces random walks on spheres and Brownian motions as their limit. In Section 5.5 mean hitting times for Brownian motions are calculated. Section 5.6 discusses the convergence of hitting times for random walks to those for Brownian motion.
Chapter 6 will tie this all together and state some results. Further possible results will be discussed in Chapter 7.

5.2 THE GRAND TOUR

Statisticians are often called upon to analyze data sets containing many variables. A good first step is always to plot the data. Thinking of a data set with \( p \) variables as a point cloud in \( \mathbb{R}^p \), this is easy for \( p \leq 2 \), difficult for \( p = 3 \) and impossible for \( p > 4 \). Many techniques have been proposed as substitute methods for "looking at" high dimensional data. The Grand Tour is one of these.

One and two dimensional projections of a data set are easy to look at. The Grand Tour, as its name suggests, is the idea of looking at a sequence of these projections with the hope of somehow gaining an overall understanding of the structure of the data set. Formally, let \( G_{1p} \) and \( G_{2p} \) be, respectively, the spaces of one and two dimensional subspaces of \( \mathbb{R}^p \). A Grand Tour is then a sequence of elements of \( G_{1p} \) and \( G_{2p} \).

Asimov (1983) discusses Grand Tours and gives criteria for a sequence of points of \( G_{1p} \) and \( G_{2p} \) to constitute a good Grand Tour. Several of Asimov's criteria are

1) The sequence should have a sense of continuity, in that adjacent points in the sequence should be close together in \( G_{1p} \) or \( G_{2p} \).
2) The sequence should be smooth in the sense that its points lie nearly along a geodesic.
3) The sequence should become dense in \( G_{1p} \) or \( G_{2p} \) as rapidly and uniformly as possible.
The criterion of becoming dense rapidly can be rephrased as follows. Suppose \((p_1, p_2, \ldots)\) is a sequence of projections constituting a Grand Tour. Given \(\varepsilon > 0\), put a cap \(p_i(\varepsilon)\) about each point \(p_i\). Let \(C_\varepsilon\) be the first integer \(n\) such that the caps \(p_1(\varepsilon), \ldots, p_n(\varepsilon)\) cover \(G_{1p}\) or \(G_{2p}\). The criterion becomes the idea that \(C_\varepsilon\) should be as small as possible. If the sequence \((p_1, p_2, \ldots)\) is nondeterministic, for example if it comes from a random walk on \(G_{1p}\) or \(G_{2p}\), then \(C_\varepsilon\) is random, and it is reasonable to try to make \(E(C_\varepsilon)\) small.

Asimov gives several suggestions for generating a Grand Tour. If the Grand Tour is required to be of a definite length, then it may be possible to select a sequence of projections that are as dense and smooth as possible. In practice, this would be difficult to optimize. If a more informal Grand Tour, of no definite length, is desired, then there are several easy ways to generate reasonable sequences of projections.

The first is what Asimov calls the Torus method. The special orthogonal group \(SO_p\) acts on \(G_{1p}\) and \(G_{2p}\) in an obvious fashion. Choose a one parameter subgroup of \(SO_p\) that becomes dense in \(SO_p\), and select a sequence of members of the subgroup. This sequence, acting on a point of \(G_{1p}\) or \(G_{2p}\), will produce a sequence in \(G_{1p}\) or \(G_{2p}\) that becomes dense asymptotically. Using this method, it is easy to find a sequence that is good by criteria 1 and 2. There is of course a conflict in trying to satisfy criteria 1 and 3 simultaneously. The rate at which a sequence of this form becomes dense is not understood. Asimov (1983) gives some simulation results.
A second technique is to choose the projections in the sequence independently and uniformly in $G_{1p}$ and $G_{2p}$. This procedure does terribly on criteria 1 and 2, but its time to become dense is comparatively well understood. Let $V_\varepsilon$ be the volume of a cap of radius $\varepsilon$ in $G_{1p}$ or $G_{2p}$, normalized so the whole space has volume 1. $C_\varepsilon$, the time caps of radius $\varepsilon$ take to cover $G_{1p}$ or $G_{2p}$, is random. As $\varepsilon \to 0$ $E(C_\varepsilon) \sim V_\varepsilon^{-1}\log(V_\varepsilon^{-1})$ and $\text{Var}(C_\varepsilon) \sim V_\varepsilon^{-2}$. Also, properly normalized, $C_\varepsilon$ has an asymptotic extreme value distribution. Flatto and Newman (1977) give results for $E(C_\varepsilon)$, while Janson (1984) contains $\text{Var}(C_\varepsilon)$ and the asymptotic distribution of $C_\varepsilon$.

A third technique is the symmetric random walk, which can be realized in two ways. First, any measure on $SO_p$ can be used to generate a random walk on $SO_p$. Let $g_1, g_2, \ldots$ be such a random walk. As in the Torus method, the images of any initial point in $G_{1p}$ or $G_{2p}$ under these rotations form a Grand Tour. If the original measure on $SO_p$ is constant on conjugacy classes, the random walk will be what I call symmetric. A second, slightly more inclusive concept is to generate a symmetric random walk on $G_{1p}$ or $G_{2p}$ directly. Let $x_i$ be the current position of the random walk, $x_{i+1}$ is chosen as follows. Let $K$ be the subgroup of $SO_p$, acting on $G_{1p}$ or $G_{2p}$, that fixes $x_i$. This partitions $G_{1p}$ or $G_{2p}$ into orbits; $y$ and $z$ are in the same orbit if $y = gz$ for some $g$ in $K$. Select an orbit according to some preselected measure on the orbits and choose $x_{i+1}$ according to the uniform distribution on that orbit. If the orbit measure is concentrated on orbits "close to" $x_i$, then the random walk will be good under criterion 1. The path of a random walk is
not smooth, so this sort of sequence fails criterion 2. Criterion 3, the rate at which such a random walk becomes dense, is the problem which motivated this thesis. Since $G_{1p}$ is isomorphic to $S_{p-1}$ with opposite points identified, this motivates consideration of the covering problem suggested in the introduction. This chapter and the next give a partial answer to the $G_{1p}$ problem; possibilities for $G_{2p}$ are discussed in Chapter 7.

Asimov also suggests a random walk that will be better in terms of criterion 2. Intuitively it can be thought of as a random walk with drift, as opposed to the symmetric random walks considered here. It is possible that the techniques of this thesis, as extended in Section 4.3 can give information about criterion 3 for this kind of random walk.

5.3 COVERING AND PACKING A SPHERE WITH CAPS

Mean covering times for random walks and Brownian motions on spheres will be contained by upper and lower bounds as follows. For a lower bound, find a set of disjoint caps with the property that if the process has not visited each cap, then it has not come within $\epsilon$ of all points of $S_{p-1}$. Then a lower bound on the time to visit all these caps is a lower bound on the covering time as well. Similarly, for an upper bound find a set of caps such that if the process has visited each cap of the set, then it has been within $\epsilon$ of every point of $S_{p-1}$.

The first problem, finding caps for a lower bound, is that of packing caps of radius $\epsilon$ without overlap onto spheres. It is of interest in coding theory and has been investigated extensively. Sloane (1982) gives
the best current results and a good bibliography. The second problem
is that of covering a sphere with caps. Here less work has been done;
Rogers (1963) is one paper on the subject. From Section 2.3 it is clear
that the time taken to visit a set of caps depends only logarithmically
on the size of the set. Because of this crude solutions to the deter-
ministic covering and packing problems are sufficient to give bounds on the
mean covering times for processes that become tight as the cap size ε
shrinks. Rather than try to use the results in the literature, it is
much easier to present the following crude construction. This construc-
tion solves both problems adequately for the needs of this thesis, but
does not give the best results for the deterministic covering or packing
problems.

Fix r, 0 < r < 1, and let rZ be the set {...,−2r,−r,0,r,2r,...}. Let
A = A(r,p) be the intersection of (rZ)^P with the closed unit ball
in \( \mathbb{R}^P \). Or, \( A = \{ (X_1,...,X_p) : X_i \in rZ, i=1,...,p \text{ and } \sum_{i=1}^{P} X_i^2 < 1 \} \).
Define B = B(r,p) ⊂ A as follows. Suppose \( (X_1,...,X_p) \in A \) and \( |X_i| \)
is maximal among \( |X_1|,...,|X_p| \). Then \( (X_1,...,X_p) \in B \) if and only if
\( (X_1,...,X_{i-1},|X_i|+r,X_{i+1},...,X_p) \notin A \). In words, \( (X_1,...,X_p) \in B \) if it is
on the boundary of A, in the sense that increasing its largest coordinate
moves it out of A. Finally, let D = D(r,p) be the projection of B
outward onto \( S_{p-1} \). D = \{ X/\|X\| : X \in B \}, \|X\| = \left( \sum_{i=1}^{P} X_i^2 \right)^{1/2} . The points
of D are fairly dense in \( S_{p-1} \) without any two being too close together.

For a ∈ R, let \( \lambda_r(a) = \lfloor \frac{a}{r} \rfloor \) be the greatest integer multiple of r
that is ≤ a. Similarly \( u_r(a) = \lceil \frac{a}{r} \rceil \) is the smallest integer multiple
r that is > a. Then \( \lambda_r(a) \leq a < u_r(a) \) and \( u_r(a) - \lambda_r(a) = r \). Suppose
\( X = (X_1,...,X_p) \in S_{p-1} \), without loss of generality \( 0 \leq X_1 \leq X_2 \leq \cdots \leq X_p \).
\((X_1, \ldots, X_p)\) lies in a \(p\)-dimensional cube with sides of length \(r\) lying along the coordinate axes, inner corner \((\ell(X_1), \ldots, \ell(X_p))\), and outer corner \((u(X_1), \ldots, u(X_p))\). Call this cube \(K(X)\).

Lemma 5.3.1). At least one corner of \(K(X)\) lies in \(B\).

Proof. The inner corner of \(K(X)\) is in \(A\); the outer corner is not. Consider the sequence

\[(\ell(X_1), \ldots, \ell(X_p)), (u(X_1), \ell(X_2), \ldots, \ell(X_p)), (u(X_1), u(X_2), \ell(X_3), \ldots, \ell(X_p)), \ldots, (u(X_1), \ldots, u(X_p)).\]

Clearly the last point in this sequence of corners that lies in \(A\) also lies in \(B\). \(\square\)

The longest diagonals of this \(p\)-dimensional cube have length \(r\sqrt{p}\).

Thus any point \(X \in S_{p-1}\) is within a distance \(r\sqrt{p}\) of a point of \(B\).

For \(b \in B\) and \(d = b/\|b\| \in D\), consider the distance from \(b\) to \(d\).

\(d\) is the point of \(S_{p-1}\) nearest to \(b\). Increasing the largest coordinate of \(b\) moves it outside \(S_{p-1}\), hence \(b\) is within a distance \(r\) of \(S_{p-1}\), and hence \(d\). By the triangle inequality it follows that any point \(X\) of \(S_{p-1}\) is within a linear distance \(r\sqrt{p+1}\) of some point of \(D\).

Note the term linear distance was used. For two points \(X\) and \(d\) of \(S_{p-1}\) there are two natural notions of the distance between them. The first, linear distance, is the length of the line segment in \(\mathbb{R}^p\) connecting them. The second, geodesic distance, is the length of the shortest geodesic in \(S_{p-1}\) connecting them. This is equivalent to the angle between the two rays from the origin to \(X\) and \(d\). For two points
in $S_{p-1}$, geodesic rather than linear distance is of interest. Thus, referring to two points on $S_{p-1}$, the unqualified term distance will refer to geodesic distance. The terms linear and geodesic will be used to avoid confusion.

Thus, in the following lemma, radius refers to geodesic radius.

Lemma (5.3.2). For $r\sqrt{p+1} < 1$, caps of radius $\arcsin(r\sqrt{p+1})$ about the points of $D(r,p)$ cover $S_{p-1}$.

Proof. Let $X$ be an arbitrary point of $S_{p-1}$ and $d$ a point of $D$ within a distance $r\sqrt{p+1}$ of $X$. Consider the isosceles triangle with corners at the origin, $X$, and $d$. The triangle has side lengths 1,1, and $\omega$, where $\omega < r\sqrt{p+1}$ is the distance from $X$ to $d$. Let $\theta$ be the interior angle at the origin, thus $\theta$ is the geodesic distance from $X$ to $d$. By the law of sines, $\frac{\sin \theta}{\omega} = \sin \left( \frac{\pi - \theta}{2} \right) < 1$ so $\sin \theta < \omega < r\sqrt{p+1}$. Thus $\theta < \arcsin(r\sqrt{p+1})$. Every point $X$ of $S_{p-1}$ is within a distance $\arcsin(r\sqrt{p+1})$ of some point of $D$, so caps of radius $\arcsin(r\sqrt{p+1})$ about the point of $D$ cover $S_{p-1}$. More precisely, it can be seen that caps of radius $2 \arctan\left( \frac{r\sqrt{p+1}}{2} \right)$ about the points of $D$ cover $S_{p-1}$. 

The next goal is to find disjoint caps with centers at the points of $D$.

Lemma (5.3.3). Caps of radius $r/2\sqrt{2}$ about the points of $D(r,p)$ are disjoint.

Proof. Let $X^*$ and $Y^*$ belong to $D(r,p)$ and let $X$ and $Y$ be the
corresponding points in $B(r,p)$. Consider two cases:

Case 1: $X$ and $Y$ differ in only one coordinate.

Case 2: $X$ and $Y$ differ in two or more coordinates.

First consider Case 1. Without loss of generality, suppose $X = (X_1, \ldots, X_p)$ with $X_i > 0$ for $i=1, \ldots, p$. Also suppose $Y = (X_1+kr, X_2, \ldots, X_p)$ for some positive integer $k$. Since $X$ and $Y$ are in $B$, $X_1$ cannot be maximal among $X_1, \ldots, X_p$, so $X_1 < 1/\sqrt{2}$. Let $Z$ be the vector $(X_1, 0, \ldots, 0)$. $X, Y, Z$ and $O$, the origin, all lie in a plane. Consider the right triangle with corners $O, X$, and $Z$. Side $OZ$ is shorter than side $ZX$, hence angle $ZOX$ is smaller than angle $ZOX$, so angle $ZOX$ is less than $\frac{\pi}{4}$. Now consider triangle $OXY$. $XY$ is parallel to $OZ$, hence angle $OXY$ is between $\frac{\pi}{2}$ and $\frac{3\pi}{4}$. Angle $XOY$ is the geodesic distance $\theta$ from $X^*$ and $Y^*$. Side $OY$ has length $< 1$. By the law of sines

$$\frac{\sin \theta}{kr} = \frac{\sin(\text{angle } OXY)}{||OY||}.$$  

Thus $\sin \theta > kr \sin(\frac{3\pi}{4}) > \frac{r}{\sqrt{2}}$, so $\theta > \sin \theta > \frac{r}{\sqrt{2}}$ as well.

Case 2: $X$ and $Y$ differ in two or more coordinates. Then $||X-Y|| > \sqrt{2} \cdot r$. Consider triangle $OXY$. Sides $OX$ and $OY$ have lengths at least $1-r$, else $X$ and $Y$ would not be in $B$. Without loss of generality suppose $||X|| > ||Y||$. Again let $\theta$ be the interior angle at $0$, which is the geodesic distance from $X^*$ to $Y^*$. Let $\gamma$ be the angle at $Y$. Since $||X|| > ||Y||$, $\gamma > \pi - \gamma - \theta$. 

86
The following shows that \( \gamma \) cannot be too large. Suppose \( \gamma > \frac{\pi}{2} \).

Then by the law of cosines

\[
\|x\|^2 = \|y\|^2 + \|x-y\|^2 - 2\|y\|\|x-y\| \cos \gamma.
\]

Thus, since \( \cos \gamma < 0 \),

\[
1 \geq (1-r)^2 + 2r^2 - 2(1-r)\sqrt{2}r \cos \gamma.
\]

Simplifying

\[
\cos \gamma \geq -\frac{1-\frac{3}{2}r}{\sqrt{2}(1-r)}, \text{ or } \gamma \leq \arccos\left(-\frac{1-\frac{3}{2}r}{\sqrt{2}(1-r)}\right).
\]

\((1-\frac{3}{2}r)/(1-r) < 1, \text{ so } \gamma < \arccos\left(-\frac{1}{\sqrt{2}}\right) = \frac{3\pi}{4} . \]

Now, to show \( \theta \) must be large, suppose \( \theta \leq a \), for some \( a \). By the law of sines

\[
\frac{\sin \theta}{\|X-Y\|} = \frac{\sin \gamma}{\|X\|}, \text{ or } \sin \theta \geq \sqrt{2}r \sin \gamma.
\]

For \( \theta \leq a \), \( (\pi-a)/2 < \gamma < \frac{3\pi}{4} \), so

\[
\sin \gamma \geq \min(\sin(\frac{3\pi}{4}), \sin(\frac{\pi-a}{2})) = \min\left(-\frac{1}{\sqrt{2}}, \cos \frac{a}{2}\right) .
\]

If \( \frac{\pi}{2} > \theta > a \), \( \sin \theta > \sin a \), so for any \( \theta \),

\[
\sin \theta \geq \min(\sin a, r, \sqrt{2}r \cos \frac{a}{2}) .
\]

This is true for any choice of \( a \). Choosing \( a = 2\arcsine\left(-\frac{r}{\sqrt{2}}\right) \), so \( \sin a = \sqrt{2}r \cos \frac{a}{2} \).
\[
\sin \theta \geq \min(r, \sqrt{2} r \sqrt{1-r^2/2}).
\]

\(r < 1\), so here \(\theta > \sin \theta \geq r\).

In either case \(\theta\), the geodesic distance from \(X^*\) to \(Y^*\) is larger than \(r/\sqrt{2}\). It follows that caps of half this radius, or \(r/2\sqrt{2}\), about the points of \(D(r,p)\) will be disjoint.

Next consider the number of points in \(D(r,p)\). Rather than trying to count them, Lemmas (5.3.2) and (5.3.3) will be used to give crude bounds.

Let \(N(r,p)\) be the number of points in \(D(r,p)\).

The standard spherical coordinates in \(S_{p-1}\) are \(\theta_1, \ldots, \theta_{p-1}\). If \(X \in S_{p-1}\) has rectangular coordinates \((x_1, \ldots, x_p)\), then \(X_i = \prod_{j=1}^{i-1} \sin \theta_j \cdot \cos \theta_i\). In spherical coordinates, Haar invariant measure is

\[
\prod_{j=1}^{p-1} (\sin \theta_j)^{p-j-1} \sin \theta_j \, d\theta_j
\]

on \(S_{p-1}\) with range \(0 \leq \theta_i \leq \pi\) for \(i=1,2,\ldots,p-2\) and \(0 \leq \theta_{p-1} < 2\pi\).

A cap of radius \(\rho\) about the North Pole of \(S_{p-1}\) is

\[\{(\theta_1, \ldots, \theta_{p-1}) : \theta_1 < \rho\}\]

and thus has volume

\[
\int_0^\rho \sin^{p-2} \theta d\theta / \int_0^{\pi} \sin^{p-2} \theta d\theta = V_{p-1}(\rho).
\]

Haar measure has been normalized here so that \(S_{p-1}\) has volume 1. To evaluate \(V_{p-1}(\rho)\), from Gradshteyn and Ryzhik (1980) formula 3.621 note

\[
\int_0^{\pi} \sin^{p-2} \theta d\theta = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)}.
\]
\[ \int_0^\rho \sin^{p-2} \theta d\theta = \int_0^\rho \theta^{p-2} (\sin \frac{\theta}{\rho})^{p-2} d\theta , \]

which is bounded below by

\[ \left( \frac{\sin \frac{\rho}{\rho}}{\rho} \right)^{p-2} \frac{\rho^{p-1}}{p-1} . \]

and above by

\[ \frac{\rho^{p-1}}{p-1} . \]

Thus

\[ (5.3.4) \quad \left( \frac{\sin \frac{\rho}{\rho}}{\rho} \right)^{p-2} \frac{\rho^{p-1}}{p-1} \frac{\Gamma \left( \frac{p-1}{2} \right) \Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{p}{2} \right)} \leq \frac{\rho^{p-1}}{p-1} \frac{\Gamma \left( \frac{p-1}{2} \right) \Gamma \left( \frac{1}{2} \right)}{\Gamma \left( \frac{p}{2} \right)} . \]

These estimates lead to bounds on \( N(r,p) \).

Proposition (5.3.5). There exist constants \( U(p) \) and \( L(p) \) such that

\[ \frac{L(p)}{r^{p-1}} \leq N(r,p) \leq \frac{U(p)}{r^{p-1}} . \]

Proof. Recall that \( N(r,p) \) is the number of points in \( D(r,p) \). Caps of radius \( r^{p+1} \) about the points of \( D(r,p) \) cover \( S_{p-1} \), so from (5.3.4)

\[ \frac{N(r,p)(p+1)}{p-1} \frac{\rho^{p-1}}{p-1} \frac{\Gamma \left( \frac{p-1}{2} \right) \left( \frac{1}{2} \right)}{\Gamma \left( \frac{p}{2} \right)} \geq 1 . \]

Similarly caps of radius \( r/2\sqrt{2} \) are disjoint, so

89
\[
\frac{N(r,\rho)}{(2\sqrt{2})^{p-1}} \frac{\sin r}{r} \leq 1.
\]

The \((\sin r/r)^{p-2}\) term causes no problem if it converges to one as \(r \to 0\).

For \(r\) large, an upper bound is trivial.

The results of this section are summarized in the following proposition.

Proposition (5.3.6). Given positive integer \(p\), there exist numbers \(U(p)\) and \(L(p)\) such that for any \(\varepsilon > 0\)

1) There exist at least \(L(p)\varepsilon^{-(p-1)}\) disjoint spherical caps of
radius \(\varepsilon\) on \(S_{p-1}\).

2) There is a set of at most \(U(p)\varepsilon^{-(p-1)}\) caps of radius \(\varepsilon\) that
cover \(S_{p-1}\).

3) The sets of caps considered above are symmetric in the origin.

Thus these sets can be considered as sets of pairs of opposite
caps with, of course, half as many elements.

Proof. 1) and 2) are Proposition (5.3.5) reworded. 3) follows from the
observation that the points of \(D(r,p)\) are symmetric in the origin.

5.4 RANDOM WALKS AND BROWNIAN MOTION

This section considers random walks and Brownian motion on spheres.
A symmetric random walk is defined first, then Brownian motion is intro-
duced as a limiting diffusion in the usual manner. Random walk in this
section will mean symmetric random walk. See Bingham (1972) for a more
general notion. Roberts and Ursell (1960) first considered the convergence of random walks on spheres to a diffusion. They derived an expansion in spherical harmonics for the distribution of this limiting diffusion, here called Brownian motion. Watson (1983) discusses this and other diffusions on spheres.

First a (symmetric) random walk on a sphere is defined. Recall $S_{p-1}$ is the unit sphere in $\mathbb{R}^p$. Let $\theta_1, \theta_2, \ldots$ be independent random variables, each uniformly distributed on $S_{p-2}$. Let $\psi_1, \psi_2, \ldots$ be independent and identically distributed on $[0, \pi]$. Further let $\psi_1, \psi_2, \ldots$ and $\theta_1, \theta_2, \ldots$ be independent. Let $X(0) \in S_{p-1}$ be the initial position of the random walk. $X(1), X(2), \ldots$ denote the successive positions of the random walk. For $i=1, 2, \ldots$, given $X(i-1)$ the distribution of $X(i)$ is determined as follows. For each $\alpha \in [0, \pi]$ and $x \in S_{p-1}$ let $0_\alpha(x)$ be the set of points of $S_{p-1}$ that are a geodesic distance $\alpha$ from $x$. $0_\alpha(x)$ is isomorphic to $S_{p-2}$ for $0 < \alpha < \pi$. For each $\alpha$ and each $x$, assume there is an isomorphism $\eta(x, \alpha)(\cdot)$ from the sphere $S_{p-2}$ where $\theta_1, \theta_2, \ldots$ are defined to $0_\alpha(x)$. Now let $X(i) = \eta(X(i-1), \psi_i)(\theta_i)$. Strictly speaking, $\eta$ is only defined for $0 < \alpha < \pi$. However, $0_0(x) = \{x\}$ and $0_\pi(x) = \{-x\}$, so setting $\eta(x, 0)(\cdot) = x$ and $\eta(x, \pi)(\cdot) = -x$ completes its definition. Informally $X(i)$ can be thought of as uniformly distributed on the set of points that are a distance $\psi_i$ from $X(i-1)$. Defining a random walk in this fashion will prove to be useful when Brownian motion is considered.

For the purpose of this thesis, it is often enough to look at one coordinate of the random walk only. Suppose $X(i)$ has coordinates
\((X^1(i),...,X^p(i))\) be some rectangular coordinate system. Let NP, for North Pole, be the point \((1,0,...,0)\), and SP, for South Pole, be the point \((-1,0,...,0)\). The geodesic distance from \(X(i)\) to NP is \(\arccos(X^1(i))\). Similarly the distance to SP is \(\pi-\arccos(X^1(i))\). Thus to study the time \(X(\cdot)\) takes to come within \(\varepsilon\) of NP, it is enough to look at \(\arccos(X^1(\cdot))\), or equivalently \(X^1(\cdot)\), only. Similarly, for the two cap problem, it is also enough to look at \(X^1(\cdot)\) only.

Lemma (5.4.1). \(\{X^1(i),i=0,1,...\}\) forms a Markov process.

Proof. NP, \(X^1(i-1)\) and \(X^1(i)\) lie on the surface of an \(S_2\) with center at the origin. Consider the spherical triangle they span. This triangle may be degenerate, but formula (5.4.3) will hold in this case as well. Side \(NPX(i-1)\) has length \(\arccos(X^1(i-1))\), side \(X(i-1)X(i)\) has length \(\psi_1\), and side \(X(i)NP\) has length \(\arccos(X^1(i))\). Consider the spherical angle \(\gamma_1\) with its vertex at \(X(i-1)\).

The geodesics from \(X(i-1)\) in the directions of NP and \(X^1(i)\) intersect \(0_{\pi/2}(X_{i-1})\) in unique points \(a\) and \(b\). \(0_{\pi/2}(X_{i-1})\) is an \(S_{p-2}\), and the geodesic distance from \(a\) to \(b\) is the angle \(\gamma_1\). Due to the uniformity of \(\theta_1\) and its independence of \(\psi_1\) and the past, \(b\) is uniformly distributed on \(0_{\pi/2}(X_{i-1})\), independently of \(a\). Thus the distance from \(a\) to \(b\), and hence \(\gamma_1\), has the density of the distance from a randomly chosen point on \(S_{p-2}\) to its north pole. This is just the first spherical coordinate on \(S_{p-2}\), so the density is
$$\frac{\sin^{p-3}(\gamma)}{\int_0^\pi \sin^{p-3}(\gamma) d\gamma} \quad 0 \leq \gamma \leq \pi.$$  

By the independence and uniformity of \(\theta_i\), \(\gamma_i\) is independent of \(X(1), \ldots, X(i-1), \psi_i\). By the law of cosines for spherical triangles

$$X^1(i) = X^1(i-1)\cos(\psi_i) + \sqrt{1-[X^1(i-1)]^2 \sin(\psi_i)\cos(\gamma_i)}.$$  

With \(\psi_i\) and \(\gamma_i\) independent of each other and the past, this completely describes the transition mechanism of \(X^1(\cdot)\). \(X^1(i)\) depends on \(X(0), \ldots, X(i-1)\) only through \(X^1(i-1)\). Thus \(X^1(\cdot)\) has the Markov property.

Next the step sizes \(\psi_i\) are shrunk and time is rescaled to obtain a limiting diffusion. For present purposes, it is sufficient to look at only \(X^1(\cdot)\), rather than the whole random walk \(X(\cdot)\). From Lemma (5.4.1), the Markov process \(X^1(\cdot)\) can be described by taking \(\{\psi_i\}\) as before, \(\{\gamma_i\}, i=1,2,\ldots\) independent of each other and \(\{\gamma_i\}\) with density (5.4.2) and letting the transitions of \(X^1(\cdot)\) be governed by (5.4.3). For each \(s > 0\), define a Markov chain \(X^1_s(\cdot)\), taking steps at the time points \(\{s, 2s, \ldots\}\) as follows \(X^1_s(0) = X^1(0)\) and

$$X^1_s(is) = X^1_s((i-1)s)\cos(\psi_i\sqrt{s}) + \sqrt{1-[X^1_s((i-1)s)]^2 \sin(\psi_i\sqrt{s})\cos(\gamma_i)}.$$  

Note that \(X^1_0(i) = X^1(i)\) for all \(i\). Intuitively \(X^1_s(\cdot)\) takes a step at each integer multiple of \(s\), but the step sizes are shrunk by a factor.
of $\sqrt{s}$. As $s \to 0$, this sequences of processes will have a diffusion $W(\cdot)$ as a weak limit.

Roberts and Ursell (1960) and Watson (1983) both discuss this diffusion and the corresponding diffusion on the sphere. If $X_s(\cdot)$ is defined to be $\eta(X_s((i-1)s), \sqrt{s} \psi_i)(\theta_i)$ and $X_s(0) = X(0)$, then as $s \to 0$, $X_s(\cdot)$ converges, in the sense of finite dimensional distributions, to symmetric Brownian motion on the sphere. $X_s^1(\cdot)$ defined by (5.4.4) has the same finite dimensional distributions as the first coordinate of $X_s(\cdot)$. Thus $W(\cdot)$ can be considered the first coordinate of Brownian motion on the sphere or, equivalently, the cosine of the geodesic distance of this Brownian motion to the North Pole. The symmetry of Brownian motion and the diffusion $W(\cdot)$ together can be used to describe Brownian motion completely. Thus, in discussing Brownian motion, Roberts and Ursell (1960) restrict their attention to $W(\cdot)$. They give an expansion for the distribution of $W(t)$, for $t > 0$, in terms of spherical harmonics.

In the remainder of this section, the drift and infinitesimal variance of $W(\cdot)$ are derived. It will not be shown until Section 5.6 that the diffusion obtained actually is a weak limit of $X_s^1(\cdot)$ and that hitting times for $X_s^1(\cdot)$ converge to those of $W(\cdot)$ as $s \to 0$.

The drift of $W(\cdot)$ should be

$$\mu_x = \lim_{s \to 0} \frac{1}{s}[E[x\cos(\psi\sqrt{s}) + \sqrt{1-x^2} \sin(\psi\sqrt{s}) \cos \gamma] - x]$$

where $\psi$ and $\gamma$ are independent and are distributed as before.
\[ E \cos \gamma = \frac{\int_0^\pi \cos \gamma \sin^{p-3} \gamma d\gamma}{\int_0^\pi \sin^{p-3} \gamma d\gamma} = 0 \]

by symmetry. For \( s \) small, \( \cos(\psi \sqrt{s}) = 1 - \frac{\psi^2 s}{2} + Z_s s^2 \), where each \( Z_s \) is a random variable with \( \{Z_s\} \) bounded uniformly in \( s \). Then

\[ \mu_x = -\frac{x E \psi^2}{2}, \] so \( \mathcal{W}(\cdot) \) has drift toward 0. Similarly,

\[ \sigma_x^2 = \lim_{x \to 0} \frac{1}{s} \left\{ E\left[ (x \cos(\psi \sqrt{s}) + \sqrt{1-x^2} \sin(\psi \sqrt{s}) \cos \gamma)^2 \right] - \right. \]

\[ \left. \frac{1}{s} \frac{2}{E(\sin(\psi \sqrt{s}) \cos \gamma)} \right\}. \]

\[ E \cos^2 \gamma = \frac{1}{p-1}, \] so \( \sigma_x^2 = \frac{(1-x^2) E \psi^2}{p-1}. \)

\( \mathcal{W}(\cdot) \) is a diffusion on \([-1,1]\] with

(5.4.5)  

\[ \text{drift } \mu_x = -\frac{\lambda x}{2} \text{ and } \]

\[ \text{variance } \sigma_x^2 = \frac{(1-x^2) \lambda}{p-1}, \] where

\[ \lambda = E \psi^2 \] is an arbitrary positive parameter. To be precise, this should be written \( \mathcal{W}_\lambda(\cdot) \). However, \( \lambda \) will enter only in inconsequential ways, so the dependence of \( \mathcal{W}(\cdot) \) on \( \lambda \) can safely be ignored.

5.5 EXPECTED HITTING TIMES FOR BROWNIAN MOTION

This section investigates the time taken by \( \mathcal{W}(\cdot) \) to leave an interval of the form \([-r_1, r_2]\), where \( r_1 \) may be 1. With the interpretation of
\( W(\cdot) \) as the first rectangular coordinate of symmetric spherical Brownian motion, this time is equivalent to the time taken by Brownian motion to enter a cap of radius \( \arccos(r_1) \) about the South Pole or a cap of radius \( \arccos(r_2) \) about the North Pole. Though the expectation of the hitting time is calculated explicitly, it is too complex to be useful.

Some simpler asymptotics, valid as \( r_1 \) and \( r_2 \) approach one, are given.

Let \( x \) be some point in \([-1,1]\) and let \( x, r_1, r_2 \) \( T \) for short, be the first time, starting from \( x \), that \( W(\cdot) \) leaves \((-r_1, r_2)\). Let \( f(x, r_1, r_2) \), \( f(x) \) when \( r_1 \) and \( r_2 \) are fixed, be \( E_T^{x, r_1, r_2} \).

Finding \( f(x, r_1, r_2) \) is a straightforward, though lengthy, exercise.

Following Karlin and Taylor (1981) Chapter 15, p. 193, \( f(x) \) satisfies

\[
(5.5.1) \quad -1 = \frac{\lambda x}{2} f'(x) + \frac{\lambda}{2} \frac{(1-x^2)}{p-1} f''(x)
\]

subject to \( f(-r_1) = f(r_2) = 0 \).

Let

\[
s(x) = \exp\{- \int_0^x 2(-\frac{\lambda y}{2}) \int \frac{\lambda (1-y^2)}{p-1} dy \} = (1-x^2)^{(p-1)/2}.
\]

\( f^x \) stands for the integral evaluated only at the upper limit \( x \). Including an arbitrary lower limit of integration would not affect the results, but would complicate the formulas.

Let \( S(x) = f^x s(y) dy \). With the trigonometric substitution \( y = \cos z \), \( S(x) \) is

\[
- \int_{\arccos(x)}^{\pi/2} \frac{dz}{(\sin z)^{p-2}}.
\]
This must be done in two cases, depending on whether $p$ is even or odd.

Formulas 5.515.1 and 5.515.2 in Gradshteyn and Ryzhik yield

\[(5.5.2) \text{ Case 1: } p \text{ even} \]

\[S(x) = \frac{x}{p-3} \sum_{k=0}^{\frac{p}{2} - 2} \frac{C_{k, p-2}}{(1-x^2)^{\frac{p-3}{2} - k}},\]

where

\[C_{k, \ell} = \frac{(2\ell-2)!! (2\ell-2k-3)!!}{(2\ell-2k-2)!! (2\ell-3)!!}\]

and

\[n!! = \begin{cases} 1 \cdot 3 \cdot 5 \cdot \cdots \cdot n & \text{n odd} \\ 2 \cdot 4 \cdot 6 \cdot \cdots \cdot n & \text{n even} \end{cases}\]

\[(5.5.3) \text{ Case 2: } p \text{ odd} \]

\[S(x) = \frac{x}{p-3} \sum_{k=0}^{\frac{p}{2} - 5} \frac{d_{k, p-3}}{(1-x^2)^{\frac{p-3}{2} - k}} - \frac{(p-4)!!}{(p-3)!!} \log \sqrt{\frac{1-x}{1+x}},\]

which, for $p=3$, is understood to be $\log \sqrt{\frac{1-x}{1+x}}$.

\[d_{k, \ell} = \frac{(2\ell-1)!! (2\ell-2k-2)!!}{(2\ell-2k-1)!! (2\ell-2)!!} \cdot\]
Next let

\[ m(x) = \left[ s(x) \sigma^2(x) \right]^{-1} = \frac{(p-1)(1-x^2)}{\lambda(1-x^2)} \]

\[ M(x) = \int_x^\infty m(y)dy \text{ is needed, also done in two cases. Gradshetyn and Ryzhik's formulas 2.511.2 and 3 are used, after a trigonometric substitution.} \]

(5.5.4) Case 1: \( p \) even

\[ M(x) = \frac{p-1}{\lambda} \left[ \frac{x}{p-2} \sum_{k=0}^{p-2} d_{k, \frac{p-1}{2}} (1-x^2)^k \right] - \frac{(p-3)!!}{(p-2)!!} \arccos(x) \]

(5.5.5) Case 2: \( p \) odd

\[ M(x) = \frac{(p-1)x}{(p-2)\lambda} \sum_{k=0}^{p-3} C_{k, \frac{p-1}{2}} (1-x^2)^k \]

which, for \( p=3 \), is \( \frac{2x}{\lambda} \).

The last integral needed is \( \int_x^\infty s(y)m(y)dy \), here called \( I(x) \).

Again there are two cases.

(5.5.6) Case 1: \( p \) even

\[ I(x) = \frac{p-1}{(p-3)\lambda} \left[ \frac{x^2}{2} - \sum_{k=1}^{p-4} C_{k, \frac{p-2}{2}} \frac{(1-x^2)^{k+1}}{2k+2} \right] \]

(5.5.7) Case 2: \( p \) odd

98
\[ I(x) = \frac{p-1}{\lambda} \left\{ \frac{p-5}{2} \sum_{k=0}^{\frac{p-3}{2}} d_{k, \beta} \frac{p-3}{p-3} \int_{x}^{\infty} y(1-y^{2})^{k} \right. \]
\[ \frac{p-3}{2} \left[ 1-y \right] \frac{1}{1+y} dy \right\} \]

which, for \( p \geq 5 \), is
\[ \frac{p-1}{\lambda} \left\{ \frac{x^{2}}{p-3} - \frac{p-5}{2} \sum_{k=1}^{\frac{p-3}{2}} d_{k, \beta} \frac{(1-x^{2})^{k+1}}{2(p-3)(k+1)} \right. \]
\[ \frac{p-3}{2} \left[ \sum_{i=0}^{p-3} (-1)^{i} \right] \frac{2i+1}{2i+1} \left[ (x^{2i+1}+1) \log(1-x) \right. \]
\[ - (x^{2i+1}+1) \log(1+x) - 2 \sum_{r=0}^{\frac{p-1}{2}} \frac{x}{2r+1} \right\} \]

and for \( p=3 \) is
\[ \frac{1}{\lambda} \left[ (1-x) \log(1-x) + (1+x) \log(1+x) \right] . \]

Let
\[ G(x,y) = \frac{2[S(x)-S(-r_{1})][S(r_{2})-S(y)]}{S(r_{2})-S(-r_{1})} \quad -r_{1} \leq x \leq y \leq r_{2} \]
\[ \frac{2[S(r_{2})-S(x)][S(y)-S(-r_{1})]}{S(r_{2})-S(-r_{1})} \quad -r_{1} \leq y \leq x \leq r_{2} \]

be the Green's function. Then, following Karlin and Taylor (1981)
\[ f(x, r_1, r_2) = \int_{-r_1}^{r_2} G(x, y)m(y) dy. \]

This turns out to be

\begin{equation}
S(r_2) - S(x) \quad 2\left(\frac{r_2 - S(r_2)}{S(r_2) - S(r_1)}\right) (S(-r_1)M(-r_1) - I(-r_1)) \\
+ \frac{S(x) - S(-r_1)}{S(r_2) - S(-r_1)} (S(r_2)M(r_2) - I(r_2)) - (S(x)M(x) - I(x)) \right].
\end{equation}

(5.5.8)

\(S(x), M(x),\) and \(I(x)\) are known from (5.5.2)-(5.5.7), so their values can be inserted in (5.5.8) to give an exact solution, but one that is difficult to use. It will be far more useful to have a comprehensible approximation to (5.5.8) as \(r_1\) and \(r_2\) approach \(1\). This is the aim of the following calculations.

Note that \(S(0)\) is always \(0\) and \(S(-x) = -S(x)\). For present purposes, it is sufficient to examine \(S(x)\) as \(x \rightarrow 1\).

(5.5.9) Case 1: \(p\) even

The dominant term in (5.5.2) as \(x \rightarrow 1\) is

\[ \frac{x}{p-3} C_{0, \frac{p-2}{2}} \frac{1}{p-3} \frac{1}{2} \frac{1}{2} (1-x^2). \]

\(C_{0, \frac{p-2}{2}} = 1\). The next term in the expansion is

\[ 0((1-x^2)^{1-p/2}). \]
Thus as $x \to 1$

$$S(x) = \frac{x}{p-3} \cdot \frac{1}{p-3} \frac{(1+0(1-x^2))}{(1-x^2)^2}.$$

(5.5.10) Case 2: $p$ even

Here the dominant term is

$$\frac{x}{p-3} d_{0, \frac{p-3}{2}} (1-x^2)^{-\left(\frac{p-3}{2}\right)} \quad \text{for } p \geq 5$$

$$\log \sqrt{\frac{1+x}{1-x}} \quad \text{for } p = 3.$$

Thus as $x \to 1$

$$S(x) = \frac{x}{p-3} \frac{1}{p-3} \frac{(1+0(1-x^2))}{(1-x^2)^2} \quad \text{for } p \geq 5$$

$$\log(-\frac{1}{1-x^2}) + 0(1) \quad \text{for } p = 3.$$

Next consider $M(x)$.

(5.5.11) Case 1: $p$ even

From (5.5.4) $M(0) = \frac{-\pi}{2\lambda} \frac{(p-1)!!}{(p-2)!!}.$

As $x \to -1$ $M(x)$ is $-\frac{\pi}{\lambda} \frac{(p-1)!!}{(p-2)!!} + 0(1-x^2)$.
For the case $x \to 1$, consider the formula (Gradshteyn and Ryzhik, 1.641)

$$\arccos(\sqrt{1-x^2}) = \arcsin(x) = \sum_{i=0}^{\infty} x^{2i+1} \frac{(2i-1)!!}{(2i)!!(2i+1)} , \quad x^2 < 1 .$$

(5.5.4) can be rewritten

$$M(x) = \frac{(p-1)!!}{\lambda(p-2)!!} \left\{ \sum_{j=0}^{\frac{p-2}{2}} \frac{j + \frac{1}{2}}{x(2j)!!} \frac{x(2j)!!}{(2j-1)!!} - \arcsin(\sqrt{1-x^2}) \right\} ,$$

from which it follows that $M(x)$ is $O((1-x^2)^{3/2})$ as $x \to 1$.

(5.5.12) Case 2: $P$ odd

$M(0) = 0$ and $M$ is antisymmetric. As $x \to 1$, from (5.5.5)

$$M(x) = \frac{(p-1)!!}{\lambda(p-2)!!} + O(1-x^2) .$$

Finally consider $I(x)$.

(5.5.13) Case 1: $p$ even

As $x \to \pm 1$, $I(x) \to \frac{p-1}{2\lambda(p-3)}$.

(5.5.14) Case 2: $p$ odd

By L'Hôpital's Rule $(x^{m-1})\log(1-x) \to 0$ as $x \to 1$ for $m \geq 1$

so $I(x)$ is bounded for $-1 \leq x \leq 1$.

102
Now expected hitting times can be approximated. Recall that
\( f(x, r_1, r_2) \) is the expected value of the time taken by the diffusion
\( W(t) \) to leave \((-r_1, r_2)\) starting from \( x \). This is also the expected
time taken by symmetric spherical Brownian motion to hit either a cap
of radius \( \arccos(r_1) \) about the South Pole or a cap of radius \( \arccos(r_2) \)
about the North Pole starting from a point a distance \( \arccos(x) \) from
the North Pole. Four specific problems will be considered. The first
two involve \( f(x, r, r) \) as \( r \to 1 \). The first of these is \( f(0, r, r) \), the
second is \( f(r', r, r) \), where \( r' \) is slightly less than \( r \). The other
two involve \( f(x, 1, r) \), defined as \( f(x, 1, r) = \lim_{r_2 \to 1} f(x, r_2, r) \). This can
easily be shown to be the time to hit a cap of radius \( \arccos(r) \) from
a point a distance \( \arccos(x) \) from NP. The first of these is
\[ \lim_{x \to -1} f(x, 1, r), \]
the second \( f(r', 1, r) \), where \( r' \) is slightly less than \( r \). Each of these has a useful interpretation.

Until now, the calculations have been done in two cases, \( p \) even and
\( p \) odd. The following proposition will unify them.

Proposition (5.5.15). If \( p \) is even

\[ \pi \left( \frac{p-1}{p-2} \right) !!! = 2^{p-2} \Gamma \left( \frac{p+1}{2} \right) / \Gamma \left( \frac{p}{2} \right). \]

If \( p \) is odd

\[ 2 \left( \frac{p-1}{p-2} \right) !!! = 2^{p-2} \Gamma \left( \frac{p+1}{2} \right) / \Gamma \left( \frac{p}{2} \right). \]

Proof. For \( N \) a positive integer
\[(2N)!! = 2^N N! = 2^N \Gamma(N+1) \text{ and } (2N-1)!! = \frac{\Gamma(N+\frac{1}{2})}{\Gamma(\frac{1}{2})} \cdot \]

Since \( \Gamma(\frac{1}{2}) = \sqrt{\pi} \), the result follows.

Consider \( f(0,r,r) \). For any \( x \in (0,r) \), \( f(0,r,r) = f(0,x,x) + f(x,r,r) \). Also, \( f(-x,r,r) = f(x,r,r) \). Thus it is clear that
\[ f(0,r,r) = \max_{-r \leq x \leq r} f(x,r,r) \cdot f(0,r,r) \text{ will be a maximal hitting time in the case where opposite points of } S_{p-1} \text{ are identified.} \]

Proposition (5.5.16).

\[
f(0,r,r) = \begin{cases} 
\frac{2\sqrt{\pi}}{\lambda (p-3)} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \left( \frac{1}{1-r^2} \right)^{\frac{p-3}{2}} (1+O(1-r^2)^{1/2}) & p \geq 4 \\
\frac{4}{\lambda} \log\left(\frac{1}{1-r^2}\right) + O(1) & p = 3.
\end{cases}
\]

as \( r \to 1 \).

Proof. Formula (5.5.8) and Proposition (5.5.15) together with the approximations in (5.5.9)-(5.5.14) give the result. For \( p \) even (5.5.8) is dominated by the term
\[
2\left(\frac{S(r)-S(0)}{S(r)-S(-r)}\right)S(-r)M(-r);
\]
other terms are an order of magnitude smaller. This is
\[
2\left(\frac{1}{p-3}\right)^{\frac{p-3}{2}} (1+O(1-r^2)^{1/2}) \left(-\frac{\pi}{\lambda} \frac{(p-1)!!}{(p-2)!!} + O\left(\frac{1}{1-r^2}\right)^{1/2}\right)
\]

104
which gives the result. For $p$ odd, the antisymmetry of $S$ and $M$
gives $f(0,r,r) = 2S(r)M(r) + o(1)$. (5.5.10) and (5.5.12) give the
result.

Next consider $f(x(r),r,r)$, where $x(r)$ is slightly less than
$r$, yet to be specified. This will be a minimal hitting time;

$$f(x(r),r,r) = \min_{|x| \leq x(r)} f(x,r,r).$$

It will turn out to be sufficient to consider $x(r) = \sqrt{1-(1-r^2)a^2(r)}$,
where $a(r) = -\log(\sqrt{1-r^2})$ for $p \geq 4$. For $p = 3$, this method of
bounding expected covering times fails somewhat. To see this
$f(x(r),r,r)$ will be calculated for $p = 3$ for $x^2(r) = 1-(1-r^2)a^2(r)$
where $a(r)$ is arbitrary.

Proposition (5.5.17). For $p \geq 4$, $x(r) = \sqrt{1-(1-r^2)\log^2(\sqrt{1-r^2})}$

$$f(x(r),r,r) = \frac{2\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma\left(\frac{b+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{p-3}{2} \left(\frac{1}{1-r^2}\right)^{\frac{p-3}{2}} (1+O(\frac{1}{\log(1-r^2)})) \text{ as } r \to 1.$$

For $p = 3$, $x(r) = (1-(1-r^2)a^2(r))^{1/2}$, $\frac{1}{1-r^2} > a^2(r) > 1$,

$$f(x(r),r,r) = \frac{4}{\lambda} \log[a^2(r)] + o(1) \text{ as } r \to 1.$$

Proof.

$$\frac{1}{1-x^2(r)} = \left(\frac{1}{1-r^2}\right)^{1/2} \cdot \frac{1}{a^2(r)},$$

105
so for \( p \geq 4 \),

\[
\frac{S(r) - S(x(r))}{S(r) - S(-r)} = \frac{\frac{r}{p-3} \left( \frac{1}{1-r^2} \right)^2 - \frac{x(r)}{p-3} \left( \frac{1}{1-r^2} \right)^2 \left( \frac{1}{a^2(r)} \right) \frac{p-3}{2}}{(1+0(1-x^2(r)))}.
\]

This is

\[
\frac{1}{2} \left( 1 - \frac{x(r)}{r} \left( \frac{1}{a^2(r)} \right)^2 \right)^{p-3} (1+0(1-x^2(r))).
\]

Using \( a(r) = -\log(1-r^2) \), this is

\[
\frac{1}{2} (1+0(\frac{1}{\log(1-r^2)})).
\]

The same calculations as Proposition (5.5.16) complete the proof for \( p \geq 4 \).

For \( p = 3 \), from (5.5.8) and the antisymmetry of \( S \) and \( M \),

\[
f(x(r), r, r) = 2[S(r)M(r) - I(r) - S(x(r))M(x(r)) - I(x(r))] =
\]

\[
2[\log(\frac{1}{1-r^2}) \frac{2}{\lambda} - \log(\frac{1}{1-x^2(r)}) \frac{2}{\lambda} + 0(1)]
\]

\[
= \frac{4}{\lambda} \left[ \log(\frac{1}{1-r^2}) - \log(\frac{1}{1-r^2}) + \log a^2(r) + 0(1) \right].
\]
Having completed the necessary calculations for the two cap problem, consider hitting times for the one cap problem. The maximal and minimal hitting times to be considered are, respectively, \( \lim_{x \to -1} f(x,1,r) \) (\( f(-1,1,r) \) for short), and \( f(r',1,r) \), where \( r' \) is slightly less than \( r \). Since \( f(x,1,r) \leq f(y,1,r) \) for \( x \geq y \), clearly

\[
f(-1,1,r) = \sup_x f(x,1,r)
\]

and

\[
f(r',1,r) = \inf_{x \leq r'} f(x,1,r).
\]

The calculations of \( f(-1,1,r) \) and \( f(r',1,r) \) do not fit into the two boundary problem whose solution is given by (5.5.8). They can be found from (5.5.8) by a limiting argument. As \( r_1 \to 1 \), clearly \( T_{x,r_1,r} \) converges to \( T_{s,1,r} \) a.s. Since the random variables \( T_{x,r_1,r} \) are increasing in \( r_1 \), the Monotone Convergence Theorem implies \( f(x,r_1,r) \) converges to \( f(x,1,r) \). Thus \( f(x,1,r) \) is given by \( \lim_{r_1 \to 1} f(x,r_1,r) \) for any \( x \leq r \). As \( r_1 \to 1 \) \( S(-r_1) \to -\infty \), so letting \( r_1 \to 1 \) in (5.5.9)

\[
f(x,1,r) = 2[S(r)M(r) - I(r) - S(x)M(x) + I(x) - M(-1)(S(r) - S(x))]
\]

(5.5.18)

\[
= 2[S(r)[M(r) - M(-1)] - S(x)[M(x) - M(-1)] - I(r) + I(x)].
\]

Consider the limit of (5.5.18) as \( x \to -1 \). From (5.5.13) and (5.5.14), \( I(x) \) remains bounded. As \( x \to -1 \), \( m(x) \) is decreasing, so
\[ |M(x)-M(-1)| \leq m(x)(1+x). \quad m(x) = \frac{p-1}{\lambda}(1-x^2)^{p-1} 3 \]

so from (5.5.9) and (5.5.10), \( S(x)[M(x)-M(-1)] \to 0 \) as \( x \to -1 \). Thus

(5.5.19) \quad f(-1,1,r) = 2[S(r)[M(r)-M(-1)]-I(r)+I(-1)] .

Proposition (5.5.20).

\[
f(-1,1,r) = \begin{cases} 
\frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma(p+1)}{\Gamma(-\frac{1}{2})(1-r^2)^{p-3}} \left(1+O((1-r^2)^{1/2})\right) & p \geq 4 \\
\frac{8}{\lambda} \log(\frac{1}{1-r^2}) + O(1) & p = 3.
\end{cases}
\]

Proof. Exactly like that of Proposition (5.5.16). Just insert (5.5.9)-(5.5.12) into (5.5.19).

Again let \( x(r) = (1-(1-r^2)a^2(r))^{1/2} \) where \( a(r) = -\log(1-r^2) \) for \( p \geq 4 \) and \( a(r) \) is arbitrary for \( p = 3 \).

Proposition (5.5.21). For \( p \geq 4 \), \( x(r) \) as above

\[
f(x(r),1,r) = \frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma(p+1)}{\Gamma(-\frac{1}{2})(1-r^2)^{p-3}} \left(1+O\left(\frac{1}{\log(1-r^2)}\right)\right) \quad \text{as} \quad r \to 1 .
\]

For \( p = 3 \), \( \frac{1}{1-r^2} > a^2(r) > 1 \), as \( r \to 1 \), \( f(x(r),1,r) = \frac{8}{\lambda} \log a^2(r) + O(1) \).

Proof. The proof again involves plugging (5.5.9)-(5.5.12) into (5.5.18). The details are simple and are omitted.
Summarizing, maximal and minimal expected hitting times have been calculated asymptotically for one and two cap problems in \( p \geq 3 \) dimensions. Propositions (5.5.16) and (5.5.17) give maximal and minimal expected hitting times for the two cap problem. Propositions (5.5.20) and (5.5.21) are the same results for the one cap problem. As might be expected, the numbers obtained in the latter two propositions are twice those obtained in the former two.

5.6 CONVERGENCE OF HITTING TIMES FOR RANDOM WALKS

This section discusses the convergence of expected hitting times for random walks to those of spherical Brownian motion as the step size of the random walk shrinks. There are four parts to the argument. First, a theorem of Skorokhod shows that the finite dimensional distributions of the one-dimensional Markov Processes \( X^1_j(i) \) converge to those of the diffusion \( W(\cdot) \), as \( s \to 0 \), as discussed in Section 5.4. Another theorem of Skorokhod shows that the time taken by \( X^1_s(\cdot) \) to leave an interval converges in distribution to the corresponding time for \( W(\cdot) \). The third step is to give a crude bound on the expected time taken by \( X^1_s(i) \) to leave an interval. This will imply that the expected times for \( X^1_s(i) \) to leave an interval converges to the expected times for \( W(\cdot) \), which were calculated in the last section. Finally, an argument is given that maximal and minimal expected times converge as well. Though stated for the one-dimensional processes \( X^1_s(i) \) and \( W(\cdot) \), these results have the alternative interpretation of the convergence of the maximal and minimal expected time taken by spherical random walks to hit caps to the corresponding quantities for spherical Brownian motion.
First consider the convergence of $X^1_s(s)$ of (5.5.5) to $W(\cdot)$, in the sense of finite dimensional distributions. This is a straightforward application of the theorem of Section 6.4 of Skorokhod (1982). Define $X^1_s(t)$ for all positive $t$ by $X^1_s(t) = X^1_s(is)$ for $is \leq t < (i+1)s$. Let $s_i = \frac{T}{i}$ for some fixed $T > 0$. The following list gives the relationship between the notation of Section 5.4 and that of Skorokhod:

**Section 5.4.**

<table>
<thead>
<tr>
<th>Skorokhod</th>
<th>Section 5.4.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi(n)$</td>
<td>$\xi(n)$</td>
</tr>
<tr>
<td>$\xi(n)(t)$</td>
<td>$\xi(n)(t)$</td>
</tr>
<tr>
<td>$a^{(n)}(t,x)$</td>
<td>$a^{(n)}(t,x)$</td>
</tr>
<tr>
<td>$b^{(n)}_1(t,x)$</td>
<td>$b^{(n)}_1(t,x)$</td>
</tr>
<tr>
<td>$f^{(n)}(t,x,u)$</td>
<td>$f^{(n)}(t,x,u)$</td>
</tr>
<tr>
<td>$a(t,x)$</td>
<td>$a(t,x)$</td>
</tr>
<tr>
<td>$b^{(n)}_1(t,x)$</td>
<td>$b^{(n)}_1(t,x)$</td>
</tr>
<tr>
<td>$f(t,x,u)$</td>
<td>$f(t,x,u)$</td>
</tr>
<tr>
<td>$X_0$(fixed)</td>
<td>$X_0$(fixed)</td>
</tr>
<tr>
<td>$\xi(n)(0)$</td>
<td>$\xi(n)(0)$</td>
</tr>
</tbody>
</table>
The quantities corresponding to \(a^{(n)}(t,x)\) and \(b^{(n)}(t,x)\) are obtained in the derivation of (5.4.5). Using these relationships, conditions a-d, e' and f of the theorem of Section 6.4 of Skorokhod are easily verified. This gives the proof for a particular sequence of values of \(s\) converging to zero. The proof can easily be extended to an arbitrary sequence of values of \(s\) decreasing to zero, and thus to all real \(s\), as \(s \to 0\), we well. Summarizing,

Proposition (5.6.1). For any \(T > 0\), \(X_s^1(0) = x = W(0)\) for all \(s\), the finite dimensional distributions of \(\{X_s^1(t), 0 \leq t \leq T\}\) converge to those of \(\{W(t), 0 \leq t \leq T\}\).

Next consider the time taken by \(X_s^1(t)\) to leave the interval \((-r_1, r_2)\) starting from \(X_s^1(0) = x\). Denote this time by \(T_{x,r_1,r_2}^s\), with \(T_{x,r_1,r_2}\) the corresponding time for \(W(\cdot)\). More abstractly, consider the path functional \(T_{x,r_1,r_2}^s\) which, for any sample path of \(X_s^1(\cdot)\) or \(W(\cdot)\) starting from \(x\) gives the first time the path leaves \((-r_1, r_2)\).

For any \(T > 0\), it is not hard to show that the functional \(\min(T, T_{x,r_1,r_2}^s)\) is what Skorokhod calls J-continuous with probability one on the sample paths of \(\{W(t), 0 \leq t \leq T\}\). The theorem of Section 6.5 of Skorokhod then immediately implies that \(\min(T, T_{x,r_1,r_2}^s)\) converges to \(\min(T, T_{x,r_1,r_2})\) in distribution as \(s \to 0\). This is true for any \(T > 0\), so the next proposition follows.

Proposition (5.6.2). \(T_{x,r_1,r_2}^s\) converges in distribution to \(T_{x,r_1,r_2}\) as \(s \to 0\).

The next goal is that \(E T_{x,r_1,r_2}^s \to E T_{x,r_1,r_2}\) as \(s \to 0\). The plan will
be to show that they are uniformly integrable which, in turn, will imply that the expectations converge.

Proposition (5.6.3). For \(-1 \leq r_1 < x < r_2 < 1\), \(\mathbf{E}^S_{x,r_1,r_2} + \mathbf{E}^T_{x,r_1,r_2}\) as \(s \to 0\).

Proof. Starting with \(W(0) = 1\), there exist \(q > 0\) and \(\delta > 0\) such that \(P(W(q) < -0.71) > 2\delta\). By Proposition (5.6.1) there exists \(s_0\) such that for \(s \geq s_0\) \(P(X_s^s(q) < -\sqrt{2}/2) > \delta\). This means that the associated spherical random walk \(X_s^s(\cdot)\) has probability at least \(\delta\) of being a geodesic distance \(\frac{3\pi}{4}\) or more from its starting point at time \(q\).

Therefore with probability at least \(\delta^2\), by time \(2q\) the random walk has moved a distance at least \(\frac{3\pi}{4}\) from its starting point \(x_0\) to some point \(x_1\), then a distance at least \(\frac{3\pi}{4}\) from \(x_1\).

Consider \(P(T^S_{x_0,r_1,r_2} < 2q)\). For \(s\) small enough and \(r_1, r_2\) fixed, this can now be bounded from below, uniformly in \(x\). Let NP be the North Pole and B a cap of radius \(r_2^* = \arccos(r_2)\) about it. Let \(x_0 \in S_{p-1}\) be arbitrary. Then it is sufficient to show that for \(s\) small enough, with probability \(\varepsilon > 0\) independent of \(x_0\), \(X_s(\cdot)\) enters \(B\) before time \(2q\). Either \(x_0\) or \(x_1\) as described above must be within a distance \(\frac{3\pi}{4}\) of NP. Thus for \(s\) small enough, with probability at least \(\delta^2\), \(X_s(\cdot)\) moves a distance \(\frac{3\pi}{4}\) from a point \(x_2\) no more than \(\frac{3\pi}{4}\) away from NP before time \(2q\). If the maximum step size of \(X_s(\cdot)\) is less than \(r_2^*\) and the distance from NP to \(x_2\) is \(\rho\), then \(X_s(\cdot)\) must hit a point \(x_3\) whose distance from \(x_2\) is between \(\rho - r_2^*\) and \(\rho + r_2^*\). By the spherical symmetry of \(X_s(\cdot)\), the probability that \(x_3\)
lies in \( B \) is bounded from below. As \( s \to 0 \), eventually the maximal step size of \( X_s(\cdot) \) will be less than \( r_2^* \). For \( s \) this small and \( s < s_0 \) from before \( P(T_{x_0,r_1,r_2}^S < 2q) \) is bounded below, uniformly in \( x_0 \).

As in Stein's Lemma, this shows that the tail of the distribution of \( T_{x_0,r_1,r_2}^S \) is exponentially small, uniformly for \( s \) small enough.

Thus \( \{T_{x_0,r_1,r_2}^S\} \), for such \( s \), are uniformly integrable and Theorem 4.5.2 of Chung (1974) implies

\[
\lim_{s \to 0} E_{x_0,r_1,r_2}^S = E_{x_0,r_1,r_2} as \ s \to 0.
\]

Finally consider maximal and minimal expected hitting times, as calculated for \( W(\cdot) \) in Section 5.5. To deal with covering times for random walks, it is necessary to show that maximal and minimal expected hitting times for random walks converge to those for \( W(\cdot) \). This appears obvious, but deserves a proof. Let \( d(x,r_1,r_2) = \min(|x-r_2|,|x+r_1|) \).

Proposition (5.6.4).

\[
\sup_{x \in [-r_1,r_2]} E_{x,r_1,r_2}^S \to \sup_{x \in [-r_1,r_2]} E_{x,r_1,r_2} \quad \text{and for all} \quad \varepsilon > 0
\]

\[
\inf_{d(x,r_1,r_2) > \varepsilon} E_{x,r_1,r_2}^S \to \inf_{d(x,r_1,r_2) > \varepsilon} E_{x,r_1,r_2} \text{ as } s \to 0.
\]
Proof. The first assertion will be proven, the second's proof is similar. The right side is the supremum of a continuous function on a compact interval, so assume the supremum is attained at some point $x^+ \in [-r_1, r_2]$. Suppose

$$\lim_{s \to 0} \sup_{x \in [-r_1, r_2]} ET^s_{x, r_1, r_2} < ET_{x^+, r_1, r_2}.$$ 

This contradicts pointwise convergence at $x^+$. Suppose that for some $\rho > 0$

$$\lim_{s \to 0} \sup_{x \in [-r_1, r_2]} ET^s_{x, r_1, r_2} = ET_{x^+, r_1, r_2} + 2\rho.$$ 

Then there is a sequence of pairs $(x_i, s_i)$ with $x_i \in [-r_1, r_2]$, $s_i \to 0$ such that $ET^s_{x_i, r_1, r_2} > ET^s_{x, r_1, r_2} + \rho$ for all $i$. Since $[-r_1, r_2]$ is compact the set $\{x_i\}$ has a limit point $x_\infty$. Without loss of generality $x_i \to x_\infty$. The following Lemma is needed.

Lemma (5.6.5). Suppose $|\arccos(x) - \arccos(y)| < \beta$. Then with

$$\arccos(L_1) = \arccos(r_1) + \beta \quad \arccos(U_1) = \arccos(r_1) - \beta$$

$$\arccos(L_2) = \arccos(r_2) + \beta \quad \arccos(U_2) = \arccos(r_2) - \beta$$

$$ET^s_{y, L_1, L_2} \leq ET^s_{x, r_1, r_2} \leq ET^s_{y, U_1, U_2}$$

for any $s$. Proof of Lemma. Only the left inequality will be proven; the other proof is identical. Consider caps of radii $\arccos(r_1)$ about $SP$ and
arccos\(r_2\) about NP. Let \(X_s(0) = (x, \sqrt{1-x^2}, 0, \ldots, 0)\) and \(Y_s(0) = (y, \sqrt{1-y^2}, 0, \ldots, 0)\). \(X_s(0) = RY_s(0)\), where \(R\) is a rotation of angle \(\arccos(x) - \arccos(y)\) in the \(e_1, e_2\) plane. The time taken to hit caps about NP and SP from \(Y_s(0)\) has the same distribution as the time taken to hit caps of the same radii about \(R(NP)\) and \(R(SP)\) from \(X^S(0)\). The caps of radii \(\arccos(r_1) + \beta\) and \(\arccos(r_2) + \beta\) about SP and NP contain the caps of radii \(\arccos(r_1)\) and \(\arccos(r_2)\) about \(R(SP)\) and \(R(NP)\), respectively. Thus \(T_{y, L_1, L_2}^S < T_{x, r_1, r_2}^S\), proving the Lemma.

Returning to the proof of the proposition, \(ET_{x, r_1, r_2}\) is continuous in \(x, r_1,\) and \(r_2\), so there exists \(\beta > 0\) such that if

\[|\arccos(x_i) - \arccos(x_\infty)| < \beta,\]

then

\[|ET_{x_i, r_1, r_2} - ET_{x_\infty, r_1, r_2}| < \frac{\rho}{3},\]

and

\[|ET_{x_\infty, r_1, r_2} - ET_{x_\infty, U_1, U_2}| < \frac{\rho}{3},\]

where \(U_1\) and \(U_2\) are as in Lemma (5.6.5). Thus

\[ET_{x_i, r_1, r_2} > ET_{x_\infty, U_1, U_2} + \frac{\rho}{3}\]

for \(x_i\) near \(x_\infty\). But for \(i\) large enough,

\[|\arccos(x_i) - \arccos(x_\infty)| < \beta,\]

implying

115
\[
\begin{align*}
&\frac{s_1}{\text{ET}_{x_1, r_1, r_2} < \text{ET}_{x_\infty, U_1, U_2}}.
\end{align*}
\]

But

\[
\begin{align*}
&\frac{s_1}{\text{ET}_{x_\infty, U_1, U_2} \rightarrow \text{ET}_{x_\infty, U_1, U_2}}.
\end{align*}
\]

giving a contradiction and completing the proof. 

CHAPTER 6
SPHERE PROBLEMS

6.1 INTRODUCTION

This chapter gives results on mean covering times for random walks or Brownian motion on $S_{p-1}$, the unit sphere in $\mathbb{R}^p$. For either process, $C_\varepsilon$ denotes the first time caps of geodesic radius $\varepsilon$ about the points visited by the process cover $S_{p-1}$. Some results are given about $EC_\varepsilon$ as $\varepsilon \to 0$ in each case. Motivated by the Grand Tour, the two cap problem is also considered. Let $C_\varepsilon^*$ be the first time the caps of radius $\varepsilon$ about the points visited by the process and the reflections of these points in the origin cover $S_{p-1}$. Results for $EC_\varepsilon^*$ analogous to those for $EC_\varepsilon$ will be given.

Section 6.2 will give bounds on $EC_\varepsilon$ and $EC_\varepsilon^*$ for Brownian motion. Section 6.3 discusses bounds for $EC_\varepsilon$ and $EC_\varepsilon^*$ for random walks. Due to the lack of a solution to the hitting time problem for random walks, the results are not as good as those for Brownian motion. Bounds on $EC_\varepsilon$ and $EC_\varepsilon^*$ can only be given asymptotically, as the step size of the random walk shrinks and time is rescaled so that the random walk approaches Brownian motion. Further possible results will be discussed in Chapter 7.

6.1 COVERING TIMES FOR BROWNIAN MOTION

Let $W(\cdot)$ be Brownian motion on $S_{p-1}$ with parameter $\lambda$. Define $C_\varepsilon$ as the smallest time $t$ such that for every point $x$ on $S_{p-1}$, for some
s \leq t, W(s) is within a geodesic distance $\varepsilon$ of x. Similarly $C^*_\varepsilon$ is the smallest time $t$ such that for every point $x$ on $S^p_{p-1}$, for some $s \leq t, \bar{W}(s)$ or $-\bar{W}(s)$ is within a geodesic distance $\varepsilon$ of x. This is equivalent to the definition given in terms of caps in Section 6.1. In this section, upper and lower bounds will be given for $EC_\varepsilon$ and $EC^*_\varepsilon$ that are asymptotically tight as $\varepsilon \to 0$ in $p > 3$ dimensions. Proofs will only be given for $EC_\varepsilon$; proofs for $EC^*_\varepsilon$ are identical. Bounds will be obtained by considering the time to visit a finite set of caps. For an upper bound, a set of caps will be presented with the property that if $\bar{W}(\cdot)$ has visited all these caps, it has been within $\varepsilon$ of all points of $S^p_{p-1}$. The expected time to visit all caps of this set furnishes an upper bound on $EC_\varepsilon$. Similarly, another set of caps will be found with the property that if $\bar{W}(\cdot)$ has not visited every cap of the set, then it has not been within $\varepsilon$ of all points of $S^p_{p-1}$. The expected time to visit all caps of this set furnishes a lower bound for $EC_\varepsilon$. The analogous constructions for $EC^*_\varepsilon$ would involve finding sets of oppositely situated pairs of caps. The bounds will be presented in asymptotic terms only. However, from the proofs and the results of Chapter 5, it will be clear how actual bounds could be obtained.

Theorem (6.2.1). As $\varepsilon \to 0$

$$EC_\varepsilon = \frac{4\sqrt{\pi}}{\lambda} \frac{\Gamma \left( \frac{p+1}{2} \right)}{\Gamma \left( \frac{p}{2} \right)} \frac{p-1}{p-3} \frac{\log \left( \frac{1}{\varepsilon} \right)}{\varepsilon^{p-3}} \left( 1 + O \left( \frac{\log \log \left( \frac{1}{\varepsilon} \right)}{\log \left( \frac{1}{\varepsilon} \right)} \right) \right)$$

for $p \geq 4$
\[ \frac{1}{2} \leq \lim \frac{\lambda \mathcal{E}_\epsilon}{8\log^2 \left( \frac{1}{\epsilon} \right)} \leq \lim \frac{\lambda \mathcal{E}_\epsilon}{8\log^2 \left( \frac{1}{\epsilon} \right)} \leq 2 \quad \text{for} \quad p = 3 \]

\[ \mathcal{E}_\epsilon^* = \frac{2\sqrt{\pi}}{\lambda} \frac{\Gamma \left( \frac{p+1}{2} \right)}{\Gamma \left( \frac{p}{2} \right)} \frac{p-1}{p-3} \frac{\log \left( \frac{1}{\epsilon} \right)}{\epsilon^{p-3}} \left( 1 + O \left( \frac{\log \log \left( \frac{1}{\epsilon} \right)}{\log \left( \frac{1}{\epsilon} \right)} \right) \right) \quad \text{for} \quad p \geq 4 \]

\[ \frac{1}{2} \leq \lim \frac{\lambda \mathcal{E}_\epsilon^*}{4\log^2 \left( \frac{1}{\epsilon} \right)} \leq \lim \frac{\lambda \mathcal{E}_\epsilon^*}{4\log^2 \left( \frac{1}{\epsilon} \right)} \leq 2 \quad \text{for} \quad p = 3 . \]

Proof. First consider a lower bound on \( \mathcal{E}_\epsilon \) for \( p \geq 4 \). By Proposition (5.3.6) there are at least

\[ L(p) \left( \epsilon + \epsilon \log \left( \frac{1}{\epsilon} \right) \right)^{-(p-1)} \]

disjoint caps of radius \( \epsilon + \epsilon \log \left( \frac{1}{\epsilon} \right) \) on \( S_{p-1} \). Consider a concentric cap of radius \( \epsilon \) inside each of these caps. There are

\[ L(p) \left( \frac{\epsilon}{\epsilon} \right)^{p-1} \left( 1 + \log \left( \frac{\epsilon}{\epsilon} \right) \right)^{-(p-1)} \]

such caps. Choosing \( x \) and \( y \) in distinct caps, the distance from \( x \) to \( y \) is at least \( 2\epsilon \log \left( \frac{1}{\epsilon} \right) \). The distance from \( x \) to the center of the cap containing \( y \) is at least \( \epsilon + 2\epsilon \log \left( \frac{1}{\epsilon} \right) \). Thus the expected time, starting from \( x \), to hit the cap containing \( y \) is at least \( f(\cos(\epsilon + 2\epsilon \log \left( \frac{1}{\epsilon} \right)), 1, \cos \epsilon) \).

Let \( r = \cos \epsilon \) and

\[ x(r) = \left( 1 - (1 - r^2) \log \sqrt{1 - r^2} \right)^{1/2} \]

For \( \epsilon \) reasonably small, \( \cos(\epsilon + 2\epsilon \log \left( \frac{1}{\epsilon} \right)) < x(\cos \epsilon) \) so
f(\cos(\varepsilon+2\varepsilon\log(\frac{1}{\varepsilon})), 1, \cos\varepsilon) > f(x(\cos\varepsilon), 1, \cos\varepsilon). f(x(\cos\varepsilon), 1, \cos\varepsilon) is

given by Proposition (5.5.21) as

\[ (6.2.3) \quad \frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{1-\cos\varepsilon} \left(1+\frac{1}{\log(1-\cos\varepsilon)}\right) \] as \( \varepsilon \to 0 \)

\[ = \frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\varepsilon^{p-3}} \left(1+\frac{1}{\log(\frac{1}{\varepsilon})}\right). \]

Returning to Section 2.3, (6.2.3) gives the minimal expected time to

hit one cap from inside another. (6.2.2) gives a lower bound on the number of
caps. To comply exactly with the conditions of Theorem (2.3.1) an initial
position \( x_0 \) should be chosen inside one of the caps and (6.2.2) decreased
by one, but this does not affect the result. Combining (6.2.2), (6.2.3)
with the second assertion of Theorem (2.3.1),

\[ \text{EC}_\varepsilon \geq \frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\varepsilon^{p-3}} \left(1+\frac{1}{\log(\frac{1}{\varepsilon})}\right) \log\left[\frac{\text{L}(p)}{\varepsilon^{p-1}(1+\log(\frac{1}{\varepsilon}))}\right] \]

\[ = \frac{4\sqrt{\pi}}{\lambda} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{\log(\frac{1}{\varepsilon})}{\varepsilon^{p-3}} \left(1+\frac{\log\log(\frac{1}{\varepsilon})}{\log(\frac{1}{\varepsilon})}\right). \]

Next consider an upper bound on \( \text{EC}_\varepsilon \) for \( p \geq 4 \). Note that the upper

bound in Theorem (2.3.1) does not require disjoint caps. By Proposition

(5.3.6) there is a set of at most

\[ (6.2.4) \quad U(p)[\frac{\varepsilon}{\log(\frac{1}{\varepsilon})}]^{-(p-1)} \]
caps of radius $\epsilon/\log(\frac{1}{\epsilon})$ that cover $S_{p-1}$. Consider a concentric cap of radius
\[ \epsilon(1 - \frac{1}{\log(\frac{1}{\epsilon})}) \]
about the center of each of these caps. If $W(\cdot)$ visits this larger cap then it is within $\epsilon$ of every point of the smaller cap. Since the small caps cover $S_{p-1}$, if $W(\cdot)$ has visited all the larger caps, it has been within $\epsilon$ of all points of $S_{p-1}$.

The maximal expected time taken to hit a cap of this radius is given by Proposition (5.5.20) as
\[ f(-1,1,\cos(\epsilon(1 - \frac{1}{\log(\frac{1}{\epsilon})})), \]

This is
\[ \frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\epsilon^{p-3}} \left(\frac{1+O\left(\frac{1}{\log(\frac{1}{\epsilon})}\right)}{}\right). \]

(6.2.5) \[ \frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\epsilon^{p-3}} \left(\frac{1+O\left(\frac{1}{\log(\frac{1}{\epsilon})}\right)}{}\right). \]

(6.2.4), (6.2.5) and the first assertion of Theorem (2.3.1) give
\[ EC_{\epsilon} \leq \frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\epsilon^{p-3}} \left(1+O\left(\frac{1}{\log(\frac{1}{\epsilon})}\right)\right) \log \left[ \frac{U(p)\log \frac{1}{\epsilon}}{p-1} \right] \]
\[ = \frac{4\sqrt{\pi}}{\lambda(p-3)} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\epsilon^{p-3}} \left(1+O\left(\frac{1}{\log(\frac{1}{\epsilon})}\right)\right) \loglog \left(\frac{1}{\epsilon}\right). \]

This completes the proof of the first assertion.

For $p=3$, the same sort of proof is used. The best lower bound obtainable by this technique is to choose caps of radius $\epsilon + \frac{1}{2} \sqrt{\epsilon}$,
then consider concentric caps of radius $\varepsilon$ inside them. This leads to the lower bound

$$EC_\varepsilon \geq f(\cos(\varepsilon+\sqrt{\varepsilon}),1,\cos \varepsilon) \log [L(2)(\frac{1}{\sqrt{\varepsilon}+\varepsilon})^2].$$

With $r = \cos \varepsilon$, $x(r)$ in Proposition (5.5.21) is $\cos(\varepsilon+\sqrt{\varepsilon})$, which corresponds to $a(r) = \sin(\varepsilon+\sqrt{\varepsilon})/\sin(\varepsilon)$.

$$\log \left( \frac{\sin(\varepsilon+\sqrt{\varepsilon})}{\sin \varepsilon} \right) = \log \left( \frac{1}{\sqrt{\varepsilon}} \right) + O(1),$$

so

$$f(\cos(\varepsilon+\sqrt{\varepsilon}),1,\cos \varepsilon) = \frac{8}{\lambda} \cdot \frac{1}{2} \log \left( \frac{1}{\varepsilon} \right) + O(1).$$

This gives a lower bound of

$$EC_\varepsilon \geq 4 \log^2 \left( \frac{1}{\varepsilon} \right) + O(1).$$

This is the largest lower bound obtainable by this method.

For an upper bound, the same argument as for the case $p > 4$ is used. Choose caps of radius $\varepsilon/\log(\frac{1}{\varepsilon})$ to cover $S_2$ and consider the time taken by $\tilde{W}(\cdot)$ to visit all the concentric caps of radius $\varepsilon(1-1/\log(\frac{1}{\varepsilon}))$. The upper bound is, from Theorem (2.3.1), Proposition (5.3.6) and (5.5.20),

122
(6.2.7) \[ EC_\varepsilon \leq \frac{8}{\lambda} \log \left( \frac{1}{\sin(\varepsilon)} \right) + O(1) \log \left[ U(2) \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \right]^2 \]

\[ = \frac{16}{\lambda} \log^2 \left( \frac{1}{\varepsilon} \right) + O(\log \frac{1}{\varepsilon} \loglog \frac{1}{\varepsilon}) . \]

(6.2.6) and (6.2.7) prove the second assertion of the theorem. The last two assertions have proofs identical to the first two.

The first assertion of Theorem (6.2.1) is quite similar to the result (1.2) of Flatto and Newman (1977) mentioned in the Introduction. The leading terms of the two expressions are different, but the presence of a \( \log \log \left( \frac{1}{\varepsilon} \right) \) term in (1.2) and the fact that the next term of (6.2.1) is \( O(\log \log \left( \frac{1}{\varepsilon} \right)) \) suggest that the next term in \( EC_\varepsilon \) may be a \( \log \log \left( \frac{1}{\varepsilon} \right) \) term. The present methods do not seem strong enough to answer this.

6.3 ASYMPTOTIC RESULTS FOR RANDOM WALKS

Expected hitting times have not been calculated or even bounded for random walks on \( S_{p-1} \). This appears to be difficult. The results of Section 5.6 show, however, that properly normalized, expected hitting times for random walks converge to those for Brownian motion. This will imply that bounds on expected covering times converge as well.

Let \( \{X_s(\cdot), 0 < s \leq 1\} \) be a family of random walks on \( S_{p-1} \) as described in Section 5.4. Recall that \( X_s(\cdot) \) takes a step every \( s \) time units. Let \( C_\varepsilon(s), C_\varepsilon^*(s) \) be expected covering times for \( X_s(\cdot) \) analogous to \( C_\varepsilon \) and \( C_\varepsilon^* \) for Brownian motion \( \tilde{W}(\cdot) \).
Theorem (6.3.1).

\[ 1+0^- \left( \frac{\log \log \left( \frac{1}{\varepsilon} \right)}{\log \left( \frac{1}{\varepsilon} \right)} \right) \leq \lim_{s \to 0} \frac{\lambda E_{\varepsilon}(s)}{4\sqrt{\pi}} \frac{\Gamma \left( \frac{P}{2} \right)}{\Gamma \left( \frac{P+1}{2} \right)} \frac{p-3}{p-1} \frac{\varepsilon^{p-3}}{\log \left( \frac{1}{\varepsilon} \right)} \]

and

\[ 1+0^+ \left( \frac{\log \log \left( \frac{1}{\varepsilon} \right)}{\log \left( \frac{1}{\varepsilon} \right)} \right) \geq \lim_{s \to 0} \frac{\lambda E_{\varepsilon}(s)}{4\sqrt{\pi}} \frac{\Gamma \left( \frac{P}{2} \right)}{\Gamma \left( \frac{P+1}{2} \right)} \frac{p-3}{p-1} \frac{\varepsilon^{p-3}}{\log \left( \frac{1}{\varepsilon} \right)} \]

for \( p \geq 4 \).

\[ 0^- \left( \frac{\log \log \left( \frac{1}{\varepsilon} \right)}{\log \left( \frac{1}{\varepsilon} \right)} \right) \quad \text{and} \quad 0^+ \left( \frac{\log \log \left( \frac{1}{\varepsilon} \right)}{\log \left( \frac{1}{\varepsilon} \right)} \right) \]

stand for the bondafide upper bounds for \( W(\cdot) \) that could have been obtained in Theorem (6.2.1), but were too messy. Similar results hold for \( p = 3 \) and \( C^*(s) \).

Proof. This is an immediate consequence of Theorem (2.3.1), the maximal and minimal expected hitting times for \( W(\cdot) \) calculated in Section 5.5, and Proposition (5.6.4) on the convergence of maximal and minimal expected hitting times for random walks to those for \( W(\cdot) \).
CHAPTER 7
CONCLUDING REMARKS

This thesis gives some results in a fairly unexplored area. There has been a good deal of research on covering problems when caps are chosen independently. For example see Solomon (1978) and its bibliography, Flatto and Newman (1977), or Janson (1984). However, other than Aldous (1983a), I know of no other work on covering problems when caps are chosen in a dependent manner.

Chapter 2 of this thesis gives some bounds applicable to covering times when the problem has a Markov structure. The bounds are fairly crude, though in the examples presented they often give asymptotically tight results. These bounds can be improved; an example of this is Section 4.3 on random walks on $\mathbb{Z}_2^k$. Many things remain to be said in the area of general bounds. What are reasonable general conditions when the bounds can be expected to be good? Here some notion of rapid mixing will probably be important. What are similar bounds for the time to visit, or come close to, every state $d > 1$ times? This problem has received a lot of consideration in the independent cases; see the references above for details.

Chapter 3 and 4 deal with covering problems for finite groups. There are many interesting problems here. The first is a relationship between the tightness of bounds and the eigenvalues of the transition matrix. Intuitively, small eigenvalues, tight bounds and rapid mixing
occur together. A precise formulation of this would be nice. All the calculations done for random walks on finite groups could in theory be done for arbitrary finite irreducible Markov chains. They would probably be much harder. What is lost in going to Markov chains is the invariance of the whole problem under translations by group elements. Explicitly and implicitly, it made many calculations much more tractable. Another question for finite group problems is the one considered for sphere problems. Namely, how long does it take a random walk to get close (in some sense) to every element of a finite group. This appears to be harder than the problem of visiting every element. Finally, all the examples of Chapter 4 involved symmetric random walks. This made their eigenvalues real, hence easier to deal with. Some examples of covering problems for nonsymmetric random walks would be interesting.

Chapters 5 and 6 deal with covering problems on spheres. Here the results are few; there are many things to do. The first is good bounds on expected hitting times for random walks on spheres. The second is to say something about processes in $G_{2p}$, the set of planes in $\mathbb{R}^p$. Finally there are questions about higher moments and generating functions. In approximating covering problems by the time to visit a set of caps, the number of caps in the set enters the answer logarithmically. This allows a great deal of room for error while still giving asymptotically tight bounds. For generating functions the number of caps will enter in a more important fashion and the crude techniques used for mean covering times will not suffice. There are numerous other possible problems in this area.
In conclusion, some significant progress has been made on covering problems for random walks. However, the surface is barely scratched; there are a great deal of problems yet to do.
REFERENCES


