Polynomial birth-death approximation of pattern occurrences in an independent, identically distributed sequence

Terence Jegaraj

For the degree of
Master of Science by research
Department of Mathematics and Statistics
University of Melbourne

October 2003
Abstract

The distribution of occurrences of a non–overlapping pattern in an independent, identically distributed sequence can be approximated well by some two–parameter polynomial birth–death distributions, at least when the pattern probability is small. The total variation distance error can be shown to approach zero as the length of the sequence being considered increases.
Acknowledgements

I am grateful to Aihua Xia for his advice, patience and good humour even during the weeks and months when there were no results. I am also grateful to Kostya Borovkov for guidance when I first started.
Declaration

The thesis comprises only my original work, except where due acknowledgement has been made in the preface and the text to all other material used.
Contents

1 Introduction 5

2 Normal approximation for patterns in a string 12

3 PBD approximation and Poisson approximation 17
   3.1 Stein method for approximating with birth–death distributions 17
   3.2 Poisson approximation 19
   3.3 Poisson approximation for patterns in a string 21

4 Two–parameter PBD approximation for patterns in a string 25
   4.1 Patterns in a circle 25
   4.2 Approximation for patterns in a circle 31

5 Another two–parameter PBD approximation 40
   5.1 Approximation with a binomial–like distribution 40

6 Numerical data for patterns of length four 49

7 Bounds for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$ 53

8 Conclusion 66

A Non-overlapping patterns in $S'$ 68
A.1 Fraction of non-overlapping patterns 68
A.2 Possible values of pattern probability $p$ for a non-overlapping pattern 69

B Bound for $\|\Delta f\|$ 70

C Unimodality of $L(W_n)$ 72

D Proof of Theorem 10 77

E Proofs of results for $l > 2$ 88
Chapter 1

Introduction

In this chapter we describe the problem that is the focus of this essay and define some notation that is used throughout the essay. We then describe the results contained in each of the following chapters and outline the Stein method for approximating distributions, in particular distributions on the non–negative integers, $\mathbb{Z}^+ = \{0, 1, 2, \ldots \}$.

We firstly need some definitions. Let $S$ be a finite set of symbols and consider the independent sequence of identically distributed (i.i.d.) random elements taking values in $S$:

$$Z_i : (\Omega, \mathcal{F}, P) \rightarrow (S, 2^S), \ i = 1, 2, \ldots$$

Here $2^S$ is the $\sigma$–algebra of all subsets of $S$. We want to approximate the distribution of the number of occurrences of a given non–overlapping pattern in the string $(Z_1, Z_2, \ldots, Z_{n+l-1})$, where $n \geq 1$. Let the given pattern, of length $l$, be $(a_1, \ldots, a_l) \in S^l$. This pattern is non–overlapping, by which we mean that one occurrence of the pattern cannot overlap another occurrence: there is no $d$ such that

$$1 \leq d < l \text{ and } (a_1+\ldots+d, \ldots, a_l) = (a_1, \ldots, a_{l-d}).$$

No matter what the length $l$, the fraction of patterns in $S^l$ that are non–overlapping
is greater than \( \frac{|S|-2}{|S|-1} \) (see Appendix A). Hence when \( S = \{a, b, c, \ldots, z\} \), more than 24/25 of patterns of any given length are non-overlapping.

Define the sequence of pattern indicators:

\[
J_i = I_{\{ (a_1, \ldots, a_l) \}}(Z_i, \ldots, Z_{i+l-1}), \quad i \geq 1.
\]

\( J_k \) has value 1 when there is a pattern in the sequence \( (Z_i) \) starting at the \( k \)th term \( Z_k \). The probability of a pattern occurrence starting at \( Z_k \) is

\[
p := E[J_k].
\]

Define the sequence of partial sums of pattern indicators:

\[
W_n = \sum_{i=1}^{n} J_i, \quad n \geq 1.
\]

\( W_n \) counts the number of pattern occurrences in the string \( (Z_1, \ldots, Z_{n+l-1}) \).

We need a measure of the error between the distribution of \( W_n \) and our chosen approximating distribution. The total variation distance between two distributions on \( \mathbb{Z}_+ \), \( \mu \) and \( \nu \), is defined by

\[
d_{TV}(\mu, \nu) := \sup_{A \subseteq \mathbb{Z}_+} |\mu(A) - \nu(A)|
\]

\[
= 1/2 \sum_{k=0}^{\infty} |\mu\{k\} - \nu\{k\}|.
\]

We write \( \mathcal{L}(W_n) \) for the distribution of \( W_n \). \( \mathcal{L}(W_n) \) can be computed by means of the difference equation in Lemma 3, however for large values of \( n \) this is time consuming. In practice we are interested in approximating distributions that are difficult to compute. Approximating \( \mathcal{L}(W_n) \) means taking a distribution, \( \pi \), of a standard type, with chosen parameters, and bounding the error between it and \( \mathcal{L}(W_n) \). If \( \pi \) is a distribution on \( \mathbb{Z}_+ \) then the error is measured by \( d_{TV}(\mathcal{L}(W_n), \pi) \).
The primary reason we are approximating $\mathcal{L}(W_n)$ is to illustrate the technique of Stein approximation with some so-called polynomial birth-death (PBD) approximating distributions. A PBD distribution is the invariant distribution of a birth-death process whose birth and death rates are polynomial functions (see Brown and Xia [5]). The sequence of pattern indicators $(J_n)$ has such a simple dependence structure that often we can use methods developed for sums of independent indicators, with small modifications.

In chapter 2 we briefly consider approximation with a normal distribution, not a PBD distribution. We show that as $n$ goes to infinity $\frac{W_n - E[W_n]}{\sqrt{\text{Var}[W_n]/2}}$ converges in distribution to a standard normal random variable. A result of Rinott [10] gives us

$$\sup_{w \in \mathbb{R}} |F_n(w) - \Phi(w)| = O(n^{-1/2}), \quad (1.1)$$

where $F_n$ is the distribution function of $\frac{W_n - E[W_n]}{\sqrt{\text{Var}[W_n]/2}}$ and $\Phi$ is the standard normal distribution function. Computational work shows that the error on the left of equation (1.1) is roughly proportional to $n^{-1/2}$ but that the rate of convergence is rather slow. To get reasonably small errors requires values of $n$ several orders of magnitude larger than those required when approximating with suitable PBD distributions. The error in normal approximation increases as pattern probability $p$ decreases, which is the opposite of what happens in PBD approximation. Our error bound, derived from Rinott’s result, badly over-estimates the size of the error. Our bound varies with $p$ approximately like $p^{-3/2}$ for small $p$, while computation suggests the error varies more like $p^{-1/2}$.

From chapter 3 onwards we approximate $\mathcal{L}(W_n)$ with PBD distributions, estimating the error measured by total variation distance.

In chapter 3 we approximate $\mathcal{L}(W_n)$ with a Poisson distribution having the same mean. Results from Lindvall [9] and Barbour, Holst and Janson [2] are used and we
exploit the fact that the pattern indicators are negatively related. Upper and lower bounds for $d_{TV}(\mathcal{L}(W_n), \text{Poi}(E[W_n]))$ are given. These show that as $n$ increases, $d_{TV}(\mathcal{L}(W_n), \text{Poi}(E[W_n]))$ tends to move to between $\frac{1}{32}(2l - 1)p$ and $(2l - 1)p$.

In chapter 4 we approximate $\mathcal{L}(W_n)$ with a two–parameter PBD distribution, $\pi_1(\alpha, \beta)$, of a type used by Brown and Xia [5] for a sum of independent indicators. The indicators $J_i$ are not independent and we look to simplify calculations by actually choosing $\alpha$ and $\beta$ so that $\pi_1(\alpha, \beta)$ approximates the distribution of the number of pattern occurrences in the circle formed by joining the ends of the string $Z_1, \ldots, Z_n$. We also get a bound for the total variation distance between the latter distribution and $\mathcal{L}(W_n)$ that is small compared to our bound for $d_{TV}(\mathcal{L}(W_n), \pi_1(\alpha, \beta))$. The main results are Theorems 4 and 5 which are new. The procedure for choosing values of $\alpha$ and $\beta$ in Theorem 5 is essentially that used by Brown and Xia [5] for approximating a sum of independent indicators. What is new here is the application and the additional manipulations needed to accommodate it. Combining Theorems 4 and 5 gives a bound for $d_{TV}(\mathcal{L}(W_n), \pi_1(\alpha, \beta))$ that is roughly $12p^2l^2d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$.

In chapter 5 we approximate $\mathcal{L}(W_n)$ with a two–parameter PBD distribution, $\pi_2(r, q)$, that is a generalisation of a binomial distribution. Again the procedure used by Brown and Xia [5] for approximating a sum of independent indicators is the template for choosing values of the parameters $r$ and $q$. However the application and the approximating distribution are new. The main result is Theorem 8. Theorem 8 and Theorem 4 together give a bound for $d_{TV}(\mathcal{L}(W_n), \pi_2(r, q))$ that is roughly $2p^2l^2d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$.

In chapter 6 we present some data showing that the error of approximating $\mathcal{L}(W_n)$ with $\pi_2(r, q)$ is significantly smaller than the error of approximating with $\pi_1(\alpha, \beta)$. Both $\pi_1(\alpha, \beta)$ and $\pi_2(r, q)$ give much smaller errors than the Poisson approximation.

In chapter 7 we present bounds for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$ that are of order
$O(n^{-1/2})$, derived using the characteristic function of $W_n$. This follows the idea of Barbour and Jensen [3] for a sum of independent indicators. In our problem the indicators $J_i$ are not independent so we derive a difference equation satisfied by the characteristic functions and attempt to bound the solution. The bounds for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$ in Theorems 10 and 11 are new. For $l > 2$ our approach yields results only for $p < (l - 1)/l^l$. Computation suggests this restriction on $p$ is a shortcoming of the method of proof, rather than reflecting the behaviour of $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$. Thus at least for small values of $p$ the errors of approximating $\mathcal{L}(W_n)$ with $\pi_1(\alpha, \beta)$ and $\pi_2(r, q)$ go to zero like $O(n^{-1/2})$ as $n$ increases.

Most of the proofs of results in chapter 7 are in Appendix D and Appendix E.

Whether we approximate $\mathcal{L}(W_n)$ with normal distribution, Poisson distribution or the two–parameter PBD distributions $\pi_1(\alpha, \beta)$ or $\pi_2(r, q)$, in each case we use the Stein method to bound the error. The Stein method was discovered by Charles Stein, for normal approximating distribution (see citation in [5], for example). For a given approximating distribution, $\pi$, we first find a difference operator (or differential operator if $\pi$ is normal), $G$, that characterises $\pi$: we have

$$E[(Gf)(X)] = 0$$

for all $f$ in an appropriate class of functions if and only if $X$ has distribution $\pi$. Then the Stein equation

$$(Gf)(x) = I_A(x) - \pi(A), \ x \in \mathbb{Z}_+ \ (x \in \mathbb{R}, \text{ if } \pi \text{ is normal}) \quad (1.2)$$

is solved for $f$. We then find bounds for the solution, $f$, and its first difference, $\Delta f$, (or first derivative if $\pi$ is normal) that do not depend on the event $A$. Let $W$ be the sum of indicators whose distribution we want to approximate with $\pi$. Substituting $W$ in (1.2) and taking the expectation gives

$$P\{W \in A\} - \pi(A) = E[(Gf)(W)].$$
Finally we estimate \( E[(Gf)(W)] \) using the bounds we derived for \( f \) and \( \Delta f \)—since these bounds do not depend on \( A \) we will have an estimate for \( d_{TV}(\mathcal{L}(W), \pi) \).

Shortly after the discovery of the Stein method for normal approximation, Chen [6] extended the method to Poisson approximating distribution. The method was extended to other approximating distributions in a somewhat ad hoc manner until Barbour, Holst and Janson [2] observed that the operator \( G \) for an approximating distribution is related to the generator of a Markov process having this invariant distribution.

Recently, Brown and Xia [5] showed that invariant distributions of birth–death processes whose birth and death rates are polynomial functions are particularly convenient to use as approximating distributions. By careful choice of the birth rate and death rate functions we can hope to get a small approximation error, that even approaches zero as the number of indicators in the sum increases.

We now remark on some features of the sequence \((J_n)\) of pattern indicators that facilitate calculations.

The sequence \((J_n)\) is a random element taking values in \((\mathbb{R}^N, \otimes_{k=1}^\infty B(\mathbb{R}))\), where \( B(\mathbb{R}) \) is the Borel \( \sigma \)–algebra of \( \mathbb{R} \) and \( \otimes_{k=1}^\infty B(\mathbb{R}) \) is the product \( \sigma \)–algebra of \( \mathbb{R}^N \). Similarly, the i.i.d. sequence of symbols \((Z_n)\) is a random element taking values in \((S^N, \otimes_{k=1}^\infty 2^S)\).

Two important features of \((J_n)\) are that it is \((l-1)\)–dependent (recall that \( l \) is the pattern length) and stationary. By \((l-1)\)–dependent we mean that the random vector \((J_1, \ldots, J_k)\) and random sequence \((J_{k+l}, J_{k+l+1}, \ldots)\) are independent for all \( k \geq 1 \). To see this, notice that \((J_1, \ldots, J_k)\) is measurable relative to the \( \sigma \)–algebra generated by \( \{Z_1, \ldots, Z_{k+l-1}\} \), while \((J_{k+l}, J_{k+l+1}, \ldots)\) is measurable relative to the \( \sigma \)–algebra generated by \( \{Z_{k+l}, Z_{k+l+1}, \ldots\} \) and these two \( \sigma \)–algebras are independent. Saying \((J_n)\) is stationary means that the random sequence \((J_k, J_{k+1}, \ldots)\) has
the same distribution as \((J_1, J_2, \ldots)\), for all \(k \geq 1\). The sequence \((J_n)\) is stationary because \((Z_n)\) is stationary. For each \(k \geq 1\), the sequence \((Z_k, Z_{k+1}, \ldots)\) has distribution that is the product of a countable infinity of copies of the distribution of \(Z_1\).

We can write

\[(J_k, J_{k+1}, \ldots) = f(Z_k, Z_{k+1}, \ldots)\]

for a function \(f : (S^N, \otimes_{k=1}^{\infty} 2^S) \rightarrow (\mathbb{R}^N, \otimes_{k=1}^{\infty} \mathcal{B}(\mathbb{R}))\) that is measurable and does not depend on \(k\). Since the distribution of \((Z_k, Z_{k+1}, \ldots)\) does not depend on \(k\), nor does the distribution of \((J_k, J_{k+1}, \ldots)\).

Since the pattern \((a_1, \ldots, a_l)\) is non-overlapping, there is in addition scope to simplify calculations due to the fact that the product \(J_i h(J_1, J_2, \ldots, J_n)\), where \(h\) is measurable and real-valued, can be replaced by a product of independent random variables:

\[J_i h(J_1, \ldots, J_n) = J_i h(J_1, \ldots, J_{i-l}, 0, \ldots, 1, 0, \ldots, J_{i+l}, \ldots, J_n)\]

with 1 in the \(i\)th slot on the right.

The simplifying features described above are used without comment in the calculations of the following chapters.
Chapter 2

Normal approximation for patterns in a string

In this chapter we show that normal approximation of the distribution of $W_n$ only becomes useful for comparatively large values of $n$. As pattern probability $p$ decreases, the error of normal approximation of $L(W_n)$ increases. However, if we are interested in $W_n$ taking values in a certain interval then we can be sure that the error of normal approximation goes to zero as $n$ increases to infinity.

The distribution of \( (W_n - E[W_n]) / \sqrt{\text{Var}[W_n]} \) converges weakly to the standard normal distribution. Hence it is natural to consider normal approximation of $L(W_n)$, at least for events that are intervals. More generally, a theorem of Rinott [10] tells us that for a stationary, $m$-dependent sequence of bounded random variables, the distribution functions of the centred and scaled partial sums converge uniformly to the standard normal distribution function, $\Phi$, with a bound for the difference that is $O(n^{-1/2})$.

**Theorem 1 (Rinott)** Let $Y_1, Y_2, \ldots, Y_n$ be an $m$-dependent sequence of random variables such that $E[Y_i] = 0$ and $|Y_i| \leq B$ for $1 \leq i \leq n$.

Let $S_i = \{j : 1 \leq j \leq n \text{ and } |j - i| \leq m\}$. Let $\sigma^2 = E\left[\left(\sum_{i=1}^{n} Y_i\right)^2\right]$ and let
\[ W = \sum_{i=1}^{n} Y_i / \sigma. \] Then
\[
\sup_{w \in \mathbb{R}} |P\{W \leq w\} - \Phi(w)| \leq \frac{4}{\sigma^2} \sqrt{E[\{\sum_{i=1}^{n} \sum_{j \in S_i} (Y_i Y_j - E[Y_i Y_j])\}^2]}
+ \frac{2}{\sigma^3} \left(1 + \frac{3}{2} \sqrt{\frac{\pi}{2}}\right) \sqrt{E[\{\sum_{i=1}^{n} |Y_i| (\sum_{j \in S_i} Y_j)^2\}^2]}
+ \frac{1}{\sqrt{2\pi}} \frac{(2m+1)B}{\sigma} + \frac{4}{\sqrt{2\pi}} \frac{n(2m+1)^2B^3}{\sigma^3}. \quad (2.1)
\]

For details of the proof see Rinott [10]. According to Rinott, much of the work was done earlier by Stein (see citation in [10]).

A random variable \( X \) is standard normal if and only if it satisfies
\[
E[f'(X) - Xf(X)] = 0
\]
for all continuous functions, \( f \), with derivative, \( f' \), defined at all but finitely many points and continuous and bounded. Therefore the Stein equation for the standard normal distribution is
\[
f'(x) - xf(x) = h(x) - \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} h(t) \exp \left(-\frac{t^2}{2}\right) dt, \quad x \in \mathbb{R}.
\]
In the proof of the theorem, \( h \) is taken to be a continuous ramp function approximating the indicator \( I_{(-\infty,w]} \); this enables a Taylor expansion for \( f \) to be used when bounding \( E[f'(W) - Wf(W)] \).

We now find crude bounds for the first two terms on the right of inequality (2.1). We will see that, in our application, the order of magnitude of the expression on the right of (2.1) is determined by the last term.

Let
\[
V_i = \sum_{j \in S_i} (E[Y_i Y_j] - Y_i Y_j), \quad i = 1, \ldots, n.
\]
\( V_i \) and \( V_k \) are independent if \(|k - i| \geq 3m+1\). Thus since \(|V_i| \leq (2m+1)2B^2\), we
have
\[ E[(\sum_{i=1}^{n} V_i)^2] = \sum_{i=1}^{n} \sum_{1 \leq k \leq n \text{ and } |k-i| \leq 3m} E[V_i V_k] \leq n(6m + 1)(2m + 1)^2 4B^4. \]

It follows that
\[ \frac{4}{\sigma^2} \sqrt{E[(\sum_{i=1}^{n} V_i)^2]} \leq \frac{8}{\sigma^2} (2m + 1)B^2(6m + 1)^{1/2} n^{1/2}. \] (2.2)

Also we have
\[ \sum_{i=1}^{n} |Y_i|((\sum_{j \in S_i} Y_j)^2 \leq nB(2m + 1)^2 B^2 \]
so
\[ \frac{2}{\sigma^3} (1 + 3/2 \sqrt{\pi/2}) \sqrt{E[(\sum_{i=1}^{n} |Y_i|((\sum_{j \in S_i} Y_j)^2)] \leq 2(1 + 3/2 \sqrt{\pi/2})(2m + 1)^2 B^3. \] (2.3)

Putting (2.2) and (2.3) and (2.1) together gives
\[ \sup_{w \in \mathbb{R}} |P\{W \leq w\} - \Phi(w)| \leq \frac{(2m + 1)B}{\sigma} \left( 8(6m + 1)^{1/2} B \frac{n^{1/2}}{\sigma} + (2 + 3 \sqrt{\pi/2} + 4 / \sqrt{2\pi})(2m + 1)B^2 \frac{n}{\sigma^2} + \frac{1}{\sqrt{2\pi}} \right). \] (2.4)

If \( Y_1, Y_2, \ldots \) is stationary as well then
\[ \sigma^2 = (n - 2m)E[\sum_{j=1}^{m+1} Y_{m+1}Y_j] + E[\sum_{i=1}^{n} \sum_{j \in S_i} Y_i Y_j + \sum_{i=m+1}^{2m} \sum_{j=i-m}^{2m} Y_i Y_j]; \] (2.5)
thus \( \sigma^2 \) is essentially proportional to \( n \) and the bound on the right of (2.4) is, in this case, \( O(n^{-1/2}) \).

In our pattern–counting application, with a non–overlapping pattern, \( m \) is \( l - 1 \) and \( Y_i = J_i - p, \ i = 1, 2, \ldots, n \) and \( B = 1 - p \). The variance of \( W_n \) is
\[ \sigma^2 = np(1 - (2l - 1)p) + l(l - 1)p^2. \] (2.6)
When we substitute these values of \( m, B \) and \( \sigma \) in the bound on the right of (2.4), we see that, for small \( p \) and large \( n \), the bound has roughly the same order of magnitude as

\[
\frac{(2m + 1)^2 B^3 n}{\sigma^3} \approx \frac{(2l - 1)^2 (1 - p)^3}{n^{1/2}(p(1 - (2l - 1)p))^{3/2}}.
\]

For this number to be less than one, it is necessary to have

\[
n > \frac{(2l - 1)^4}{p^3}.
\]  

(2.7)

Thus, since pattern probability \( p \) for a non–overlapping pattern is at most \( \frac{1}{l} \left( \frac{l-1}{l} \right)^{l-1} \) (see Appendix A), the bound we get from (2.4) is only useful for comparatively large values of \( n \). For \( l = 4 \) and \( p = \frac{1}{4} \left( \frac{l-1}{l} \right)^{l-1} = 27/256 \), inequality (2.7) becomes: \( n > 2046542 \).

Table 2.1 shows some computed actual values of

\[
\sup_{w \in \mathbb{R}} |P\{(W_n - np)/(Var[W_n])^{1/2} \leq w\} - \Phi(w)|
\]

for patterns of length \( l = 4 \) and various values of \( p \) and \( n \). Also shown, in parentheses, are the values of

\[
\sup_{w \in \mathbb{R}} |P\{(W_n - np)/(Var[W_n])^{1/2} \leq w\} - \Phi(w)| \times (np(1 - (2l - 1)p))^{1/2}.
\]

The data shows that \( \sup_{w \in \mathbb{R}} |P\{(W_n - np)/(Var[W_n])^{1/2} \leq w\} - \Phi(w)| \) is approximately proportional to \( (np(1 - (2l - 1)p))^{-1/2} \) for \( np > 1 \) and small values of \( p \). In particular, the value of the sup is reduced by a factor of approximately 1/2 when \( p \) increases by a factor of 4. This \( p \) dependence indicates that, for our problem, there is room for improvement in the bound in (2.1), particularly in the last term.

In following chapters, we will see that when approximating with suitable PBD distributions, one can bound the total variation distance error. The error bounds
Table 2.1: Values of $\sup_{w \in \mathbb{R}} \{ P \left( \frac{(W_n - np)}{\text{Var}[W_n]} \leq w \} - \Phi(w) \}$ and, in parentheses, the product of this quantity with $(np(1 - (2l-1)p))^{1/2}$. Pattern length $l$ is four.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$1/1024$</th>
<th>$1/512$</th>
<th>$1/256$</th>
<th>$1/128$</th>
<th>$1/64$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0.99812$</td>
<td>$0.96661$</td>
<td>$0.83852$</td>
<td>$0.60041$</td>
<td>$0.45320$</td>
</tr>
<tr>
<td></td>
<td>(0.31084)</td>
<td>(0.42426)</td>
<td>(0.51686)</td>
<td>(0.51598)</td>
<td>(0.53463)</td>
</tr>
<tr>
<td>1000</td>
<td>$0.50949$</td>
<td>$0.37354$</td>
<td>$0.27106$</td>
<td>$0.19721$</td>
<td>$0.14604$</td>
</tr>
<tr>
<td></td>
<td>(0.50176)</td>
<td>(0.51845)</td>
<td>(0.52835)</td>
<td>(0.53593)</td>
<td>(0.54478)</td>
</tr>
<tr>
<td>10000</td>
<td>$0.17089$</td>
<td>$0.12160$</td>
<td>$0.086659$</td>
<td>$0.062670$</td>
<td>$0.046288$</td>
</tr>
<tr>
<td></td>
<td>(0.53220)</td>
<td>(0.53371)</td>
<td>(0.53416)</td>
<td>(0.53857)</td>
<td>(0.54604)</td>
</tr>
<tr>
<td>100000</td>
<td>$0.054104$</td>
<td>$0.038450$</td>
<td>$0.027474$</td>
<td>$0.019838$</td>
<td>$0.014643$</td>
</tr>
<tr>
<td></td>
<td>(0.53283)</td>
<td>(0.53367)</td>
<td>(0.53553)</td>
<td>(0.53911)</td>
<td>(0.54625)</td>
</tr>
</tbody>
</table>

are useful for relatively small values of $n$ and, unlike normal approximation, the error decreases as $p$ decreases. In our approximation with two–parameter PBD distributions, the error also has the desirable property of shrinking to zero as $n$ increases.
Brown and Xia [5] described the Stein method for a large class of approximating distributions that are invariant distributions of birth–death processes. In this chapter, following Brown and Xia [5], we first give a brief description of approximating with birth–death distributions. Then we focus on approximation with a Poisson distribution and apply it to our problem. Poisson approximation is a well–established method and was first developed by Chen [6]. We give upper and lower bounds for the total variation distance between $L(W_n)$ and the Poisson distribution with the same mean.

### 3.1 Stein method for approximating with birth–death distributions

Let $\pi$ be the invariant distribution of a positive recurrent birth–death process with birth rate $\alpha_i$ in state $i$ and death rate $\beta_i$ in state $i$, $i \in \mathbb{Z}_+$. We assume that $\beta_0 = 0$ and $\beta_i > 0$ for all $i > 0$. Then, writing $\pi_i$ for $\pi\{i\}$, we have

$$\alpha_i \pi_i = \beta_{i+1} \pi_{i+1} \text{ for all } i \in \mathbb{Z}_+. \quad (3.1)$$
From which it follows that

$$\pi_i = \frac{\alpha_0 \alpha_1 \cdots \alpha_{i-1}}{\beta_1 \beta_2 \cdots \beta_i} \pi_0 \text{ for all } i \geq 1$$  \hspace{1cm} (3.2)

and

$$\pi_0 = \left( 1 + \sum_{i=1}^{\infty} \frac{\alpha_0 \cdots \alpha_{i-1}}{\beta_1 \cdots \beta_i} \right)^{-1}. \hspace{1cm} (3.3)$$

For more information on invariant distributions of birth–death processes see, for example, [7, pages 249–250].

From now on we restrict ourselves to birth rates, $\alpha_i$, that are non-increasing and death rates, $\beta_i$, that are non-decreasing and $\beta_0 = 0$ and $\beta_i > 0$ for $i > 0$.

Let $g : \mathbb{Z}_+ \rightarrow \mathbb{R}$ be any bounded function. Multiplying both sides of (3.1) by $g(i + 1)$ and summing over $i$, we have:

$$\sum_{i=0}^{\infty} g(i+1)\alpha_i \pi_i = \sum_{i=0}^{\infty} g(i+1)\beta_{i+1} \pi_{i+1}.$$  

Thus if $W$ is a random variable with distribution $\pi$ then

$$E[\alpha_W g(W+1) - \beta_W g(W)] = 0$$ \hspace{1cm} (3.4)

for all bounded $g : \mathbb{Z}_+ \rightarrow \mathbb{R}$. Conversely one can show that if (3.4) holds for all bounded functions $g$ then $W$ has distribution $\pi$. Therefore we have the Stein equation for $\pi$–approximation:

$$\alpha_i g(i + 1) - \beta_i g(i) = I_A(i) - \pi(A), \hspace{0.5cm} i \in \mathbb{Z}_+. \hspace{1cm} (3.5)$$

Here $I_A$ is the indicator of $A \subset \mathbb{Z}_+$. By convention we take $g(0) = 0$. Brown and Xia [5] showed that when $\alpha_i > 0$ for all $i \in \mathbb{Z}_+$ then the solution $g$ of (3.5) is bounded and

$$\|\Delta g\| := \sup_{i \geq 1} |g(i+1) - g(i)| \leq \sup_{i \geq 1} \min \{1/\alpha_i, 1/\beta_i\}. \hspace{1cm} (3.6)$$
As mentioned in chapter 1, when birth rates $\alpha_i$ and death rates $\beta_i$ are polynomial functions, we call $\pi$ a polynomial birth–death (PBD) distribution. A particularly simple PBD distribution is the Poisson distribution with mean $\lambda$; in this case we can take $\alpha_i = \lambda$ and $\beta_i = i$, $i \in \mathbb{Z}_+$.

3.2 Poisson approximation

Let $Y_1, \ldots, Y_n$ be indicator random variables defined on some probability space and $E[Y_i] = p_i > 0$, $1 \leq i \leq n$. Let $W = \sum_{i=1}^n Y_i$. When $Y_1, \ldots, Y_n$ satisfy certain conditions, which will be defined below, they are said to be negatively related and we can get a useful lower bound as well as an upper bound for the total variation distance between $\mathcal{L}(W)$ and the Poisson distribution with mean $\lambda = \sum_{i=1}^n p_i$.

In the first part of our discussion, we follow Lindvall [9, pages 60–64] and derive an upper bound. A similar treatment can also be found in [2, pages 23–26]. Adapting the Stein equation in (3.5), the Stein equation for approximation with a Poisson distribution having mean $\lambda$, $\text{Poi}(\lambda)$, is:

$$\lambda h(i + 1) - ih(i) = I_A(i) - \text{Poi}(\lambda)(A), \quad i \in \mathbb{Z}_+. \tag{3.7}$$

Here $I_A$ is the indicator of $A \subset \mathbb{Z}_+$. By (3.6), the solution $h$ of (3.7) satisfies

$$\|\Delta h\| := \sup_{i \geq 1} |h(i + 1) - h(i)| \leq \min\{1/\lambda, 1\}. \tag{3.8}$$

Set $\lambda = \sum_{i=1}^n p_i$. We shall bound $E[\lambda h(W + 1) - Wh(W)]$ using some couplings. A coupling of two distributions on $\mathbb{Z}_+$, $\mu$ and $\nu$, is when we introduce two random elements defined on the same probability space and having respective distributions $\mu$ and $\nu$.

For each $i$, $1 \leq i \leq n$, define on the same probability space

$$U_i \overset{d}{=} W \quad \text{and} \quad V_i + 1 \overset{d}{=} P(W \in \cdot | Y_i = 1). \tag{3.9}$$
Here \( d \) means 'having the same distribution' and \( P(W \in \cdot | Y_i = 1) \) denotes the distribution of \( W \) relative to the probability measure \( P(\cdot \cap \{ Y_i = 1\})/P\{ Y_i = 1\} \); the latter probability measure has density \( Y_i/p_i \) relative to \( P \). We have:

\[
E[\lambda h(W + 1) - Wh(W)] = \sum_{i=1}^{n} p_i (E[h(W + 1)] - E[h(W)Y_i/p_i])
\]

\[
= \sum_{i=1}^{n} p_i (E[h(U_i + 1)] - E[h(V_i + 1)]). \tag{3.10}
\]

By writing the difference as a telescoping sum, one can show that

\[
|h(U_i + 1) - h(V_i + 1)| \leq \|\Delta h\| |U_i - V_i|.
\]

Hence from equation (3.10) we get

\[
|E[\lambda h(W + 1) - Wh(W)]| \leq \|\Delta h\| \sum_{i=1}^{n} p_i E[|U_i - V_i|]. \tag{3.11}
\]

Now we put additional specifications on \( U_i \) and \( V_i \). Suppose for each \( i, \, 1 \leq i \leq n \), we can define on the same probability space random vectors

\[
(Y_{i1}, \ldots, Y_{in}) \overset{d}{=} (Y_1, \ldots, Y_n) \text{ and } (Y'_{i1}, \ldots, Y'_{in}) \overset{d}{=} P((Y_1, \ldots, Y_n) \in \cdot | Y_i = 1)
\]

such that

\[
Y_{ij} \geq Y'_{ij} \text{ for all } j \in \{1, \ldots, n\} - \{i\}. \tag{3.12}
\]

When such random vectors exist, \( Y_1, \ldots, Y_n \) are said to be \textit{negatively related}. We set

\[
U_i = \sum_{j=1}^{n} Y_{ij} \text{ and } V_i + 1 = \sum_{j=1}^{n} Y'_{ij}, \, i = 1, \ldots, n.
\]

Then \( U_i \) and \( V_i + 1 \) have distributions as in (3.9) and also

\[
U_i - V_i = \sum_{1 \leq j \leq n \text{ and } j \neq i} (Y_{ij} - Y'_{ij}) + Y_{ii} - (Y'_{ii} - 1) \geq 0
\]
with probability 1, due to inequality (3.12). Thus when $Y_1, \ldots, Y_n$ are negatively related we can simplify the right hand side of (3.11):

\[
\sum_{i=1}^{n} p_i E[|U_i - V_i|] = \sum_{i=1}^{n} p_i E[U_i - V_i] = \sum_{i=1}^{n} p_i (E[W] - E[V_i + 1]), \quad \text{from (3.9)}
\]

\[
= \lambda^2 + \sum_{i=1}^{n} p_i (1 - E[WY_i]/p_i), \quad \text{from (3.9)}
\]

\[
= \lambda^2 + \lambda - E[\sum_{i=1}^{n} W]
\]

\[
= \lambda^2 + \lambda - E[W^2]
\]

\[
= \lambda - Var[W]. \quad (3.13)
\]

Substituting this into (3.11) and using (3.8), we have for negatively related $Y_1, \ldots, Y_n$,

\[
d_{TV}(L(W), Pois(\lambda)) \leq \min \{1, 1/\lambda\}(\lambda - Var[W]). \quad (3.14)
\]

### 3.3 Poisson approximation for patterns in a string

We now turn to our application and show that the pattern indicators for a non–overlapping pattern, $J_1, \ldots, J_n$, are negatively related. For each $k = 1, \ldots, n$ define

\[
J'_{kj} := \begin{cases} 
J_j & \text{if } |j - k| \geq l, \\
1 & \text{if } j = k, \\
0 & \text{if } 0 < |j - k| \leq l - 1,
\end{cases}, \quad 1 \leq j \leq n.
\]

Take $(i_1, \ldots, i_n) \in \{0, 1\}^n$. We will show that

\[
P\{(J'_{k1}, \ldots, J'_{kn}) = (i_1, \ldots, i_n)\} = P\{(J_1, \ldots, J_n) = (i_1, \ldots, i_n)\| J_k = 1\}.
\]

We have
\[
P\{(J_1, \ldots, J_n) = (i_1, \ldots, i_n) \text{ and } J_k = 1\}/P\{J_k = 1\}
= \begin{cases} 
0 & \text{if } i_k = 0 \text{ or } i_j = 1 \text{ for some } j \text{ such that } 0 < |j - k| \leq l - 1, \\
P\{J_j = i_j \text{ for all } j \text{ such that } 1 \leq j \leq n \text{ and } |j - k| \geq l\} & \text{if } i_k = 1 \text{ and } i_j = 0 \text{ for all } j \text{ such that } 0 < |j - k| \leq l - 1.
\end{cases}
\]

Thus \((J'_{k1}, \ldots, J'_{kn}) \overset{d}{=} P((J_1, \ldots, J_n) \in \cdot\cdot\cdot | J_k = 1), k = 1, \ldots, n\). Also for each \(k\) we have \(J'_{kj} \leq J_j\) for all \(j \in \{1, \ldots, n\} - \{k\}\). Thus defining \(J_{kj} := J_j\) for \(j = 1, \ldots, n\) and \(k = 1, \ldots, n\), we see that \(J_1, \ldots, J_n\) are negatively related.

We can use (3.14) to get an upper bound for \(d_{TV}(\mathcal{L}(W_n), \text{Poi}(\lambda))\). Here \(\lambda = E[W_n] = np\) and \(\text{Var}[W_n] = np(1 - (2l - 1)p) + l(l - 1)p^2\). Putting these values into (3.14) we get:

\[
d_{TV}(\mathcal{L}(W_n), \text{Poi}(np)) \leq \min\{1, 1/(np)\}((2l - 1)n - l(l - 1))p^2.
\] (3.15)

As \(n\) increases, this bound increases to the limit \((2l - 1)p\).

Barbour, Holst and Janson [2, page 60] have derived a useful lower bound for \(d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda))\) when \(W := \sum_{i=1}^n Y_i\) is a sum of negatively related indicators with \(\lambda = E[W]\). They noted that if \(g\) is any bounded real valued function defined on \(\mathbb{Z}_+\) and \(X\) has Poisson distribution with mean \(\lambda\) then \(E[\lambda g(X + 1) - Xg(X)] = 0\); thus

\[
|E[\lambda g(W + 1) - Wg(W)]| = \left| E[\lambda g(W + 1) - Wg(W)] - E[\lambda g(X + 1) - Xg(X)] \right|
= \left| \sum_{x=0}^{\infty} (\lambda g(x + 1) - x g(x)) (P\{W = x\} - \text{Poi}(\lambda)\{x\}) \right|
\leq \sup_{j \in \mathbb{Z}_+} |\lambda g(j + 1) - jg(j)| 2 d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda))
\]

and

\[
d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)) \geq \frac{|E[\lambda g(W + 1) - Wg(W)]|}{2 \sup_{j \in \mathbb{Z}_+} |\lambda g(j + 1) - jg(j)|}.
\]
Barbour, Holst and Janson made a clever choice of the function $g$ and proved that when $Y_1, \ldots, Y_n$ are negatively related then

$$d_{TV}(L(W), Pois(\lambda)) \geq \frac{1 - Var[W]/\lambda}{11 + 3 \max \{0, (\lambda - E[(W - \lambda)^4] + 3(Var[W])^2)/(\lambda(\lambda - Var[W]))\}}.$$  

(3.16)

According to Barbour, Holst and Janson, the result was first proved for independent indicators by Barbour and Hall (see citation in [2]).

We now apply (3.16) to our problem and find a lower bound for $d_{TV}(L(W_n), Pois(np))$. By straightforward calculation we have:

$$1 - Var[W]/\lambda = p(2l - 1) - l(l - 1)p/n$$  

(3.17)

and

$$(\lambda - E[(W_n - \lambda)^4] + 3(Var[W_n])^2)/\lambda$$

$$= 7(2l - 1)p + p^2[2(2l - 1)(4l - 1)(4l - 3)p - 6(3l - 2)(3l - 1) + 3l^2(l - 1)^2 p/n].$$  

(3.18)

Now since $p \leq 1/(2l - 1)$, for $n \geq l/2$ the expression in square brackets in (3.18) is negative:

$$2(2l - 1)(4l - 1)(4l - 3)p - 6(3l - 2)(3l - 1) + 3l^2(l - 1)^2 p/n < 0.$$

Thus from (3.17) and (3.18) we have, for $n \geq l/2$:

$$((\lambda - E[(W_n - \lambda)^4] + 3(Var[W_n])^2)/\lambda)/((\lambda - Var[W_n])) < \frac{7(2l - 1)p}{np(p(2l - 1) - l(l - 1)p/n)} = \frac{7}{np \left(1 - \frac{l(l-1)}{(2l-1)n}\right)}.$$  

(3.19)

Finally, from (3.16) we get:

$$d_{TV}(L(W_n), Pois(np)) > \frac{1}{11 + 21/ \left(1 - \frac{l(l-1)}{(2l-1)n}\right)} \min \{1, 1/(np)\} \{((2l-1)n - l(l-1))p^2.}$$  

(3.20)
The upper bound for $d_{TV}(\mathcal{L}(W_n), Poi(np))$ in (3.15) shows that Poisson approximation is good when $p$ is small, however the lower bound in (3.20) increases towards $\frac{1}{32}(2t - 1)p$ as $n$ increases, showing that the error in the approximation does not shrink to zero as $n$ increases.
Chapter 4

Two–parameter PBD approximation for patterns in a string

In this chapter we approximate $\mathcal{L}(W_n)$ with a two–parameter PBD distribution, $\pi_1(\alpha, \beta)$.

Firstly we bound the total variation distance between the distributions of $W_n$ and $T_n$, where $T_n$ is the number of pattern occurrences in a circular arrangement of the $n$ i.i.d. random elements $Z_1, \ldots, Z_n$. In the second section we bound the error of approximating $\mathcal{L}(T_n)$ with the PBD distribution $\pi_1(\alpha, \beta)$. The advantage of working with $T_n$ instead of $W_n$ directly is that one can make better use of the stationarity of the sequence $J_1, J_2, \ldots$ when bounding the error.

4.1 Patterns in a circle

When approximating $\mathcal{L}(W_n)$ with some two–parameter PBD distributions, it is convenient to proceed by approximating the distribution of pattern occurrences in a circular arrangement of the random elements $Z_1, \ldots, Z_n$.

Recall from chapter 1 that $(Z_i)$ is an i.i.d. sequence of random elements in $S$ and $(a_1, \ldots, a_l)$ is a non–overlapping pattern we are interested in. $J_i = I_{\{(a_1, \ldots, a_l)\}}(Z_{i-1}, \ldots, Z_{i+l-1})$
is the indicator for a pattern occurrence in the sequence, starting at $Z_i$, $i \geq 1$. We defined $W_n := \sum_{i=1}^{n} J_i$, the number of patterns in the string $(Z_1, \ldots, Z_{n+l-1})$. Now we shall need to make a few more definitions. For fixed $n$, define

$$J^n_i = I_{\{a_1, \ldots, a_l\}}(Z_i, \ldots, Z_{i+l-1}) = J_i, \quad 1 \leq i \leq n - l + 1$$

and

$$J^n_{n-l+2} = I_{\{a_1, \ldots, a_l\}}(Z_{n-l+2}, \ldots, Z_n, Z_1), \ldots, J^n_n = I_{\{a_1, \ldots, a_l\}}(Z_n, Z_1, \ldots, Z_{l-1}).$$

$J^n_i$ is the indicator for a pattern starting at $Z_i$ in the circular arrangement of $Z_1, \ldots, Z_n$, with $Z_n$ looping round to join $Z_1$ and form the circle, as in Figure 4.1.

Define

$$T_n = \sum_{i=1}^{n} J^n_i.$$ 

This is the number of pattern occurrences in the circular arrangement of $Z_1, \ldots, Z_n$.

Let $\pi$ be a PBD distribution approximating $\mathcal{L}(T_n)$. Then

$$d_{TV}(\mathcal{L}(W_n), \pi) \leq d_{TV}(\mathcal{L}(T_n), \pi) + d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n)). \quad (4.1)$$

Figure 4.1: Circular arrangement of $Z_1, Z_2, \ldots, Z_n$. 

$Z_{n-1}$ $Z_n$ $Z_1$ $Z_2$ $Z_3$ $\cdots$
Estimating $d_{TV}(\mathcal{L}(T_n), \pi)$ is done simultaneously with the choice of values of parameters that appear in the formulas for the birth rates and death rates that determine $\pi$. Estimating $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n))$ only depends on the characteristics of the sequence $(J_i)$.

We estimate $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n))$ first. We shall need a couple of recurrence relations. The first recurrence relates the distribution of patterns in a circle to distributions of patterns in strings. The second describes a relationship between distributions of patterns in strings.

**Lemma 2** We have

$$P\{W_{n-l+1} = k\} - P\{T_n = k\} = (l - 1)p(P\{W_{n-2l+1} = k\} - P\{W_{n-2l+1} = k - 1\}),$$

for $n \geq l$ and $k \geq 0$.

**Lemma 3** We have

$$P\{W_n = k\} - P\{W_{n-1} = k\} = p(P\{W_{n-l} = k - 1\} - P\{W_{n-l} = k\}),$$

for $n \geq 1$ and $k \in \mathbb{Z}$.

In these two lemmas and in what follows, we interpret $W_i$ as $W_0 \equiv 0$ if $i < 0$. The proofs of Lemmas 2 and 3 are postponed until after we have bounded $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n))$.

**Theorem 4** We have the bound

$$d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n)) \leq p^2l(l - 1)\max_{n-3l+2 \leq j \leq n-2l}d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j + 1)).$$

**Proof:** From Lemma 2 we have

$$P\{T_n = k\} - P\{W_{n-l+1} = k\} = (l - 1)p(P\{W_{n-2l+1} = k - 1\} - P\{W_{n-2l+1} = k\})$$

(4.2)
and from Lemma 3 we have

\[ P\{W_n = k\} - P\{W_{n-l+1} = k\} = \sum_{j=n-l+1}^{n-1} (P\{W_{j+1} = k\} - P\{W_j = k\}) \]

\[ = p \sum_{j=n-l+1}^{n-1} (P\{W_{j+1-l} = k-1\} - P\{W_{j+1} = k\}) \]

\[ = p \sum_{r=0}^{l-2} (P\{W_{n-l-r} = k-1\} - P\{W_{n-l-r} = k\}) \]  \hspace{1cm} \text{(4.3)}

We now subtract equation (4.2) from (4.3):

\[ P\{W_n = k\} - P\{T_n = k\} \]

\[ = -p \sum_{r=0}^{l-2} (P\{W_{n-l-r} = k\} - P\{W_{n-l-r} = k-1\}) \]

\[ - (P\{W_{n-2l+1} = k\} - P\{W_{n-2l+1} = k-1\})) \]

\[ = -p \sum_{r=0}^{l-2} (P\{W_{n-l-r} = k\} - P\{W_{n-2l+1} = k\}) \]

\[ - (P\{W_{n-l-r} = k-1\} - P\{W_{n-2l+1} = k-1\})) \]

\[ = -p \sum_{r=0}^{l-2} \sum_{j=n-2l+1}^{n-l-r-1} (P\{W_{j+1} = k\} - P\{W_j = k\}) \]

\[ - (P\{W_{j+1} = k-1\} - P\{W_j = k-1\})) \]

\[ = p^2 \sum_{r=0}^{l-2} \sum_{j=n-2l+1}^{n-l-r-1} (P\{W_{j+1-l} = k-1\} - P\{W_{j+1} = k-1\}) \]

\[ - (P\{W_{j+1-l} = k-1\} - P\{W_{j+1-l} = k-2\})) \], \hspace{1cm} \text{by Lemma 3,} \]

\[ = p^2 \sum_{r=0}^{l-2} \sum_{j=n-3l+2}^{n-2l-r} (P\{W_j = k\} - P\{W_j = k-1\}) \]

\[ - (P\{W_j = k-1\} - P\{W_j = k-2\})) \], \hspace{1cm} \text{just simplifying subscripts,} \]

\[ = p^2 \sum_{j=n-3l+2}^{n-2l} \sum_{r=0}^{n-2l-j} (P\{W_j = k\} - P\{W_j = k-1\}) \]

\[ - (P\{W_j = k-1\} - P\{W_j = k-2\})) \]

\[ = p^2 (n-2l-j+1) (P\{W_j = k\} - P\{W_j = k-1\}) \]

\[ - (P\{W_j = k-1\} - P\{W_j = k-2\})) \]  \hspace{1cm} \text{(4.4)}
Thus
\[ |P\{W_n = k\} - P\{T_n = k\}| \leq p^2 \sum_{j=n-3l+2}^{n-2l} (n-2l-j+1)(|P\{W_j = k\} - P\{W_j = k-1\}| + |P\{W_j = k-1\} - P\{W_j = k-2\}|). \]

Now sum over \( k \) and multiply by \( \frac{1}{2} \) to get
\[ d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n)) \leq 2p^2 \sum_{j=n-3l+2}^{n-2l} (n-2l-j+1)d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j + 1)) \]
\[ \leq p^2 l(l-1) \max_{n-3l+2 \leq j \leq n-2l} d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j + 1)). \]

This completes the proof of Theorem 4.

We now prove Lemmas 2 and 3.

**Proof of Lemma 2** We write events mostly in words instead of symbols \( T_i, W_i, J_i \) etc. to (hopefully) make the proof clearer.

Let \( n \geq l \). When we form a circle from the string \( Z_1, \ldots, Z_n \) there can arise one additional pattern occurrence, thus for \( k \geq 0 \) we have
\[ P\{\text{string } Z_1, \ldots, Z_n \text{ has } k \text{ patterns}\} \]
\[ = P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k \text{ patterns and all start in } Z_1, \ldots, Z_{n-l+1}\} \]
\[ + P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k+1 \text{ patterns and one starts in } Z_{n-l+2}, \ldots, Z_n\}. \] (4.5)

The event on the left is the union of the events on the right. For \( k \geq 0 \) we also have
\[ P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k+1 \text{ patterns and one starts in } Z_{n-l+2}, \ldots, Z_n\} \]
\[ = (l-1)P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k+1 \text{ patterns and one starts at } Z_1\}. \] (4.6)

This follows from partitioning the left event according to where a pattern starts in \( Z_{n-l+2}, \ldots, Z_n \). Now \( P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k+1 \text{ patterns and one starts at } Z_1\} = pP\{\text{string } Z_1, \ldots, Z_{n-1} \text{ has } k \text{ patterns}\} \) so by equation (4.6),
\[ P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k + 1 \text{ patterns and one starts in } Z_{n-l+2}, \ldots, Z_n \} \]
\[ = (l - 1)pP\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\}. \tag{4.7} \]

For \( k = 0 \) equations (4.5) and (4.7) give

\[ P\{\text{string } Z_1, \ldots, Z_n \text{ has } 0 \text{ patterns}\} \]
\[ = P\{\text{circle } Z_1, \ldots, Z_n \text{ has } 0 \text{ patterns}\} \]
\[ + (l - 1)pP\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } 0 \text{ patterns}\}. \tag{4.8} \]

For \( k > 0 \) equation (4.5) gives

\[ P\{\text{string } Z_1, \ldots, Z_n \text{ has } k \text{ patterns}\} - P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k \text{ patterns}\} \]
\[ = P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k + 1 \text{ patterns and one starts in } Z_{n-l+2}, \ldots, Z_n \} \]
\[ - P\{\text{circle } Z_1, \ldots, Z_n \text{ has } k \text{ patterns and one starts in } Z_{n-l+2}, \ldots, Z_n \} \]
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \}
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \]
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \]
\[ = (l - 1)p \left( P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k \text{ patterns}\} \right. \]
\[ - P\{\text{string } Z_1, \ldots, Z_{n-l} \text{ has } k - 1 \text{ patterns}\} \]
where the last line follows from equation (4.7). Using the notation defined earlier we can write equations (4.8) and (4.9) as

\[ P\{W_{n-l+1} = k\} - P\{T_n = k\} = (l - 1)p \left( P\{W_{n-2l+1} = k\} - P\{W_{n-2l+1} = k - 1\} \right), \tag{4.10} \]

for \( n \geq l \) and \( k \geq 0 \). This completes the proof of Lemma 2.

**Proof of Lemma 3** For \( n \geq 1 \) we have

\[ P\{W_n = k\} = P\{W_n = k \text{ and } J_n = 1\} + P\{W_n = k \text{ and } J_n = 0\} \]
\[ = P\{W_{n-l} = k - 1 \text{ and } J_n = 1\} + P\{W_n = k \text{ and } J_n = 0\} \]
\[ = pP\{W_{n-l} = k - 1\} + P\{W_n = k \text{ and } J_n = 0\} \tag{4.11} \]
and

\[
P\{W_{n-1} = k\} = P\{W_n = k \text{ and } J_n = 0\} + P\{W_n = k+1 \text{ and } J_n = 1\} \\
= P\{W_n = k \text{ and } J_n = 0\} + P\{W_{n-l} = k \text{ and } J_n = 1\} \\
= P\{W_n = k \text{ and } J_n = 0\} + pP\{W_{n-l} = k\}.
\] (4.12)

Equation (4.11) minus equation (4.12) gives

\[
P\{W_n = k\} - P\{W_{n-1} = k\} = p(P\{W_{n-l} = k-1\} - P\{W_{n-l} = k\}), \quad (4.13)
\]

for \(n \geq 1\). This completes the proof of Lemma 3.

### 4.2 Approximation for patterns in a circle

Theorem 4 provides a bound for \(d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n))\), assuming we can estimate \(d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j+1))\). This section and the next chapter are about bounding the other term on the right of (4.1): \(d_{TV}(\mathcal{L}(T_n), \pi)\), where \(\pi\) is a two-parameter PBD distribution approximating \(\mathcal{L}(T_n)\). We look at one type of two-parameter PBD approximating distribution now and another type, which is more adapted to the nature of our problem, in the next chapter.

Let \(\pi_1(\alpha, \beta)\) be the invariant distribution of a birth–death process with constant birth rate, \(\alpha\), and death rate when in state \(i\) of \(\beta i + i(i-1)\), \(i \in \mathbb{Z}_+\). \(\alpha\) and \(\beta\) are the parameters and we shall choose positive values for them while bounding \(d_{TV}(\mathcal{L}(T_n), \pi_1(\alpha, \beta))\). Brown and Xia [5] have used this type of PBD distribution to approximate the distribution of a sum of independent but not identically distributed indicators. We follow their procedure for choosing \(\alpha\) and \(\beta\) and bounding \(d_{TV}(\mathcal{L}(T_n), \pi_1(\alpha, \beta))\). The idea behind using \(\alpha\) and \(\beta\) and bounding \(d_{TV}(\mathcal{L}(T_n), \pi_1(\alpha, \beta))\) is that death rates can increase fast with \(i\), reducing the tail of the distribution and leading to a better fit than in Poisson approximation.
When \( \alpha_i = \alpha \) and \( \beta_i = \beta i + i(i - 1) \), Stein equation (3.5) becomes

\[
\alpha g(i + 1) - (\beta i + i(i - 1))g(i) = I_A(i) - \pi(A), \quad i \in \mathbb{Z}_+.
\] (4.14)

Here \( A \) is a subset of \( \mathbb{Z}_+ \). Since \( \alpha \) and \( \beta \) are both positive, (3.6) tells us the solution, \( g \), of equation (4.14) satisfies

\[
\|\Delta g\| := \sup_{i \geq 1} |g(i + 1) - g(i)| \leq \frac{1}{\alpha}.
\] (4.15)

We will substitute \( T_n \) for \( i \) in equation (4.14) and bound the expected value of the left hand side. The following telescoping sum result is useful more than once when deriving the bound. Let \( h : \mathbb{Z}_+ \to \mathbb{R} \) be a real valued function and \( m \) be some number greater than the pattern length \( l \). We have

\[
E[h(T_{n+1}) - h(W_{n-m} + 1)]
\]

\[
= E \left[ \sum_{k=n-m}^{n-l-1} (h(W_{k+1} + 1) - h(W_k + 1))J_{k+1} \right.
\]

\[
+ \sum_{k=n-l}^{n-1} (h(W_{n-l-1} + J_{n-l} + \cdots + J_{k+1} + 1) - h(W_{n-l-1} + J_{n-l} + \cdots + J_k + 1))J_{k+1} \right]
\]

\[
= pE \left[ \sum_{k=n-m}^{n-l-1} (h(W_{k+1-l} + 2) - h(W_{k+1-l} + 1)) \right]
\]

\[
+ pE \left[ \sum_{k=n-l}^{n-1} (h(W_{n-2l+1} + 2) - h(W_{n-2l+1} + 1)) \right]
\]

\[
= lpE[\Delta h(W_{n-2l+1} + 1)] + p \sum_{k=n-m}^{n-l-1} E[\Delta h(W_{k+1-l} + 1)].
\] (4.16)

**Theorem 5** Suppose \( p < \frac{1}{4l-2} \) and set \( \alpha = \frac{n(n-2l+1)p(1-(2l-1)p)}{(2l-1)} \) and \( \beta = \frac{(n-2l+1)(1-p(4l-2))}{(2l-1)} \).

Then

\[
d_{TV}(\mathcal{L}(T_n), \pi_1(\alpha, \beta))
\]

\[
\leq \frac{p^2}{1 - p(2l - 1)}(11l(l-1) - 2l + 4 - p(10l^3 - 15l^2 + 9l - 2)) \max_{n-6l+4 \leq k \leq n-3l+1} d_{TV}(\mathcal{L}(W_{k+1}), \mathcal{L}(W_k)).
\]

32
Proof We must bound
\[
E[\alpha g(T_n + 1) - (\beta T_n + T_n(T_n - 1))g(T_n)]
\]
\[
= \alpha E[g(T_n + 1)] - \beta E[T_n g(T_n)] - E[T_n(T_n - 1)g(T_n)]
\]
\[
= \alpha E[g(T_n + 1)] - \beta \sum_{i=1}^{n} E[J_i^T g(T_n)] - \sum_{i=1}^{n} E[J_i^T (T_n - 1)g(T_n)]
\]
\[
= \alpha E[g(T_n + 1)] - \beta npE[g(W_{n-2l+1} + 1)] - npE[W_{n-2l+1}g(W_{n-2l+1} + 1)]
\]
\[
= \alpha E[g(T_n + 1)] - \beta npE[g(W_{n-2l+1} + 1)] - npE[T_{n-2l+1}g(T_{n-2l+1} + 1)]
\]
\[
+ np(E[T_{n-2l+1}g(T_{n-2l+1} + 1)] - E[W_{n-2l+1}g(W_{n-2l+1} + 1)]). \quad (4.17)
\]

We shall bound this in two parts.

First part

First we bound the sum of the first three terms on the right of equation (4.17):
\[
\alpha E[g(T_n + 1)] - \beta npE[g(W_{n-2l+1} + 1)] - npE[T_{n-2l+1}g(T_{n-2l+1} + 1)]
\]
\[
= \alpha E[g(T_n + 1)] - \beta npE[g(W_{n-2l+1} + 1)] - np \sum_{j=1}^{n-2l+1} E[J_j^T g(T_{n-2l+1} + 1)]
\]
\[
= \alpha E[g(T_n + 1)] - \beta npE[g(W_{n-2l+1} + 1)] - n(n - 2l + 1)p^2 E[g(W_{n-2l+2} + 2)]
\]
\[
= (\alpha - \beta np) E[g(T_n + 1)] + \beta npE[g(T_n + 1) - g(W_{n-2l+1} + 1)]
\]
\[
- n(n - 2l + 1)p^2 E[g(W_{n-2l+2} + 2)]
\]

where we have used equation (4.16) in the last line.

Now set
\[
\alpha - \beta np = n(n - 2l + 1)p^2. \quad (4.18)
\]

We have

33
\[
\alpha E[g(T_n + 1)] - \beta npE[g(W_{n-2l+1} + 1)] - npE[T_{n-2l+1}g(T_{n-2l+1} + 1)]
\]

\[
= \beta np^2 \left( \sum_{i=1}^{l-1} E[\Delta g(W_{n-2l+1-i} + 1)] + lE[\Delta g(W_{n-2l+1} + 1)] \right)
+ n(n - 2l + 1)p^2 E[g(T_n + 1) - g(W_{n-4l+2} + 2)]
\]

\[
= \beta np^2 \left( \sum_{i=1}^{l-1} E[\Delta g(W_{n-2l+1-i} + 1)] + lE[\Delta g(W_{n-2l+1} + 1)] \right)
+ n(n - 2l + 1)p^2 \{ E[g(T_n + 1) - g(W_{n-4l+2} + 1)] - E[g(W_{n-4l+2} + 2) - g(W_{n-4l+2} + 1)] \}
\]

\[
= \beta np^2 \left( \sum_{i=1}^{l-1} E[\Delta g(W_{n-2l+1-i} + 1)] + lE[\Delta g(W_{n-2l+1} + 1)] \right)
+ n(n - 2l + 1)p^3 \left\{ lE[\Delta g(W_{n-2l+1} + 1)] + \sum_{i=1}^{3l-2} E[\Delta g(W_{n-2l+1-i} + 1)] \right\}
- n(n - 2l + 1)p^2 E[\Delta g(W_{n-4l+2} + 1)].
\]

Equation (4.16) was used in the last line. Now take \( \beta \) so that

\[
\beta np^2(2l - 1) + n(n - 2l + 1)p^3(4l - 2) = n(n - 2l + 1)p^2,
\]

that is,

\[
\beta = \frac{(n - 2l + 1)(1 - p(4l - 2))}{2l - 1}.
\]

(4.19)

Note that, since \( \beta \) must be positive, we require \( p < \frac{1}{4l-2} \). This choice of \( \beta \) ensures that we add and subtract the same multiple of the term \( E[\Delta g(W_{n-2l+1} + 1)] \) in order to get second differences of \( g \). Thus

\[
\alpha E[g(T_n + 1)] - \beta npE[g(W_{n-2l+1} + 1)] - npE[T_{n-2l+1}g(T_{n-2l+1} + 1)]
\]

\[
= -\beta np^2 \sum_{i=1}^{l-1} E[\Delta g(W_{n-2l+1} + 1) - \Delta g(W_{n-2l+1-i} + 1)]
- n(n - 2l + 1)p^3 \sum_{i=1}^{3l-2} E[\Delta g(W_{n-2l+1} + 1) - \Delta g(W_{n-2l+1-i} + 1)]
+ n(n - 2l + 1)p^2 E[\Delta g(W_{n-2l+1} + 1) - \Delta g(W_{n-4l+2} + 1)]
\]

34
\[
= -\beta np^2 \sum_{i=1}^{l-1} \sum_{k=n-2l+1-i}^{n-2l} E[(\Delta g(W_{k+1} + 1) - \Delta g(W_k + 1))J_{k+1}]
\]

\[
- n(n - 2l + 1)p^3 \sum_{i=1}^{3l-2} \sum_{k=n-2l+1-i}^{n-2l} E[(\Delta g(W_{k+1} + 1) - \Delta g(W_k + 1))J_{k+1}]
\]

\[
+ n(n - 2l + 1)p^2 \sum_{k=n-4l+2}^{n-2l} E[(\Delta g(W_{k+1} + 1) - \Delta g(W_k + 1))J_{k+1}]
\]

\[
= -\beta np^3 \sum_{i=1}^{l-1} \sum_{k=n-2l+1-i}^{n-2l} E[\Delta^2 g(W_{k+1-l} + 1)]
\]

\[
- n(n - 2l + 1)p^4 \sum_{i=1}^{3l-2} \sum_{k=n-2l+1-i}^{n-2l} E[\Delta^2 g(W_{k+1-l} + 1)]
\]

\[
+ n(n - 2l + 1)p^3 \sum_{k=n-4l+2}^{n-2l} E[\Delta^2 g(W_{k+1-l} + 1)]
\]

\[
= -\beta np^3 \sum_{k=n-3l+2}^{n-2l} (3l - 1 + k - n)E[\Delta^2 g(W_{k+1-l} + 1)]
\]

\[
- n(n - 2l + 1)p^4 \sum_{k=n-5l+3}^{n-2l} (5l - 2 + k - n)E[\Delta^2 g(W_{k+1-l} + 1)]
\]

\[
+ n(n - 2l + 1)p^3 \sum_{k=n-4l+2}^{n-2l} E[\Delta^2 g(W_{k+1-l} + 1)]
\]

\[
= -n(n - 2l + 1)p^3 \left\{ \frac{1 - p(2l - 1)}{2l - 1} \sum_{k=n-3l+2}^{n-2l} (l + k - n)E[\Delta^2 g(W_{k+1-l} + 1)]
\right.
\]

\[
+ \sum_{k=n-4l+2}^{n-3l+1} (p(5l - 2 + k - n) - 1)E[\Delta^2 g(W_{k+1-l} + 1)]
\]
\[ + \sum_{k=n-5l+3}^{n-4l+1} (5l - 2 + k - n)E[\Delta^2 g(W_{k+1-l} + 1)] \right). \quad (4.20) \]

Now the sum of absolute values of the coefficients of \( E[\Delta^2 g(W_{k+1-l} + 1)] \), as \( k \) varies from \( n - 5l + 3 \) to \( n - 2l \), inside the braces in the right hand side of equation (4.20) is \( l(1 - lp) + \frac{1}{2}(l - 1)(3l - 2)(\frac{1}{2l-1} - p) \). Thus

\[ |\alpha E[g(T_n + 1)] - \beta npE[g(W_{n-2l+1} + 1)] - npE[T_{n-2l+1}g(T_{n-2l+1} + 1)]| \]

\[ \leq n(n - 2l + 1)p^3(l(1 - lp) + \frac{1}{2}(l - 1)(3l - 2)(\frac{1}{2l-1} - p)) \max_{n-6l+4 \leq k \leq n-3l+1} \frac{d_{TV}(\mathcal{L}(W_k + 1), \mathcal{L}(W_k))}{\alpha} \]

\[ \leq 2|\Delta g|n(n - 2l + 1)p^3(l(1 - lp) + \frac{1}{2}(l - 1)(3l - 2)(\frac{1}{2l-1} - p)) \max_{n-6l+4 \leq k \leq n-3l+1} \frac{d_{TV}(\mathcal{L}(W_k + 1), \mathcal{L}(W_k))}{\alpha} \]

\[ \leq \frac{p^2}{1 - p(2l - 1)}(2l(2l - 1)(1 - lp) + (l - 1)(3l - 2)(1 - p(2l - 1))) \max_{n-6l+4 \leq k \leq n-3l+1} d_{TV}(\mathcal{L}(W_k + 1), \mathcal{L}(W_k)) \]

where in the last step we have substituted our chosen value of \( \alpha \):

\[ \alpha = \frac{n(n - 2l + 1)p(1 - p(2l - 1))}{2l - 1}. \]

This value comes from equations (4.19) and (4.18).

**Second part**

We now bound the rest of the right hand side of equation (4.17):

\[
np(E[T_{n-2l+1}g(T_{n-2l+1} + 1)] - E[W_{n-2l+1}g(W_{n-2l+1} + 1)])
\]

\[
= np \left( \int_{Z_+} xg(x + 1) \left[ P_{T_{n-2l+1}}^{-1}(x) - P_{W_{n-2l+1}}^{-1}(x) \right] dx \right)
\]

\[
= np \sum_{x=1}^{m} xg(x + 1)(P\{T_{n-2l+1} = x\} - P\{W_{n-2l+1} = x\}),
\]

36
where \( m = \left\lfloor \frac{n-1}{l} \right\rfloor \), the largest integer not greater than \( \frac{n-1}{l} \),

\[
= np \left[ g(m+1) \sum_{x=1}^{m} x(P\{T_{n-2l+1} = x\} - P\{W_{n-2l+1} = x\}) \right.
- \left. \sum_{x=1}^{m-1} \left( \sum_{k=1}^{x} k(P\{T_{n-2l+1} = k\} - P\{W_{n-2l+1} = k\}) \right) \Delta g(x+1) \right]
\]  
(4.22)

where we have used the summation by parts formula (this can be proved by induction):

\[
\sum_{k=1}^{m} u_{k}v_{k} = v_{m} \sum_{j=1}^{m} u_{j} - \sum_{j=1}^{m-1} \left( \sum_{k=1}^{j} u_{k} \right) (v_{k+1} - v_{k}).
\]

We notice that the first sum in equation (4.22) is \( E[T_{n-2l+1}] - E[W_{n-2l+1}] = 0 \). Thus

\[
np(E[T_{n-2l+1}g(T_{n-2l+1} + 1)] - E[W_{n-2l+1}g(W_{n-2l+1} + 1)])
= np \sum_{x=1}^{m-1} \left[ \sum_{k=1}^{x} k(P\{W_{n-2l+1} = k\} - P\{T_{n-2l+1} = k\}) \right] \Delta g(x+1).  
\]  
(4.23)

Consider the sum inside square brackets in (4.23). Using equation (4.4) we can write the summand as

\[
k(P\{W_{n-2l+1} = k\} - P\{T_{n-2l+1} = k\})
= p^{2} \sum_{j=n-5l+3}^{n-4l+1} (n-4l+2-j)k[P\{W_{j} = k\} - P\{W_{j} = k-1\} - (P\{W_{j} = k-1\} - P\{W_{j} = k-2\})].
\]

Thus

\[
\sum_{k=1}^{x} k(P\{W_{n-2l+1} = k\} - P\{T_{n-2l+1} = k\})
= p^{2} \sum_{j=n-5l+3}^{n-4l+1} (n-4l+2-j) \sum_{k=1}^{x} k[P\{W_{j} = k\} - P\{W_{j} = k-1\} - (P\{W_{j} = k-1\} - P\{W_{j} = k-2\})]
\]

and the inner sum reduces to

\[
- \sum_{k=1}^{x} (P\{W_{j} = k-1\} - P\{W_{j} = k-2\}) + x(P\{W_{j} = x\} - P\{W_{j} = x-1\}).
\]

Hence

37
\[ np|E[T_{n-2l+1}g(T_{n-2l+1} + 1)] - E[W_{n-2l+1}g(W_{n-2l+1} + 1)]| \]
\[ \leq np \sum_{x=1}^{m-1} \sum_{k=1}^{x} k(P\{W_{n-2l+1} = k\} - P(T_{n-2l+1} = k)) \left| \Delta g(x + 1) \right| \]
\[ \leq np \sum_{x=1}^{m-1} \||\Delta g||p^2 \left( \sum_{j=n-5l+3}^{n-4l+1} (n - 4l + 2 - j) \sum_{x=1}^{m-1} \sum_{k=1}^{x} |P\{W_j = k - 1\} - P\{W_j = k - 2\}| \\
+ \sum_{j=n-5l+3}^{n-4l+1} (n - 4l + 2 - j) x |P\{W_j = x\} - P\{W_j = x - 1\}| \right) \]
\[ \leq \frac{(2l-1)p^2}{(n-2l+1)(1-p(2l-1))} \left( \sum_{j=n-5l+3}^{n-4l+1} (n - 4l + 2 - j) \sum_{x=1}^{m-1} \sum_{k=1}^{x} |P\{W_j = k - 1\} - P\{W_j = k - 2\}| \\
+ \sum_{j=n-5l+3}^{n-4l+1} (n - 4l + 2 - j) x |P\{W_j = x\} - P\{W_j = x - 1\}| \right), \text{ by inequality (4.15)} \]
\[ \leq \frac{(2l-1)p^2}{(n-2l+1)(1-p(2l-1))} \left( \sum_{j=n-5l+3}^{n-4l+1} (n - 4l + 2 - j)(m - 1)2d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j + 1)) \\
+ \sum_{j=n-5l+3}^{n-4l+1} (n - 4l + 2 - j)(m - 1)2d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j + 1)) \right) \]
\[ = \frac{4(m - 1)(2l - 1)p^2}{(n-2l+1)(1-p(2l-1))} \sum_{j=n-5l+3}^{n-4l+1} (n - 4l + 2 - j)d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j + 1)) \]
\[ \leq \frac{4(2l - 1)p^2}{l(1 - p(2l - 1))} \frac{l(l - 1)}{2} \max_{n-5l+3 \leq j \leq n-4l+1} d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j + 1)) \]
\[ = \frac{p^2}{1 - p(2l - 1)} 2(2l - 1)(l - 1) \max_{n-5l+3 \leq j \leq n-4l+1} d_{TV}(\mathcal{L}(W_j), \mathcal{L}(W_j + 1)). \tag{4.24} \]

Now from equation (4.17) we see that combining inequalities (4.21) and (4.24) gives us
\[ d_{TV}(\mathcal{L}(T_n), \pi_1(\alpha, \beta)) \]
\[ \leq \frac{p^2}{1 - p(2l - 1)} [2l(2l - 1)(1 - lp) + (l - 1)(3l - 2)(1 - p(2l - 1)) + 2(2l - 1)(l - 1)] \]
\[ \times \max_{n-6l+4 \leq k \leq n-3l+1} d_{TV}(\mathcal{L}(W_k + 1), \mathcal{L}(W_k)). \tag{4.25} \]

Rearrangement of the expression in square brackets in (4.25) completes the proof of Theorem 5.
We postpone the job of bounding $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1}))$ until the end but remark that, for small $p$, one can obtain bounds that are $O(k^{-1/2})$.

The bound for $d_{TV}(\mathcal{L}(T_n), \pi_1(\alpha, \beta))$ in Theorem 5 is about eleven times the bound for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n))$ in Theorem 4. This reassures us that, for this type of approximating distribution, working with $T_n$ instead of $W_n$ directly is justified.

In the next chapter we show that by using a different type of two-parameter PBD approximating distribution we can get a bound for the error of approximating $\mathcal{L}(T_n)$ that is of roughly the same size as the bound for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n))$ in Theorem 4.
Chapter 5

Another two–parameter PBD approximation

In this chapter we approximate the distribution of $T_n$ with a type of two–parameter PBD distribution, $\pi_2(r, q)$, that reduces to the binomial distribution $\text{Bin}(r, q)$ when $r$ is a positive integer. We will obtain a bound for $d_{TV}(\mathcal{L}(T_n), \pi_2(r, q))$ of roughly the same size as that for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n))$ in Theorem 4. The main result is Theorem 8.

5.1 Approximation with a binomial–like distribution

In Poisson approximation we choose the mean of the approximating Poisson distribution to be the same as that of the distribution we wish to approximate. In two–parameter PBD approximation we can try to match both the mean and the variance of the approximating distribution to those of the distribution being approximated.

We have $E[T_n] = np$ and $\text{Var}[T_n] = np(1 - (2l - 1)p)$. If $n$ is a multiple of $(2l - 1)$ then the distribution of $T_n$ has the same mean and variance as $\text{Bin}(n/(2l - 1), (2l - 1)p)$, the distribution of the number of successes in $n/(2l - 1)$ independent Bernoulli
trials, each with success probability \((2l - 1)p\). Thus, in this case, it is natural to approximate the distribution of \(T_n\) with \(Bin(n/(2l - 1), (2l - 1)p)\).

Consider the two-parameter PBD distribution \(\pi_2(r, q)\) which we define as having birth rates

\[
\alpha_i = \begin{cases} 
q(r - i) & , 0 \leq i \leq \lfloor r \rfloor \\
0 & , i > \lfloor r \rfloor 
\end{cases}
\]  

(5.1)

and death rates

\[
\beta_i = (1 - q)i, \ i \geq 0.
\]  

(5.2)

Here \(r\) is any positive real number and \(\lfloor r \rfloor\) is the largest integer not greater than \(r\). \(q\) satisfies \(0 < q < 1\). We hope, motivated by the discussion of the preceding paragraph, that taking

\[
r = n/(2l - 1)
\]  

(5.3)

and

\[
q = (2l - 1)p
\]  

(5.4)

makes \(\pi_2(r, q)\) a good approximation for the distribution of \(T_n\) for all values of \(n\).

The Stein equation, (3.5), for \(\pi_2(r, q)\) is

\[
q(r - i)g(i + 1) - (1 - q)ig(i) = I_A(i) - \pi_2(r, q)(A), \ 0 \leq i \leq \lfloor r \rfloor \\
and \ -(1 - q)ig(i) = I_A(i) - \pi_2(r, q)(A), \ i \geq \lfloor r \rfloor + 1.
\]  

(5.5)

Take the solution \(g\) of (5.5) and define

\[
f(i) = \begin{cases} 
g(i) & , 0 \leq i \leq \lfloor r \rfloor + 1 \\
0 & , i \geq \lfloor r \rfloor + 2
\end{cases}
\]  

(5.6)

Then \(f\) satisfies the equation

\[
q(r - i)f(i + 1) - (1 - q)if(i) = \begin{cases} 
I_A(i) - \pi_2(r, q)(A) & , 0 \leq i \leq \lfloor r \rfloor + 1 \\
0 & , i \geq \lfloor r \rfloor + 2
\end{cases}
\]  

(5.7)

Substituting \(T_n\) for \(i\) in (5.7), we have

\[
q(r - T_n)f(T_n + 1) - (1 - q)T_n f(T_n) = (I_A(T_n) - \pi_2(r, q)(A))I_{[0, \lfloor r \rfloor + 1]}(T_n)
\]
and thus
\[ E[I_A(T_n) - \pi_2(r, q)(A)] = E[q(r - T_n)f(T_n + 1) - (1 - q)T_n f(T_n)] \]
\[ + E[(I_A(T_n) - \pi_2(r, q)(A))I_{(|r|+1,\infty)}(T_n)]. \tag{5.8} \]

The second term on the right of equation (5.8) satisfies
\[ |E[(I_A(T_n) - \pi_2(r, q)(A))I_{(|r|+1,\infty)}(T_n)]| \leq P\{T_n > |r| + 1\}. \tag{5.9} \]

When \(|r| + 1 > E[T_n]\), as is certainly the case when \(r = n/(2l - 1)\), Chebychev’s inequality gives the crude bound
\[ P\{T_n > |r| + 1\} \leq \frac{1}{(|r| + 1 - E[T_n])^2 \text{Var}[T_n]}. \tag{5.10} \]

This bound makes poor use of the nature of the indicators \(J_i\). At the end of this chapter we use an elementary argument to show that if \(r = n/(2l - 1)\) and \(p < \frac{1}{l(l-1)}\) then \(P\{T_n > |r| + 1\}\) shrinks exponentially fast as \(n\) increases.

When finding a bound for the first term on the right of (5.8), we will need a bound for \(\|\Delta f\| := \sup_{i \geq 1} |f(i+1) - f(i)|\). By the definition of \(f\) in (5.6), we need to examine the solution \(g\) of equation (5.5). Lemma 6 gives the solution of a more general equation.

**Lemma 6** If \(\alpha_i > 0\) for \(0 \leq i \leq m\) and \(\alpha_i = 0\) for \(i > m\) and \(\beta_0 = 0\) and \(\beta_i > 0\) for \(i > 0\) then
\[ g(i) := \frac{1}{\beta_i \pi_i} (\pi(A)\pi\{i, i+1, \ldots\} - \pi(A \cap \{i, i+1, \ldots\})) \quad 1 \leq i \leq m + 1 \tag{5.11} \]
\[ - \frac{1}{\beta_i} (I_A(i) - \pi(A)) \quad i \geq m + 2 \tag{5.12} \]

is the solution of equation (3.5):
\[ \alpha_i g(i+1) - \beta_i g(i) = I_A(i) - \pi(A), \quad i \in \mathbb{Z}_+ \]
and \(g(0) = 0\).
The proof of Lemma 6 is straightforward. To get (5.11) use induction with the rearranged equation (3.5):

\[ g(i + 1) = \frac{1}{\alpha_i} (\beta g(i) + I_A(i) - \pi(A)), \quad 0 \leq i \leq m. \]

Then put \( i = m + 1 \) in (5.11) and rearrange to show

\[ -\beta_{m+1} g(m + 1) = I_A(m + 1) - \pi(A), \]

so (3.5) is satisfied even when \( i = m + 1 \). Finally, when \( i \geq m + 2 \) we have \( \alpha_i = 0 \) so (5.12) follows from (3.5).

By examining the solution in Lemma 6 carefully, one can show that if birth rates \( \alpha_i \) are non-increasing and \( \alpha_i > 0 \) for \( 0 \leq i \leq m \) and \( \alpha_i = 0 \) for \( i > m \) and death rates \( \beta_i \) are non-decreasing and \( \beta_0 = 0 \) and \( \beta_i > 0 \) for \( i > 0 \), then the solution \( g \) of (3.5) satisfies

\[ \|\Delta g\| := \sup_{i \geq 1} |g(i + 1) - g(i)| \leq \sup_{i \geq 1} \min \{1/\alpha_i, 1/\beta_i\}. \quad (5.13) \]

Here we take \( \frac{1}{\alpha} = \infty \). We omit the proof of this result, which follows the method used by Brown and Xia [5] in their proof of (3.6).

From the definition of \( f \) in (5.6), Lemma 6 and (5.13), we have:

\[ \|\Delta f\| := \sup_{i \geq 1} |f(i + 1) - f(i)| \leq \frac{1}{rq(1-q)}. \quad (5.14) \]

This is proved in Appendix B.

We can now bound the first term on the right of (5.8), taking \( r = n/(2l - 1) \) and \( q = (2l - 1)p \).

**Lemma 7** We have

\[
|E[(2l - 1)p(n \frac{p}{2l-1} - T_n)f(T_n + 1) - (1 - (2l - 1)p)T_n f(T_n)]| \\
\leq \frac{p^2}{1 - (2l - 1)p} l(l - 1) \max_{n-4l+3 \leq k \leq n-3l+1} d_{TV}(\mathcal{L}(W_k + 1), \mathcal{L}(W_k)).
\]
Proof

\[ E[(2l - 1)p\left(\frac{n}{2l-1} - T_n\right)f(T_n + 1) - (1 - (2l - 1)p)T_n f(T_n)] \]

\[ = E[(np - (2l - 1)pT_n)f(T_n + 1) - (1 - (2l - 1)p)T_n f(T_n)] \]

\[ = npE[f(T_n + 1)] - (2l - 1)p \sum_{i=1}^{n} E[J_i^n f(T_n + 1)] - (1 - (2l - 1)p) \sum_{i=1}^{n} E[J_i^n f(T_n)] \]

\[ = npE[f(T_n + 1)] - (2l - 1)pnE[J_{n-l+1}f(W_{n-2l+1} + 2)] \]

\[ - (1 - (2l - 1)p)pnE[J_{n-l+1}f(W_{n-2l+1} + 1)] \]

\[ = npE[f(T_n + 1)] - (2l - 1)p^2 nE[f(W_{n-2l+1} + 2)] - (1 - (2l - 1)p)pnE[f(W_{n-2l+1} + 1)] \]

\[ = npE[f(T_n + 1) - f(W_{n-2l+1} + 1)] - (2l - 1)p^2 nE[f(W_{n-2l+1} + 2) - f(W_{n-2l+1} + 1)] \]

and using (4.16) for the first term

\[ = np^2 \left( \sum_{k=n-2l+1}^{n-l-1} E[\Delta f(W_{k+1-l} + 1)] + lE[\Delta f(W_{n-2l+1} + 1)] \right) \]

\[ - (2l - 1)p^2 nE[\Delta f(W_{n-2l+1} + 1)] \]

\[ = np^2 \sum_{j=1}^{l-1} E[\Delta f(W_{n-2l+1-j} + 1)] - (l - 1)p^2 nE[\Delta f(W_{n-2l+1} + 1)] \]

\[ = -np^2 \sum_{j=1}^{l-1} E[\Delta f(W_{n-2l+1} + 1) - \Delta f(W_{n-2l+1-j} + 1)] \]

\[ = -np^2 \sum_{j=1}^{l-1} \sum_{k=1}^{j} E[(\Delta f(W_{n-2l+1-j+k} + 1) - \Delta f(W_{n-2l+1-j+k-1} + 1))J_{n-2l+1-j+k}] \]

\[ = -np^3 \sum_{j=1}^{l-1} \sum_{k=1}^{j} E[\Delta^2 f(W_{n-3l+1-j+k} + 1)]. \]

So

\[ |E[(2l - 1)p\left(\frac{n}{2l-1} - T_n\right)f(T_n + 1) - (1 - (2l - 1)p)T_n f(T_n)]| \]

\[ \leq np^3 \sum_{j=1}^{l-1} \sum_{k=1}^{j} |E[\Delta^2 f(W_{n-3l+1-j+k} + 1)]| \]

\[ \leq np^3 \sum_{j=1}^{l-1} \sum_{k=1}^{j} 2\|\Delta f\|_{TV}(L(W_{n-3l+1-j+k} + 1), L(W_{n-3l+1-j+k})) \]

44
\[
\leq \frac{2p^2}{1 - (2l - 1)p} \frac{l(l - 1)}{2} \max_{n - 4l + 3 \leq k \leq n - 3l + 1} d_{TV}(\mathcal{L}(W_k + 1), \mathcal{L}(W_k)).
\]

In the last line we have used (5.14) with \( r = n/(2l - 1) \) and \( q = (2l - 1)p \). This completes the proof of Lemma 7.

**Theorem 8** Take \( r = n/(2l - 1) \) and \( q = (2l - 1)p \). Then
\[
d_{TV}(\mathcal{L}(T_n), \pi_2(r, q)) \leq \frac{p^2}{1 - (2l - 1)p} \frac{l(l - 1)}{n - 4l + 3 \leq k \leq n - 3l + 1} d_{TV}(\mathcal{L}(W_k + 1), \mathcal{L}(W_k)) + P\{T_n > \lfloor r \rfloor + 1\}. \tag{5.15}
\]

**Proof** This follows from (5.8), (5.9) and Lemma 7.

The bound for \( d_{TV}(\mathcal{L}(T_n), \pi_2(n/(2l - 1), (2l - 1)p)) \) in Theorem 8 is roughly the same size as that for \( d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n)) \) in Theorem 4 provided that \( P\{T_n > \lfloor n/(2l - 1) \rfloor + 1\} \) is negligible. To end this chapter we show that if \( p < \frac{1}{l} \left( \frac{l-1}{l-1} \right)^{l-1} \) then \( P\{T_n > \lfloor n/(2l - 1) \rfloor + 1\} \) shrinks exponentially as \( n \) increases.

**Lemma 9** For \( n \geq 5(2l - 1) \) we have
\[
P\{T_n > \lfloor n/(2l - 1) \rfloor + 1\} \leq \frac{1}{\sqrt{2\pi}} \frac{(2l - 1)^{l/2}}{1 - (l - 1)p} \left( \frac{l}{l - 1 - \frac{2l - 1}{n}} \right)^{l-\frac{1}{2}} \exp \left( \frac{2l - 1}{12ln} + \frac{5(l - 1)}{5l - 6} + \frac{l - 1}{l} \right) n^{-\frac{1}{2}} \left[ \frac{p}{l} \left( \frac{l-1}{l-1} \right)^{-1} \right]^{\frac{n}{2l - 1}}.
\]

**Proof** We have the inclusion
\[
\{T_n > \lfloor n/(2l - 1) \rfloor + 1\} \subset \{W_n \geq \lfloor n/(2l - 1) \rfloor + 1\} \tag{5.16}
\]
so we bound \( P\{W_n \geq \lfloor n/(2l - 1) \rfloor + 1\} \). Define \( N(n, k; l) \) to be the number of increasing \( k \)-tuples, \((i_1, \ldots, i_k)\), from \( \{1, \ldots, n\} \) such that the difference between successive coordinates is at least \( l \):
\[
1 \leq i_1 < i_2 < \ldots < i_k \leq n \text{ and } i_j - i_{j-1} \geq l \text{ for } 2 \leq j \leq k.
\]
One can show that
\[ N(n, k; l) = \binom{n - (k - 1)(l - 1)}{k} \]
by substitution into the partial difference equation
\[ N(n, k; l) = N(n - l, k - 1; l) + N(n - 1, k; l) \]
and also checking that it satisfies the obvious boundary conditions. For each such
\( k \)-tuple \( (i_1, \ldots, i_k) \),
\[ P\{J_{i_1} = 1, J_{i_2} = 1, \ldots, J_{i_k} = 1 \text{ and } W_n = k\} \leq p^k. \]
Thus
\[ P\{W_n = k\} \leq \binom{n - (k - 1)(l - 1)}{k} p^k \]
and
\[ P\{W_n \geq \left\lfloor \frac{n}{2l - 1} \right\rfloor + 1\} \leq \sum_{k = \left\lfloor \frac{n}{2l - 1} \right\rfloor + 1}^{\left\lfloor \frac{n}{l} \right\rfloor} \binom{n - (k - 1)(l - 1)}{k} p^k. \]
Notice that the terms in the above sum decrease as \( k \) increases. For \( k \geq \frac{n}{2l - 1} \) we have
\[ \binom{n - k(l - 1)}{k + 1} p^{k+1} \leq \binom{n - (k - 1)(l - 1)}{k} p^k \left( \frac{n - kl}{k + 1}\right) \]
and the factor
\[ \frac{n - kl}{k + 1} p \leq \frac{n - \frac{n}{2l - 1} l}{\frac{n}{2l - 1} + 1} = \frac{n(l - 1)}{n + 2l - 1} p < (l - 1)p. \]
Thus
\[ P\{W_n \geq \left\lfloor \frac{n}{2l - 1} \right\rfloor + 1\} < \binom{n - \left\lfloor \frac{n}{2l - 1} \right\rfloor (l - 1)}{\left\lfloor \frac{n}{2l - 1} \right\rfloor + 1} p^{\left\lfloor \frac{n}{2l - 1} \right\rfloor + 1} \frac{1}{1 - (l - 1)p}. \quad (5.17) \]
Now we just have to estimate the right hand side using Stirling’s formula. We shall use the following bounds:
\[ e^{\frac{1}{12n+1}} \leq \frac{n!}{\sqrt{2\pi n} n^{n+1/2} e^{-n}} \leq e^{\frac{1}{12n}}. \quad (5.18) \]
This is proved in [4, page 325].

For notational convenience set \( t = \lfloor \frac{n}{2l-1} \rfloor \). We bound the product of the first two factors on the right of (5.17):

\[
\left( \frac{n - t(l - 1)}{t + 1} \right) p^{t+1}
\]

\[
= \frac{(n - t(l - 1))!}{(t + 1)!(n - t(l - 1) - t - 1)!}p^{t+1}
\]

\[
\leq \frac{e^{\frac{2l-1}{2\pi n}}}{\sqrt{2\pi}} \frac{(n - t(l - 1))^{n-t(l-1)+1/2}}{(t + 1)^{t+1+1/2}(n - t(l - 1) - t - 1)^n} p^{t+1}, \quad \text{by (5.18)}
\]

\[
= \frac{e^{\frac{2l-1}{2\pi n}}}{\sqrt{2\pi}} n^{-1/2} \left( 1 - \frac{t}{n} \right)^{n-t(l-1)+1/2} \left( 1 - \frac{t+1}{n} \right)^{n-t-1/2} p^{t+1}
\]

\[
\leq \frac{e^{\frac{2l-1}{2\pi n}}}{\sqrt{2\pi}} n^{-1/2} \left( 2l - 1 \right)^{1/2} \left[ 1 - \frac{t}{n} \left( l - 1 \right) \right]^{n-t-1/2} \left( (2l - 1)p \left( 1 - \frac{t}{n} \left( l - 1 \right) \right) \right)^{t+1}. \quad (5.19)
\]

We have the following bounds:

\[
\left[ 1 - \frac{t}{n} \left( l - 1 \right) \right]^{n-t-1/2} \leq \left( \frac{l}{l-1} \right)^{l-1/2} \left( \frac{l}{l-1} - \frac{2l-1}{n} \right)^{l-1/2}, \quad n \geq 5(2l-1), \quad (5.20)
\]

and

\[
(2l - 1)p \left( 1 - \frac{t}{n} \left( l - 1 \right) \right)^{t+1} \leq \left( pl \right)^{\frac{n}{2\pi}} e^{\frac{l-1}{2l+1}}. \quad (5.21)
\]

The proofs of (5.20) and (5.21) are given at the end. Using the bounds (5.20) and (5.21) in (5.19) we get

\[
\left( \frac{n - t(l - 1)}{t + 1} \right) p^{t+1}
\]

\[
\leq \frac{1}{\sqrt{2\pi}} (2l - 1)^{1/2} \left( \frac{l}{l-1} - \frac{2l-1}{n} \right)^{l-1/2} \exp \left( \frac{2l - 1}{12ln} + \frac{5(l - 1)}{5l - 6} + \frac{l - 1}{l} \right) n^{-1/2} \left[ \frac{p}{l^{(l-1)/(l-1)}} \right]^{\frac{n}{2\pi - 1}},
\]

for \( n \geq 5(2l-1) \). Finally, using this bound in (5.17) gives us the desired bound for \( P\{W_n \geq \lfloor \frac{n}{2l-1} \rfloor + 1 \} \) and hence for \( P\{T_n > \lfloor \frac{n}{2l-1} \rfloor + 1 \} \).
We now prove (5.20) and (5.21).

**Proof of (5.20)** First we bound the quantity inside square brackets on the left of (5.20):

\[
\frac{1 - \frac{t}{n}(l - 1)}{1 - \frac{t}{n}(l + 1)} \leq \frac{1 - \frac{1}{2\ell - 1}(l - 1)}{1 - \frac{l}{2\ell - 1} - \frac{1}{n}} = \frac{l}{l - 11 - \frac{2\ell - 1}{n(l - 1)}}.
\]

Thus since \( n - tl - 1/2 \leq n - (\frac{n}{2\ell - 1} - 1)l - 1/2 = n(\frac{l - 1}{2\ell - 1}) + l - 1/2 \), we have

\[
\left[ \frac{1 - \frac{t}{n}(l - 1)}{1 - \frac{t}{n}(l + 1)} \right]^{n-tl^{1/2}} \leq \left[ \frac{1 - \frac{1}{n}(l - 1)}{1 - \frac{t}{n}(l + 1)} \right]^{n(\frac{l - 1}{2\ell - 1})} \left[ \frac{1 - \frac{t}{n}(l - 1)}{1 - \frac{l}{n}(l - 1)} \right]^{l-1/2} \leq \left( \frac{l}{l - 1} \right)^{(\frac{l - 1}{2\ell - 1})} \left( 1 - \frac{2l - 1}{n(l - 1)} \right)^{-n(\frac{l - 1}{2\ell - 1})} \left( \frac{l}{l - 1 - \frac{2\ell - 1}{n}} \right)^{l-1/2}.
\]

Now the middle factor on the right,

\[
\left( 1 - \frac{2l - 1}{n(l - 1)} \right)^{-n(\frac{l - 1}{2\ell - 1})},
\]

decreases to the limit \( e \) as \( n \) increases and for \( n \geq 5(2l - 1) \) it is less than \( e^{\frac{5(l - 1)}{2\ell - 6}} \). Hence we get (5.20).

**Proof of (5.21)** We have

\[
((2l - 1)p(1 - \frac{t}{n}(l - 1)))^{t+1} \leq \left( (2l - 1)p \left( 1 - \frac{\frac{n}{2\ell - 1} - 1}{n} (l - 1) \right) \right)^{\frac{n}{2\ell - 1}} = \left( (2l - 1)p \left( \frac{l}{2\ell - 1} + \frac{1}{n} \right) \right)^{\frac{n}{2\ell - 1}} = (pl)^{\frac{n}{2\ell - 1}} (1 + \frac{(l - 1)(2l - 1)}{n(2\ell - 1)})^{\frac{n}{2\ell - 1}} \leq (pl)^{\frac{n}{2\ell - 1}} e^{\frac{l - 1}{2\ell - 1}}.
\]

This completes the proof of Lemma 9.

Lemma 9 shows that for \( p \ll \frac{1}{l} (\frac{l - 1}{2\ell - 1})^{l-1} \) the bound for \( d_{TV}(\mathcal{L}(T_n), \pi_2(\frac{n}{2\ell - 1}, (2l - 1)p)) \) in Theorem 8 is of roughly the same size as the bound for \( d_{TV}(\mathcal{L}(W_n), \mathcal{L}(T_n)) \) in Theorem 4.
Chapter 6

Numerical data for patterns of length four

In this chapter we present some computed values of the errors of approximating the distribution of $W_n$ with a Poisson distribution and with two–parameter PBD distributions of the types introduced in the previous two chapters. The data is for patterns of length four. We also compare these computed actual errors with our error estimates.

Table 6.1 shows the errors of approximating the distribution of $W_n$ with $\text{Poi}(np)$ and the two–parameter PBD distributions $\pi_1(\alpha, \beta)$ as in Theorem 5, and $\pi_2(r, q)$ as in Theorem 8. Thus we take

$$\alpha = \frac{n(n - 2l + 1)p(1 - (2l - 1)p)}{2l - 1} \quad \text{and} \quad \beta = \frac{(n - 2l + 1)(1 - p(4l - 2))}{2l - 1}$$

and

$$r = \frac{n}{2l - 1} \quad \text{and} \quad q = (2l - 1)p.$$

This choice of $\alpha$, $\beta$, $r$ and $q$ enables us to use Theorems 5 and 8 with Theorem 4 to estimate errors. Values of $P\{W_n = k\}$ were computed using the difference equation in Lemma 3. The PBD approximating distributions were computed using (3.2) and (3.3).
Figure 6.1: Actual total variation distance approximation errors. The pattern has length four and $p = 1/256$.

The first column of data in Table 6.1 contains computed values of the actual total variation distance approximation errors for a pattern of length four and $p = 1/256$. The value of $n$ increases from 1000 to 128000 going down the column. These data also appear in Figure 6.1. We see that the Poisson approximation errors are always about 0.0067. The two–parameter PBD approximation errors decrease as $n$ increases, approximately like $\propto n^{-1/2}$. The error in $\pi_1(\alpha, \beta)$ approximation is greater than the error in $\pi_2(r, q)$ approximation by a factor of about seven.

The second column of data contains corresponding values of our error estimates, to two significant figures. Poisson approximation error estimates do not vary as $n$ increases and we have an error range, $8.5 \times 10^{-4} - 0.027$, calculated from (3.15) and (3.20). The calculated upper bound for the error is about four times the actual error. Errors for $\pi_1(\alpha, \beta)$ approximation are calculated from Theorem 4 and Theorem 5 using estimates for $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_k + 1))$ from Theorem 11. Errors for $\pi_2(r, q)$ approximation are calculated from Theorem 4, Theorem 8 and Lemma 9.
using estimates for $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1}))$ from Theorem 11. The calculated error estimates for $\pi_1(\alpha, \beta)$ and $\pi_2(r, q)$ approximation are greater than the actual errors by a factor of about ten.

The third and fourth columns of data give actual approximation errors and error estimates, respectively, for a pattern of length four and $p = 1/16$. The actual error in $\pi_1(\alpha, \beta)$ approximation is nearly four times that in $\pi_2(r, q)$ approximation. The error estimates for $p = 1/16$ in the fourth column of data do not vary with $n$ because here we use the trivial bound $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1})) \leq 1$. We have no proven bounds for $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1}))$ when $l = 4$ and $p \geq 3/256$. When $l = 4$, Theorem 11 only applies for $p < 3/256$. Thus a convincing demonstration of the approximating abilities of $\pi_1(\alpha, \beta)$ and $\pi_2(r, q)$ really does depend on getting good estimates for $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1}))$; this problem is considered in the next chapter.
Table 6.1: Actual total variation distance errors, $d_{TV}(\mathcal{L}(W_n), \pi)$, and the corresponding estimated values. $\mathcal{L}(W_n)$ is approximated with $\pi = Poi(np)$, $\pi_1(\alpha, \beta)$ and $\pi_2(r, q)$. Pattern length is four and there are two different values of $p$: $p = 1/256$ and $p = 1/16$.

<table>
<thead>
<tr>
<th></th>
<th>$\pi$</th>
<th>$p = 1/256$</th>
<th>$p = 1/16$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual error</td>
<td>Estimate</td>
<td>Actual error</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1000</td>
<td>$Poi(np)$</td>
<td>0.00693037</td>
<td>$8.5 \times 10^{-4} - 0.027$</td>
</tr>
<tr>
<td></td>
<td>$\pi_1(\alpha, \beta)$</td>
<td>9.46099 $\times 10^{-5}$</td>
<td>$9.2 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_2(r, q)$</td>
<td>1.38956 $\times 10^{-3}$</td>
<td>$1.6 \times 10^{-4}$</td>
</tr>
<tr>
<td>2000</td>
<td>$Poi(np)$</td>
<td>0.00678042</td>
<td>$8.5 \times 10^{-4} - 0.027$</td>
</tr>
<tr>
<td></td>
<td>$\pi_1(\alpha, \beta)$</td>
<td>6.25006 $\times 10^{-3}$</td>
<td>$6.5 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_2(r, q)$</td>
<td>9.11759 $\times 10^{-6}$</td>
<td>$1.1 \times 10^{-4}$</td>
</tr>
<tr>
<td>4000</td>
<td>$Poi(np)$</td>
<td>0.00670484</td>
<td>$8.5 \times 10^{-4} - 0.027$</td>
</tr>
<tr>
<td></td>
<td>$\pi_1(\alpha, \beta)$</td>
<td>4.31429 $\times 10^{-3}$</td>
<td>$4.6 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_2(r, q)$</td>
<td>6.25203 $\times 10^{-6}$</td>
<td>$7.8 \times 10^{-5}$</td>
</tr>
<tr>
<td>8000</td>
<td>$Poi(np)$</td>
<td>0.00672083</td>
<td>$8.5 \times 10^{-4} - 0.027$</td>
</tr>
<tr>
<td></td>
<td>$\pi_1(\alpha, \beta)$</td>
<td>3.00288 $\times 10^{-3}$</td>
<td>$3.2 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_2(r, q)$</td>
<td>4.33398 $\times 10^{-6}$</td>
<td>$5.5 \times 10^{-5}$</td>
</tr>
<tr>
<td>16000</td>
<td>$Poi(np)$</td>
<td>0.00670845</td>
<td>$8.5 \times 10^{-4} - 0.027$</td>
</tr>
<tr>
<td></td>
<td>$\pi_1(\alpha, \beta)$</td>
<td>2.12329 $\times 10^{-3}$</td>
<td>$2.3 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_2(r, q)$</td>
<td>3.05706 $\times 10^{-6}$</td>
<td>$3.9 \times 10^{-5}$</td>
</tr>
<tr>
<td>32000</td>
<td>$Poi(np)$</td>
<td>0.00671504</td>
<td>$8.5 \times 10^{-4} - 0.027$</td>
</tr>
<tr>
<td></td>
<td>$\pi_1(\alpha, \beta)$</td>
<td>1.49590 $\times 10^{-3}$</td>
<td>$1.6 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_2(r, q)$</td>
<td>2.15573 $\times 10^{-6}$</td>
<td>$2.8 \times 10^{-5}$</td>
</tr>
<tr>
<td>64000</td>
<td>$Poi(np)$</td>
<td>0.00670974</td>
<td>$8.5 \times 10^{-4} - 0.027$</td>
</tr>
<tr>
<td></td>
<td>$\pi_1(\alpha, \beta)$</td>
<td>1.05635 $\times 10^{-3}$</td>
<td>$1.1 \times 10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_2(r, q)$</td>
<td>1.52179 $\times 10^{-6}$</td>
<td>$2.0 \times 10^{-5}$</td>
</tr>
<tr>
<td>128000</td>
<td>$Poi(np)$</td>
<td>0.00670964</td>
<td>$8.5 \times 10^{-4} - 0.027$</td>
</tr>
<tr>
<td></td>
<td>$\pi_1(\alpha, \beta)$</td>
<td>7.46339 $\times 10^{-6}$</td>
<td>$8.1 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_2(r, q)$</td>
<td>1.07514 $\times 10^{-6}$</td>
<td>$1.4 \times 10^{-5}$</td>
</tr>
</tbody>
</table>
Chapter 7

Bounds for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_{n+1}))$

Error estimates for two–parameter PBD approximation of the distribution of $W_n$ depend on good estimates for $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1}))$. In this chapter we firstly present some data suggesting $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1}))$ is bounded by $0.4/(\text{Var}[W_n])^{1/2}$. Motivated by this we try to bound $d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1}))$ by bounding the characteristic function of $W_k$; this follows the idea of Barbour and Jensen [3] for a sum of independent indicators. Bounds that are $O(k^{-1/2})$ are obtained for $l = 2$. There are only partial results for $l > 2$. Bounds that are $O(k^{-1/2})$ are obtained for $l = 3$ and $l = 4$, but only for small $p$: $p < (l-1)/l^l$; however the method can be extended to larger values of $l$ under the condition $p < (l-1)/l^l$.

Some computed values of $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_{n+1}))$ are shown in Table 7.1 for patterns with length eight and $p$ value 1/256, length four and $p$ value 1/256 and length four and $p$ value 1/16. The numbers in parentheses are values of $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_{n+1})) \times (np(1 - (2l - 1)p))^{1/2}$. Recall that $\text{Var}[W_n] = np(1 - (2l - 1)p) + l(l-1)p^2$. The data suggest that

$$d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_{n+1})) \leq 0.4(np(1 - (2l - 1)p))^{-1/2}. \quad (7.1)$$
Table 7.1: Values of $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$. Also shown, in parentheses, are values of the product $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1)) \times (np(1 - (2l - 1)p))^{1/2}$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$l = 8$ and $p = \frac{1}{256}$</th>
<th>$l = 4$ and $p = \frac{1}{256}$</th>
<th>$l = 4$ and $p = \frac{1}{16}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.2031624 (0.389594)</td>
<td>0.2013701 (0.392514)</td>
<td>0.06701932 (0.397376)</td>
</tr>
<tr>
<td>2000</td>
<td>0.1461315 (0.396303)</td>
<td>0.1442264 (0.397575)</td>
<td>0.04753202 (0.398568)</td>
</tr>
<tr>
<td>4000</td>
<td>0.1040351 (0.399005)</td>
<td>0.1024643 (0.399450)</td>
<td>0.03362598 (0.398754)</td>
</tr>
<tr>
<td>8000</td>
<td>0.07355263 (0.398944)</td>
<td>0.07237435 (0.399015)</td>
<td>0.02378274 (0.398849)</td>
</tr>
<tr>
<td>16000</td>
<td>0.05202445 (0.399058)</td>
<td>0.05119250 (0.399141)</td>
<td>0.01681891 (0.398895)</td>
</tr>
<tr>
<td>32000</td>
<td>0.03675153 (0.398675)</td>
<td>0.03615643 (0.398676)</td>
<td>0.01189346 (0.398917)</td>
</tr>
<tr>
<td>64000</td>
<td>0.02599597 (0.398808)</td>
<td>0.02557499 (0.398808)</td>
<td>0.008410195 (0.398932)</td>
</tr>
<tr>
<td>128000</td>
<td>0.01838501 (0.398875)</td>
<td>0.01808727 (0.398874)</td>
<td>0.005946993 (0.398930)</td>
</tr>
</tbody>
</table>
It is shown in Appendix C that the distribution of $W_n$ is unimodal, thus we have
\[
d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n+1)) := 1/2 \sum_{k=0}^{\infty} |P\{W_n = k\} - P\{W_n = k-1\}| = \max_{k \geq 0} P\{W_n = k\},
\]
the maximum of the density of $W_n$. Barbour and Jensen [3] showed that if $W$ is a sum of independent indicator random variables then
\[
\max_{k \geq 0} P\{W = k\} \leq \frac{1}{2} (\text{Var}[W])^{-1/2}.
\]
This looks similar to the proposed bound in (7.1). Barbour and Jensen [3] worked with the characteristic function of the sum $W$ and the Fourier inversion formula and we do the same here with $W_n$.

Define for each $k \geq 0$
\[
\phi_k(t) = E[e^{itW_k}], \text{ for all } t \in \mathbb{R},
\]
the characteristic function of $W_k$. We have the Fourier inversion formula:
\[
P\{W_k = m\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_k(t) e^{-itm} dt, \text{ } m \geq 0.
\]
Thus we have the bound
\[
\max_{m \geq 0} P\{W_k = m\} \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_k(t)| dt. \tag{7.4}
\]
The problem then is to find a nice bounding function for the modulus of the characteristic function, $|\phi_k|$. Barbour and Jensen [3] made the useful observation that
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\tau(1-\cos t)} dt \leq \frac{1}{2} \tau^{-1/2} \text{ when } \tau > 0. \tag{7.5}
\]
They used (7.5) to obtain (7.3). We also write our bounds for $|\phi_k(t)|$ in terms of the expression $e^{-\tau(1-\cos t)}$, where $\tau$ is a positive number.

Proceeding as described in the previous paragraph, we get the following two theorems.
Theorem 10 For patterns of length \( l = 2 \) we have

\[
d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1})) \leq \begin{cases} 
\frac{9}{16} (p(1 - 4p))^{-1/2} k^{-1/2} + \frac{9}{8} \left(\frac{1}{\sqrt{2}}\right)^k, & 0 < p < \frac{3}{32}, \\
\frac{41}{80} (p(1 - 4p))^{-1/2} k^{-1/2} + \frac{41}{30} \left(\frac{1}{\sqrt{2}}\right)^k + \frac{3}{5} k \left(\frac{5}{6}\right)^k, & \frac{3}{32} \leq p \leq \frac{5}{32}, \\
\frac{5}{2} (p(1 - 4p))^{-1/2} k^{-1/2} + \left(\frac{1}{\sqrt{2}}\right)^k, & \frac{5}{32} < p < \frac{13}{64}, \\
2k^{-1/2} + \left(\frac{1}{\sqrt{2}}\right)^k, & \frac{13}{64} \leq p \leq \frac{1}{4}.
\end{cases}
\]

Theorem 10 covers all possible values of \( p \) for patterns of length two.

Theorem 11 For patterns of length \( l = 3 \) or \( l = 4 \) we have

\[
d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_{k+1})) \leq \begin{cases} 
0.77(p(1 - \frac{27}{2}p))^{-1/2} k^{-1/2} + (0.33 + 0.61k) \left(\frac{1}{3}\right)^k, & l = 3 \text{ and } 0 < p \leq \frac{1}{27}, \\
0.68(p(1 - \frac{256}{3}p))^{-1/2} k^{-1/2} + (0.04 + 0.05k + 0.1k^2) \left(\frac{1}{3}\right)^k, & l = 4 \text{ and } 0 < p < \frac{3}{256}.
\end{cases}
\]

Notice that Theorem 11 applies only for small values of \( p \). A consolation is that the method of proof should work for larger values of \( l \), subject to the condition \( p < (l - 1)/l \). Details of the proofs of Theorems 10 and 11 are in Appendix D and Appendix E, respectively.

One might well wonder at the contrast between the neat–looking conjectured bound (7.1) and the messy bounds in Theorems 10 and 11. We now describe how Theorems 10 and 11 were arrived at.

Recall that we want to bound \( d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1)) \) by bounding \( \phi_n \), the characteristic function of \( W_n \), and using (7.4). For \( n \geq 1 \) we have

\[
\phi_n(t) = E[J_n e^{itW_n} + (1 - J_n)e^{itW_n}] \\
= E[J_n e^{it(W_{n-1}+1)} + (1 - J_n)e^{itW_{n-1}}] \quad (W_{n-l} \text{ is interpreted as 0 if } n - l < 0) \\
= pe^{it}E[e^{itW_{n-l}}] + E[e^{itW_{n-1}}] - E[J_n e^{itW_{n-l}}] \\
= \phi_{n-1}(t) + p(e^{it} - 1)\phi_{n-l}(t). \quad (7.6)
\]
Here is one approach to bounding $\phi_n$. Looking at (7.3), one gets the idea that the bound in (7.1) is like that expected for the sum of $\frac{n}{2l-1}$ (assuming this is an integer) independent indicators, each with success probability $(2l-1)p$. The characteristic function of such an indicator is $1 + (2l-1)p(e^{it} - 1)$. Applying (7.6) recursively $(2l-1)$ times gives

$$
\phi_n(t) = \phi_{n-2l+1}(t) + p(e^{it} - 1) \sum_{i=l}^{3l-2} \phi_{n-i}(t)
$$

Set

$$
w_k(t) = \frac{p(e^{it} - 1)}{1 + (2l-1)p(e^{it} - 1)} \sum_{i=l}^{3l-2} \left( \frac{\phi_{k(2l-1)-i}(t)}{\phi_{(k-1)(2l-1)}(t)} - 1 \right), \quad k = 1, 2, \ldots.
$$

Then for $m \geq 1$, we get from (7.7):

$$
\phi_{m(2l-1)}(t) = (1 + (2l-1)p(e^{it} - 1))^m \prod_{i=1}^{m} [1 + w_i(t)].
$$

The factor $(1 + (2l-1)p(e^{it} - 1))^m$ on the right of this equation is the characteristic function of the sum of $m$ independent indicators, each with success probability $(2l-1)p$. Thus we expect that

$$
\left| \prod_{i=1}^{m} [1 + w_i(t)] \right| \leq K, \quad t \in \mathbb{R},
$$

for some constant $K$ not much different from 1. Unfortunately, I do not have a proof for (7.8). A more direct approach yields bounds for $\phi_n$ that can be used in (7.4).

By equation (7.6) we have

$$
\phi_{k+l}(t) - \phi_{k+l-1}(t) - p(e^{it} - 1)\phi_k(t) = 0, \quad \text{for } k \geq 0 \text{ and } t \in (-\pi, \pi].
$$

(7.9)
This is an \( l \)th order linear difference equation with constant coefficients for each fixed \( t \in (-\pi, \pi] \setminus \{0\} \). For each \( t \in (-\pi, \pi] \setminus \{0\} \) we can associate with the difference equation (7.9) a polynomial equation of degree \( l \) called its characteristic equation:

\[
(z(t))^l - (z(t))^{l-1} - p(e^{it} - 1) = 0.
\] (7.10)

If for a particular \( t \in (-\pi, \pi] \setminus \{0\} \) this equation has \( l \) distinct roots: \( z_1(t), \ldots, z_l(t) \), then for this value of \( t \) the general solution of equation (7.9) has the form:

\[
\phi_k(t) = c_1(t)(z_1(t))^k + \cdots + c_l(t)(z_l(t))^k, \quad \text{for } k \geq 0,
\] (7.11)

where coefficients \( c_1(t), \ldots, c_l(t) \) can be determined if we know \( \phi_0(t), \ldots, \phi_{l-1}(t) \).

Solving linear difference equations with constant coefficients is discussed in, for example, [8, pages 19—22].

If for all but finitely many \( t \) in \((-\pi, \pi]\) equation (7.10) has \( l \) distinct roots: \( z_1(t), \ldots, z_l(t) \), then when bounding \( |\phi_k(t)| \) for use in (7.4), we can concentrate on bounding the right hand side of (7.11):

\[
|c_1(t)(z_1(t))^k + \cdots + c_l(t)(z_l(t))^k|.
\]

We do this in two stages: firstly bound the roots \( z_i(t) \) of (7.10) and then bound the coefficients \( c_i(t) \) in (7.11). When \( l = 2 \), (7.10) is a quadratic equation and the procedure is straightforward. The results for \( l = 2 \) are in Theorem 10, which is proved in Appendix D.

One might look to use the results for \( l = 2 \) in Theorem 10 to get results for \( l > 2 \). For example, consider \( W_{m(l-1)} \). We can form subsums of \( l - 1 \) successive indicators:

\[
\mathcal{J}_1 := J_1 + \cdots + J_{l-1}, \quad \mathcal{J}_2 := J_1 + \cdots + J_{2l-2}, \ldots, \mathcal{J}_l := J_{(l-1)(l-1)+1} + \cdots + J_{l(l-1)}, \ldots.
\]

This gives a 1-dependent sequence of indicators \( \{\mathcal{J}_i\} \) and we have

\[
W_{m(l-1)} = \sum_{i=1}^{m} \mathcal{J}_i.
\]
Now define \((I_i)\) to be an i.i.d. sequence of indicators with success probability \((l - 1)p\) and set \(V_m := \sum_{i=1}^{m} I_i\). Also define \((K_i)\) to be a sequence of indicators for a non-overlapping pattern with length two and probability \((l - 1)p\) and set \(X_m := \sum_{i=1}^{m} K_i\). If \(J_i\) has value 1 then \(J_{i+1}\) (or \(J_{i-1}\)) has reduced, but not necessarily zero, probability of taking the value 1. This effect is between the 'extreme' effects of \(I_i\) and \(K_i\) on their respective neighbours. Thus it is not surprising that computational results suggest that

\[
P\{W_{m(l-1)} = k\} \text{ lies between } P\{V_m = k\} \text{ and } P\{X_m = k\} \text{ for all } k \geq 0. \quad (7.12)
\]

See for example Table 7.2 which shows some values of \(P\{W_{m(l-1)} = k\} - P\{V_m = k\}\) and \(P\{X_m = k\} - P\{V_m = k\}\) for \(l = 4, m = 400\) and \(p = 1/256\). The signs of the two differences are the same for all \(k\) but the magnitude of \(P\{X_m = k\} - P\{V_m = k\}\) is always larger, as shown in Figure 7.1. Using Lemma 3 one can prove by induction that

\[
P\{X_m = 0\} \leq P\{W_{m(l-1)} = 0\} \leq P\{V_m = 0\}.
\]

However, going on to prove (7.12) seems much harder. A proof of (7.12) would yield bounds for \(d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_k + 1))\) via Theorem 10 that are valid for \(l > 2\) and \(0 < p \leq \frac{1}{4(l-1)}\).

To get Theorem 11 we have to work with equations (7.10) and (7.11), just as for \(l = 2\). The formulas for the roots of polynomial equations of degree \(l = 3\) or 4 are not as simple as the quadratic formula. For \(l > 4\) there is no general formula for the roots. Hence for \(l > 2\) we have bounds for the roots of the characteristic equation (7.10) only for \(p\) values in the restricted range \(0 < p < \frac{l-1}{l}\).

The rest of this chapter is an outline of our procedure for bounding the right hand side of (7.11) when \(l > 2\). Most of the proofs are in Appendix E.
Figure 7.1: Differences for $l = 4$, $m = 400$ and $p = \frac{1}{256}$.

Table 7.2: Here we take $l = 4$, $m = 400$ and $p = \frac{1}{256}$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$P{W_{m(l-1)} = k} - P{V_{m} = k}$</th>
<th>$P{X_{m} = k} - P{V_{m} = k}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$-3.31 \times 10^{-4}$</td>
<td>$-4.945 \times 10^{-4}$</td>
</tr>
<tr>
<td>1</td>
<td>$-9.105 \times 10^{-4}$</td>
<td>$-1.3669 \times 10^{-3}$</td>
</tr>
<tr>
<td>2</td>
<td>$-9.054 \times 10^{-4}$</td>
<td>$-1.3663 \times 10^{-3}$</td>
</tr>
<tr>
<td>3</td>
<td>$2.62 \times 10^{-5}$</td>
<td>$3.39 \times 10^{-5}$</td>
</tr>
<tr>
<td>4</td>
<td>$1.113 \times 10^{-3}$</td>
<td>$1.6755 \times 10^{-3}$</td>
</tr>
<tr>
<td>5</td>
<td>$1.459 \times 10^{-3}$</td>
<td>$2.202 \times 10^{-3}$</td>
</tr>
<tr>
<td>6</td>
<td>$9.875 \times 10^{-4}$</td>
<td>$1.4887 \times 10^{-3}$</td>
</tr>
<tr>
<td>7</td>
<td>$2.396 \times 10^{-4}$</td>
<td>$3.569 \times 10^{-4}$</td>
</tr>
<tr>
<td>8</td>
<td>$-2.886 \times 10^{-4}$</td>
<td>$-4.403 \times 10^{-4}$</td>
</tr>
<tr>
<td>9</td>
<td>$-4.634 \times 10^{-4}$</td>
<td>$-7.014 \times 10^{-4}$</td>
</tr>
<tr>
<td>10</td>
<td>$-4.026 \times 10^{-4}$</td>
<td>$-6.065 \times 10^{-4}$</td>
</tr>
<tr>
<td>11</td>
<td>$-2.651 \times 10^{-4}$</td>
<td>$-3.975 \times 10^{-4}$</td>
</tr>
<tr>
<td>12</td>
<td>$-1.449 \times 10^{-4}$</td>
<td>$-2.1613 \times 10^{-4}$</td>
</tr>
</tbody>
</table>
Using Rouché’s theorem from complex analysis, we can get a rough idea of the locations of the roots of equation (7.10). This special case of Rouché’s theorem suffices.

**Lemma 12 (Rouché’s theorem)** Let \( \gamma \) be a circle in the complex plane and suppose that \( f(z) \) and \( g(z) \) are polynomials satisfying

\[
|f(z) - g(z)| < |g(z)| \quad \text{for all } z \text{ on } \gamma.
\]

Then \( f(z) \) and \( g(z) \) have the same number of zeros, counting multiplicities, inside \( \gamma \).

For a proof see, for example, [1, page 153]. Using Rouché’s theorem we can prove the next lemma.

**Lemma 13** All roots of \( z^l - z^{l-1} - p(e^{it} - 1) = 0 \) lie inside the circle in the complex plane with centre \( \frac{1}{l} \) and radius \( \frac{l-1}{l} \) when \( 0 < p < \frac{l-1}{l} \) and \( t \in (-\pi, \pi) - \{0\} \).

Lemma 13 tells us that any root \( z \) of the characteristic equation (7.10) satisfies \( |z - \frac{1}{l}| \leq \frac{l-1}{l} \) when \( p < \frac{l-1}{l} \). We can use this knowledge to get a bound for the roots of (7.10), as stated in Lemma 14.

**Lemma 14** All roots of \( z^l - z^{l-1} - p(e^{it} - 1) = 0 \) satisfy

\[
|z(t)| \leq 1 - \frac{1}{l} \left( \frac{1}{l-1} \right)^{l-1} p(1 - p/(\frac{l-1}{l})) (1 - \cos t)
\]

when \( 0 \leq p < \frac{l-1}{l} \).

Lemmas 13 and 14 are both proved in Appendix E. We will see that when \( 0 < p < \frac{l-1}{l} \) there are \( l \) distinct roots of equation (7.10), \( z_1(t), \ldots, z_l(t) \), for all \( t \in (-\pi, \pi) \).
\{0\}. Thus for 0 < p < \frac{l-1}{l} equation (7.11) for \(\phi_k(t)\) holds for all \(t \in (-\pi, \pi] - \{0\}\) and Lemma 14 tells us that for any root \(z_j\) of equation (7.10) we have

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |z_j(t)|^k dt \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - \frac{1}{l} \frac{l-1}{l} p(1 - p/(\frac{l-1}{l}))(1 - \cos t))^k dt \\
= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{k \ln(1 - \frac{1}{l} \frac{l-1}{l} p(1 - p/(\frac{l-1}{l}))(1 - \cos t))} dt \\
\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-[k \frac{l-1}{l} \frac{l-1}{l} p(1 - p/(\frac{l-1}{l}))(1 - \cos t)]} dt, \quad \text{since} \ln(1 + x) \leq x \quad \text{for} \quad x > -1,
\]

\[
\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\frac{k}{2} \frac{l-1}{l} \frac{l-1}{l} p(1 - p/(\frac{l-1}{l}))(1 - \cos t)}} dt \\
= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\frac{k}{2} \frac{l-1}{l} \frac{l-1}{l} p(1 - p/(\frac{l-1}{l}))(1 - \cos t)}} dt, \quad \text{since} \ln(1 + x) \leq x \quad \text{for} \quad x > -1,
\]

(7.13)

where the last line follows from (7.5).

Next we find expressions for the coefficients \(c_1(t), \ldots, c_l(t)\) in equation (7.11) and find constant bounds for them. We have \(\phi_0(t) = 1\) and, by equation (7.6), \(\phi_k(t) = 1 + kp(e^t - 1)\) for \(1 \leq k \leq l - 1\). Thus equation (7.11) gives us a system of linear equations in the unknowns \(c_1(t), \ldots, c_l(t)\):

\[
\begin{align*}
&c_1(t) + \cdots + c_l(t) = 1 \\
c_1(t)z_1(t) + \cdots + c_l(t)z_l(t) = 1 + c_l(t)(e^t - 1) \\
c_1(t)(z_1(t))^2 + \cdots + c_l(t)(z_l(t))^2 = 1 + 2p(e^t - 1) \\
&\vdots \\
c_1(t)(z_1(t))^{l-1} + \cdots + c_l(t)(z_l(t))^{l-1} = 1 + (l-1)p(e^t - 1).
\end{align*}
\]

(7.15)

The coefficient matrix for this system:

\[
A := \begin{bmatrix}
1 & 1 & \cdots & 1 \\
\frac{z_1}{1} & \frac{z_2}{2} & \cdots & \frac{z_l}{l} \\
\frac{z_1^2}{1^2} & \frac{z_2^2}{2^2} & \cdots & \frac{z_l^2}{l^2} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{z_1^{l-1}}{1^{l-1}} & \frac{z_2^{l-1}}{2^{l-1}} & \cdots & \frac{z_l^{l-1}}{l^{l-1}}
\end{bmatrix}
\]

is a Vandermonde matrix of order \(l\). Its determinant, \(\det A = \prod_{1 \leq j < k \leq l}(z_k - z_j)\), is
non-zero when \( z_1, \ldots, z_l \) are distinct. Set

\[
A_j = \begin{bmatrix}
1 & \cdots & 1 & 1 & \cdots & 1 \\
z_1 & \cdots & z_{j-1} & 1 + p(e^{it} - 1) & z_{j+1} & \cdots & z_l \\
z_1^2 & \cdots & z_{j-1}^2 & 1 + 2p(e^{it} - 1) & z_{j+1}^2 & \cdots & z_l^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
z_1^{l-1} & \cdots & z_{j-1}^{l-1} & 1 + (l-1)p(e^{it} - 1) & z_{j+1}^{l-1} & \cdots & z_l^{l-1}
\end{bmatrix}, \quad j = 1, \ldots, l.
\]

(7.16)

\( A_j \) is just formed from the coefficient matrix \( A \) by replacing the \( j \)th column with the column vector of right hand sides of equation (7.15). By Cramer’s rule,

\[
c_j = \frac{\det A_j}{\det A} = \frac{h_j(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_l)}{\prod_{1 \leq i < j}(z_j - z_i) \prod_{k \leq i}(z_k - z_j)}, \quad j = 1, \ldots, l,
\]

(7.17)

where \( h_j \) is a polynomial of degree one in each of \( z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_l \).

To bound \( c_j \) we need some more information about the roots of equation (7.10), \( z_1, \ldots, z_l \).

**Lemma 15**

1. Let \( 0 < s \leq \frac{l-1}{l} \). If \( 0 < p < \frac{1}{2}s^{l-1}(1 - s) \) and \( t \in (-\pi, \pi] - \{0\} \) then the equation \( z^l - z^{l-1} - p(e^{it} - 1) = 0 \) has \( l \) distinct roots and \( l - 1 \) of them have modulus less than \( s \).

2. If \( p = \frac{r}{2}(l-1)^{-1} \), where \( 0 \leq r < 1 \), then there is a root \( z_1 \) of the equation \( z^l - z^{l-1} - p(e^{it} - 1) = 0 \) that satisfies

\[
|z_1 - 1| \leq \frac{r}{l}
\]

and

\[
|z_1 - (1 + p(e^{it} - 1))| \leq \frac{r}{1-r}|p(e^{it} - 1)|.
\]

Lemma 15 is proved in Appendix E. Recall that the bound in Lemma 14 is valid for \( p < \frac{l-1}{l} \) while Lemma 15 is useful for \( p < \frac{1}{2}(l-1)^{-1} \). For \( l \) greater than two
we have $\frac{l-1}{p^2} \leq \frac{1}{2}(\frac{l-1}{l})^{l-1}$, so we only consider $p < \frac{l-1}{l^2}$. For $0 < p < \frac{l-1}{l}$ and $t \in (-\pi, \pi] - \{0\}$, Lemma 15 tells us the roots $z_1(t), \ldots, z_l(t)$ of the characteristic equation (7.10) are distinct. In addition, if $p \ll \frac{1}{2}(\frac{l-1}{l})^{l-1}$ then one root, $z_1(t)$, is near 1 and separated by some distance from the other roots, which are near 0.

With the information in Lemma 15, we can return to equation (7.17) and bound the coefficients $c_j$. Getting a small bound for $c_1$ is important since $z_1$ is the biggest root. Each of the other roots has modulus less than $\frac{l-1}{l}$ so $|z_j|^k < (\frac{l-1}{l})^k$ for $j = 2, \ldots, l$. From (7.17) we have

$$c_1(t) = \frac{\det A_1}{\det A} \quad \text{and by (7.16),}$$

$$= \frac{1}{\det A} \left( \det \begin{bmatrix} 1 + p(e^{it} - 1) & z_2 & \cdots & z_l \\ (1 + p(e^{it} - 1))^2 & z_2^2 & \cdots & z_l^2 \\ \vdots & \vdots & \ddots & \vdots \\ (1 + p(e^{it} - 1))^{l-1} & z_2^{l-1} & \cdots & z_l^{l-1} \end{bmatrix} \right)$$

$$- \det \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 0 & z_2 & \cdots & z_l \\ p^2(e^{it} - 1)^2 & z_2^2 & \cdots & z_l^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=2}^{l-1} \binom{l-1}{k} p^k(e^{it} - 1)^k & z_2^{l-1} & \cdots & z_l^{l-1} \end{bmatrix}$$

$$= \frac{\prod_{j=2}^{l}(z_j - (1 + p(e^{it} - 1)))}{\prod_{j=2}^{l}(z_j - z_1)} - \frac{p^2(e^{it} - 1)^2 h(z_2, \ldots, z_l)}{\prod_{j=2}^{l}(z_j - z_1)}.$$  \hspace{1cm} (7.18)

In equation (7.18) $h$ is a polynomial in $z_2, \ldots, z_l$ of degree one in each variable and for small $p$ the second term on the right is small. The first term on the right is close to 1 because it can be written in the form

$$\prod_{j=2}^{l} \left( 1 + \frac{z_1 - (1 + p(e^{it} - 1))}{z_j - z_1} \right)$$

and by Lemma 15 $|z_1 - (1 + p(e^{it} - 1))|$ is small compared to $|z_j - z_1|$, for small $p$.

For $2 \leq j \leq l$, the expression for $c_j$ in equation (7.17) will have at least one factor
$(z_k - z_j)$ in the denominator with $k \geq 2$. Such a factor can get very small since $z_2(t), \ldots, z_l(t)$ all approach 0 as $t$ approaches 0, so rather than trying to bound $c_j$ for $2 \leq j \leq l$, we do some algebraic manipulation of the expression $c_2 z_2^k + \cdots + c_l z_l^k$ in (7.11) to show that its contribution to $\phi_k$ becomes negligible quickly as $k$ increases.

The scheme just outlined is used to prove Theorem 11. Details are in Appendix E. When $l > 2$, the bounds for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$ we get by this brute-force approach are disappointing in view of the bound in (7.1), suggested by numerical work. The bounds we get are valid for small $p$ only, $0 < p < \frac{l-1}{l}$, with the bound becoming bad as $p$ nears the upper end of the interval.
Chapter 8

Conclusion

This essay has attempted to illustrate the Stein method for approximation with PBD distributions. The application was to the problem of estimating the distribution of $W_n$.

We started by briefly looking at normal approximation of the distribution of $W_n$. It was shown that, for our problem, there is significant room for improvement in the error bound from Rinott’s theorem. In particular, Rinott’s theorem gives a bound that increases much too quickly as $p$, the pattern probability, decreases. Nevertheless, for $p$ values that are not too small and very large values of $n$, normal approximation can give useful results for events that are intervals. In fact the error is $O(n^{-1/2})$.

Approximation of the distribution of $W_n$ with PBD distributions is, in a sense, complementary to normal approximation. The error decreases as $p$ decreases and the PBD distributions become more difficult to evaluate as $np$ becomes very large. Poisson approximation gives good results for small $p$ values but the error does not decrease below $\frac{1}{52}(2l - 1)p$, as $n$ increases. The two-parameter PBD distributions $\pi_1(\alpha, \beta)$ and $\pi_2(r, q)$, with the parameter values specified in Theorem 5 and Theorem 8 respectively, approximate the distribution of $W_n$ with considerably smaller errors than Poisson approximation and, at least for small $p$, the errors are $O(n^{-1/2})$. 

66
Our error bounds for two-parameter PBD approximation involve the factor $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$. Thus good error bounds depend on good bounds for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$. Bounds for $d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_n + 1))$ that are $O(n^{-1/2})$ were found for patterns of length $l = 2$ and also for $l = 3$ and 4, but subject to $0 < p < (l - 1)/l$ when $l > 2$. These bounds are disappointing, especially those for $l > 2$. Computation suggests that for any value of $l$ and all possible values of $p$ one should aim for a bound like $C(np(1 - (2l - 1)p))^{-1/2}$, where $C$ is a number close to $1/2$. Perhaps one could get such a bound by working to prove (7.8). A less ambitious but reasonably good outcome would follow from proving (7.12).

An obvious extension of the problem studied here would be to consider a pattern that can overlap. Brown and Xia [5] have already considered negative binomial approximation of the number of pattern occurrences for a pattern of two identical symbols. When the pattern can overlap, the job of getting a small error bound is harder than the work in this essay since it is not just a matter of modifying methods that work for sums of independent indicators.
Appendix A

Non-overlapping patterns in $S^l$

A.1 Fraction of non-overlapping patterns

We shall find a lower bound for the fraction of non-overlapping patterns in $S^l$. Let $(s_1, \ldots, s_l)$ be an overlapping pattern in $S^l$ and suppose $t$ is the smallest number such that $s_1 \ldots s_t = s_{l-t+1} \ldots s_l$. $t$ satisfies $1 \leq t \leq \lfloor \frac{l}{2} \rfloor$. Here $\lfloor \frac{l}{2} \rfloor$ means the largest integer which is less than or equal to $l/2$.

Define $r_i$ to be the number of non-overlapping patterns in $S^i$, $i = 1, 2, \ldots$. There are $|S|^{l-2t} r_t$ patterns like $(s_1, \ldots, s_i)$. Thus the number of non-overlapping patterns of length $l$ is

$$r_l = |S|^l - \sum_{t=1}^{\lfloor \frac{l}{2} \rfloor} |S|^{l-2t} r_t$$

and

$$\frac{r_l}{|S|^l} = 1 - \sum_{t=1}^{\lfloor \frac{l}{2} \rfloor} |S|^{-t} \frac{r_t}{|S|^l}$$

$$> 1 - \sum_{t=1}^{\infty} |S|^{-t} = \frac{|S| - 2}{|S| - 1}.$$
A.2 Possible values of pattern probability $p$ for a non-overlapping pattern

Consider a non-overlapping pattern of length $l \geq 2$ and containing the distinct symbols $s_1, \ldots, s_k \in S$. $k$ is at least two. Suppose that $q := P\{Z_1 = s_i\}$ is the largest of the numbers:

$$P\{Z_1 = s_j\}, \quad j = 1, \ldots, k$$

and that the symbol $s_i$ occurs $m$ times in the pattern. Then

$$p \leq (1 - q)^{l-m} q^m.$$ 

Suppose $q \geq 1 - q$ (the same result will follow if we take the reverse inequality). Then

$$p \leq (1 - q)q^{l-1}.$$ 

The right hand side of the last equation has a maximum value when $q = \frac{l-1}{l}$. Thus

$$p \leq \frac{1}{l} (\frac{l-1}{l})^{l-1}.$$ 

The upper bound in this equation would be attained for a pattern like

$$(s_1, s_2, \ldots, s_2)$$

when $P\{Z_1 = s_1\} = \frac{1}{l}$ and $P\{Z_1 = s_2\} = \frac{l-1}{l}$. 

69
Appendix B

Bound for $\|\Delta f\|$ 

We shall show that

$$\|\Delta f\| := \sup_{i \geq 1} |f(i+1) - f(i)| \leq \frac{1}{rq(1-q)}.$$  \hspace{1cm} (B.1)

Here $f$ is defined as in equation 5.6:

$$f(i) = \begin{cases} g(i), & 0 \leq i \leq \lfloor r \rfloor + 1 \\ 0, & i \geq \lfloor r \rfloor + 2 \end{cases}$$ \hspace{1cm} (B.2)

and $g$ is the solution of the Stein equation in (3.5).

Recall from (5.1) and (5.2) that the birth rates are

$$\alpha_i = \begin{cases} q(r-i), & 0 \leq i \leq \lfloor r \rfloor \\ 0, & i > \lfloor r \rfloor \end{cases}$$

and the death rates are

$$\beta_i = (1-q)i, \quad i \geq 0.$$

By equation (5.13),

$$\|\Delta g\| := \sup_{i \geq 1} |g(i+1) - g(i)| \leq \sup_{i \geq 1} \min \left\{ \frac{1}{\alpha_i}, \frac{1}{\beta_i} \right\},$$ \hspace{1cm} (B.3)

where we take $\frac{1}{0} = \infty$.

For $1 \leq i \leq \lfloor r \rfloor$ we have

$$\min \left\{ \frac{1}{\alpha_i}, \frac{1}{\beta_i} \right\} = \min \left\{ \frac{1}{q(r-i)}, \frac{1}{(1-q)i} \right\} \leq \frac{1}{qr(1-q)}.$$
For $i > \lfloor r \rfloor$ we have

$$\min \left\{ \frac{1}{\alpha_i}, \frac{1}{\beta_i} \right\} = \frac{1}{(1-q)i} < \frac{1}{qr(1-q)}.$$  

Thus from (B.3),

$$\| \Delta g \| \leq \frac{1}{qr(1-q)}. \tag{B.4}$$

From the definition of $f$ in (B.2), the result in (B.1) will follow if

$$|g(\lfloor r \rfloor + 1)| \leq \frac{1}{qr(1-q)}. \tag{B.5}$$

There are two cases to consider. When $r$ is an integer we have $\alpha_i > 0$ for $0 \leq i \leq r-1$ and $\alpha_i = 0$ for $i \geq r$. Hence the number $m$ appearing in Lemma 6 is $r - 1$ and by equation (5.12) we have

$$|g(\lfloor r \rfloor + 1)| = |g(r + 1)| = \left| -\frac{1}{\beta_{r+1}}(I_A(r + 1) - \pi(A)) \right| \leq \frac{1}{(1-q)(r + 1)}. \tag{B.6}$$

When $r$ is not an integer, the number $m$ in Lemma 6 is $\lfloor r \rfloor$ and by equation (5.11) we have

$$|g(\lfloor r \rfloor + 1)| = \left| \frac{1}{\beta_{\lfloor r \rfloor + 1}\pi_{\lfloor r \rfloor + 1}}(\pi(A)\pi_{\lfloor r \rfloor + 1} - I_A(\lfloor r \rfloor + 1)\pi_{\lfloor r \rfloor + 1}) \right| \leq \frac{1}{(1-q)(\lfloor r \rfloor + 1)}. \tag{B.7}$$

The numbers on the right of (B.6) and (B.7) are less than $\frac{1}{qr(1-q)}$, hence (B.5) holds and so does (B.1).
Appendix C

Unimodality of $\mathcal{L}(W_n)$

In this section we show $\mathcal{L}(W_n)$ is unimodal.

For notational convenience define

$$p(n, k) := P\{W_n = k\}, \quad n, k \in \mathbb{Z}.$$  

As mentioned earlier, we interpret $W_n$ as $W_0 \equiv 0$ if $n < 0$. Thus $p(n, k)$ is interpreted as $p(0, k)$ if $n < 0$.

We shall use Lemma 3:

$$p(n, k) = p(n - 1, k) - p(p(n - l, k) - p(n - l, k - 1)), \quad n \geq 1, \quad k \in \mathbb{Z} \quad (C.1)$$

and another partial difference equation which can be derived from (C.1) but which can be found more simply as follows. For $n, k \geq 1$ we have

$$\{W_n = k\} = \bigcup_{i=1}^{n} \{W_n = k \text{ and the first pattern occurrence starts at } i\}$$

$$= \bigcup_{i=1}^{n} \{W_{i-l} = 0 \text{ and } J_i = 1 \text{ and } \sum_{j=i+l}^{n} J_j = k - 1\},$$

where we interpret $\sum_{j=i+l}^{n} J_j$ as 0 if $i + l > n$. Thus

$$p(n, k) = p\sum_{i=1}^{n} p(i - l, 0)p(n - i - l + 1, k - 1), \quad n, k \geq 1. \quad (C.2)$$
Define
\[ \Delta_1 p(n, k) := p(n + 1, k) - p(n, k), \]
i.e. \( \Delta_1 \) means forward difference in the first coordinate.

The following lemma is the key to proving \( \mathcal{L}(W_n) \) is unimodal.

**Lemma 16** For each \( k \geq 0 \),
\[ p(n + 1, k) - p(n, k) < 0 \quad \text{for some } n \geq 0 \]
implies
\[ p(m + 1, k) - p(m, k) < 0 \quad \text{for all } m \geq n. \]

The proof is by induction and relies on these properties of \( p(\cdot, 0) \):

(i) \( r(n) := \frac{p(n+1, 0)}{p(n, 0)}, \quad n \geq -l + 1 \), is non-increasing;
(ii) \( p(n, 0) > p(n + 1, 0) \) for \( n \geq 0 \).

These properties will be proved first and then we will prove Lemma 16.

**Proof of (i)** First note that \( p(m, 0) > 0 \) for all \( m > 0 \) because the event \( \{Z_1 = a_1, \ldots, Z_{m+l-1} = a_1\} \) always has positive probability and gives no occurrences of the non-overlapping pattern \((a_1, \ldots, a_l)\) in the string \((Z_1, \ldots, Z_{m+l-1})\) when \( l \geq 2 \). Hence \( r \) is well-defined.

For \(-l + 1 \leq n \leq -1\), \( \frac{p(n+1, 0)}{p(n, 0)} = 1 \) so for \( n \) in this range \( r(n) \) is non-increasing.

For \( n \geq 0 \), by (C.1):
\[
\frac{p(n + 1, 0)}{p(n, 0)} = 1 - p \frac{p(n - l + 1, 0)}{p(n, 0)} = 1 - p \frac{p(n - l + 1, 0) p(n - l + 2, 0)}{p(n - l + 2, 0) p(n - l + 3, 0) \cdots} \frac{p(n - 1, 0)}{p(n, 0)} \cdot
\]
\[ \text{\( l-1 \) factors} \quad (C.3) \]

For \( n = 0 \) we get:
\[
\frac{p(1, 0)}{p(0, 0)} = 1 - p \frac{p(-l + 1, 0)}{p(-l + 2, 0)} \cdots \frac{p(-1, 0)}{p(0, 0)} = 1 - p < \frac{p(0, 0)}{p(-1, 0)}.
\]
Suppose for some \( n \geq 0 \) we have
\[
\frac{p(-l+2,0)}{p(-l+1,0)} \geq \ldots \geq \frac{p(n,0)}{p(n-1,0)} \geq \frac{p(n+1,0)}{p(n,0)}.
\] (C.4)

Then
\[
\frac{p(n+2,0)}{p(n+1,0)} = 1 - p \frac{p(n-l+2,0) p(n-l+3,0) \ldots p(n-1,0)}{p(n-l+3,0) p(n-l+4,0) \ldots p(n,0)} \frac{p(n,0)}{p(n+1,0)}
\] (C.5)

and from (C.4) we have \( \frac{p(n,0)}{p(n+1,0)} \geq \frac{p(n-l+1,0)}{p(n-l+2,0)} \), so comparing (C.3) and (C.5) we see that
\[
\frac{p(n+2,0)}{p(n+1,0)} \leq 1 - p \frac{p(n-l+1,0) p(n-l+2,0) \ldots p(n-1,0)}{p(n-l+3,0) p(n-l+4,0) \ldots p(n,0)} = \frac{p(n+1,0)}{p(n,0)}.
\]

Hence by induction,
\[
\frac{p(-l+2,0)}{p(-l+1,0)} \geq \frac{p(-l+3,0)}{p(-l+2,0)} \geq \ldots \geq \frac{p(n+1,0)}{p(n,0)} \quad \text{for all } n \geq 0.
\]

Thus \( r \) is non-increasing.

Proof of (ii) Let \( n \geq 0 \). We have from (C.1):
\[
p(n+1,0) = p(n,0) - p p(n-l+1,0) < p(n,0).
\]

Proof of Lemma 16 For \( n \geq 0 \) and \( k \geq 0 \), from equation C.2 we get:
\[
p(n,k+1) = p \sum_{j=1}^{n} p(j-l,0) p(n-j-l+1,k) \quad \text{(if } n = 0 \text{ this sum is interpreted as } 0),
\] (C.6)

\[
p(n+1,k+1) = p \sum_{j=1}^{n+1} p(j-l,0) p(n-j-l+2,k),
\] (C.7)

and
\[
p(n+2,k+1) = p \sum_{j=1}^{n+2} p(j-l,0) p(n-j-l+3,k).
\] (C.8)
Fix \( k \geq 0 \). Suppose \( p(\cdot, k) \) is such that once it decreases it is decreasing thereafter, i.e.

\[
\Delta_1 p(i, k) < 0 \text{ implies } \Delta_1 p(j, k) < 0 \text{ for all } j \geq i.
\]

We will show \( p(\cdot, k + 1) \) has the same property. Suppose for some \( n \geq 0 \) we have (using equations (C.6) and (C.7)):

\[
\Delta_1 p(n, k + 1) = p \sum_{j=1}^{n} p(j - l, 0) \Delta_1 p(n - j - l + 1, k) + pp(n + 1 - l, 0)p(-l + 1, k) < 0. 
\] (C.9)

Then clearly \( n \) must be at least 1 and for some \( i, 1 \leq i \leq n \), we have

\[
\Delta_1 p(n - j - l + 1, k) < 0 \text{ for } 1 \leq j \leq i, \text{ and } \Delta_1 p(n - j - l + 1, k) \geq 0 \text{ for } i < j \leq n.
\]

We have, using equations (C.7) and (C.8):

\[
\Delta_1 p(n + 1, k + 1)
\]

\[
= pp(-l + 1, 0)\Delta_1 p(n - l + 1, k) + p \sum_{j=2}^{n+1} p(j - l, 0)\Delta_1 p(n - j - l + 2, k)
\]

\[
+ pp(n + 2 - l, 0)p(-l + 1, k)
\]

\[
= p\Delta_1 p(n - l + 1, k) + p \sum_{j=1}^{n} p(j + 1 - l, 0)\Delta_1 p(n - j - l + 1, k)
\]

\[
+ pp(n + 2 - l, 0)p(-l + 1, k)
\]

\[
= p\Delta_1 p(n - l + 1, k) + p \sum_{j=1}^{n} r(j - l)p(j - l, 0)\Delta_1 p(n - j - l + 1, k)
\]

\[
+ pr(n + 1 - l)p(n + 1 - l, 0)p(-l + 1, k)
\]

\[
< 0
\]

because \( \Delta_1 p(n - l + 1, k) < 0 \) and because of the inequality (C.9) and the fact \( r \) is non-increasing.

Since \( p(\cdot, 0) \) has the property that once it decreases it is decreasing thereafter, it follows by induction on \( k \) that \( p(\cdot, k) \) has this property for all \( k \geq 0 \). This completes the proof of Lemma 16.
Lemma 16 says that once $p(\cdot,k)$ decreases, it keeps on decreasing. An equivalent statement is: for each $k \geq 0$, $p(n+1,k) - p(n,k) \geq 0$ for some $n \geq 0$ implies $p(m+1,k) - p(m,k) \geq 0$ for all $m$ such that $0 \leq m \leq n$. We shall use this statement and equations (C.1) and (C.2) to prove the next theorem.

**Theorem 17** $L(W_n)$ is unimodal for $n \geq 0$.

**Proof** From equation (C.1) we have

$$-\frac{1}{p}(p(n+l,k+1) - p(n+l-1,k+1)) = p(n,k+1) - p(n,k), \text{ for } n \geq -l+1 \text{ and } k \geq 0.$$  

(C.10)

Fix $n \geq 1$ and suppose that for some $k \geq 0$ we have $p(n,k+1) - p(n,k) \leq 0$. Then from equation (C.10) it follows that $p(n+l,k+1) - p(n+l-1,k+1) \geq 0$ so $p(m,k+1) - p(m-1,k+1) \geq 0$ for $1 \leq m \leq n+l$, by Lemma 16. But by (C.10) this implies

$$p(m-l,k+1) - p(m-l,k) \leq 0 \text{ for } 1 \leq m \leq n+l.$$  

(C.11)

From equation (C.2) we have

$$p(n,k+1) = p \sum_{j=1}^{n} p(j-l,0)p(n-j-l+1,k)$$  

(C.12)

and

$$p(n,k+2) = p \sum_{j=1}^{n} p(j-l,0)p(n-j-l+1,k+1)$$  

(C.13)

and, by inequality (C.11), $p(m,k+1) \leq p(m,k)$ for $-l+1 \leq m \leq n$. Thus for each $j$ from 1 to $n$ the $j$th summand in the sum in equation (C.13) is less than or equal to the $j$th summand in the sum in equation (C.12). Hence $p(n,k+2) \leq p(n,k+1)$. Thus $L(W_n)$ is unimodal. This completes the proof of Theorem 17.
Appendix D

Proof of Theorem 10

Theorem 10

For a pattern of length \( l = 2 \),

\[
d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_{n+1})) \leq \begin{cases} 
\frac{9}{16} (p(1 - 4p))^{-1/2} n^{-1/2} + \frac{9}{8} (\frac{1}{\sqrt{2}})^n, & 0 < p < \frac{3}{32} \\
\frac{41}{80} (p(1 - 4p))^{-1/2} n^{-1/2} + \frac{41}{40} (\frac{1}{\sqrt{2}})^n + \frac{3}{5} n(\frac{5}{6})^n, & \frac{3}{32} \leq p \leq \frac{5}{32} \\
\frac{1}{2} (p(1 - 4p))^{-1/2} n^{-1/2} + (\frac{1}{\sqrt{2}})^n, & \frac{5}{32} < p < \frac{13}{64} \\
2n^{-1/2} + (\frac{1}{\sqrt{2}})^n, & \frac{13}{64} \leq p \leq \frac{1}{4}.
\end{cases}
\]

Proof

We prove the theorem using (7.4). For \( l = 2 \) equation (7.9) becomes

\[
\phi_{n+2}(t) - \phi_{n+1}(t) - p(e^{it} - 1)\phi_n(t) = 0, \quad n \geq 0.
\]

The corresponding characteristic equation for \( t \in (-\pi, \pi] - \{0\} \) is

\[
z^2 - z - p(e^{it} - 1) = 0,
\]

where \( z \) is understood to be a function of \( t \). Completing the square we have:

\[
(z - \frac{1}{2})^2 = \frac{1}{4} (1 + 4p(e^{it} - 1)).
\]

So the roots of the characteristic equation are

\[
z_1 = \frac{1}{2} \left( 1 + \sqrt{1 + 4p(e^{it} - 1)} \right), \quad z_2 = \frac{1}{2} \left( 1 - \sqrt{1 + 4p(e^{it} - 1)} \right)
\]
where $\pm \sqrt{1 + 4p(e^{it} - 1)}$ are the two distinct square roots of $1 + 4p(e^{it} - 1)$ when $(p, t) \neq (\frac{1}{8}, \pi)$. When $p = \frac{1}{8}$ and $t = \pi$ there is a single root, $z = \frac{1}{2}$, with multiplicity two. Therefore, provided $(p, t) \neq (\frac{1}{8}, \pi)$, the solution of (D.1) is

$$
\phi_n(t) = c_1(t)(z_1(t))^n + c_2(t)(z_2(t))^n \quad \text{for } n \geq 0, \tag{D.5}
$$

where $c_1(t)$ and $c_2(t)$ are determined by knowledge of $\phi_0(t)$ and $\phi_1(t)$.

When bounding $\frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_n(t)| \, dt$ in (7.4), our job is simplified a little by the fact that at least one of the roots in (D.4) has modulus not greater than $\frac{1}{\sqrt{2}}$ for all $t \in (-\pi, \pi]$ and $p \in (0, \frac{1}{4}]$. To see this firstly note that $4p(e^{it} - 1)$ lies in the closed disk with centre $(-1, 0)$ and radius 1 in the complex plane, i.e. $|4p(e^{it} - 1) + 1| \leq 1$. Hence $|\frac{1}{2}\sqrt{1 + 4p(e^{it} - 1)}| \leq \frac{1}{2}$ and for each $t$ at least one of the numbers $\pm \frac{1}{2}\sqrt{1 + 4p(e^{it} - 1)}$ has real part lying in the interval $[-\frac{1}{2}, 0]$. Thus at least one of the numbers $\frac{1}{2} \pm \frac{1}{2}\sqrt{1 + 4p(e^{it} - 1)}$ has real part lying in $[0, \frac{1}{2}]$ and imaginary part in $[-\frac{1}{2}, \frac{1}{2}]$ and such a complex number has modulus not greater than $\frac{1}{\sqrt{2}}$.

As we proceed with the proof of the theorem we shall use some inequalities which can be verified in a straightforward way using calculus:

$$
(1 + x)^{1/r} \leq \frac{1}{r}x + 1 \quad \text{for } r > 1 \text{ and } x \geq -1; \tag{D.6}
$$

$$
\ln(1 + x) \leq x \quad \text{for } x > -1; \tag{D.7}
$$

$$
\tan^{-1}x \leq x \quad \text{for } x \geq 0; \tag{D.8}
$$

$$
\sin x \leq x \quad \text{for } x \geq 0. \tag{D.9}
$$

Our first step will be to find bounds for the moduli of the roots $z_1$ and $z_2$ and bounds for the integrals $\frac{1}{2\pi} \int_{-\pi}^{\pi} |z_i(t)|^n \, dt, \quad i = 1, 2$. Then we shall find constant bounds for the moduli of $c_1$ and $c_2$ in (D.5).

From equation (D.3),

$$
|z - \frac{1}{2}| = \frac{1}{2} \sqrt{|1 + 4p(e^{it} - 1)|}.
$$
Since
\[ |1 + 4p(e^t - 1)| = \sqrt{1 - 2(4p)(1 - 4p)(1 - \cos t)}, \]
we have
\[
\sqrt{|1 + 4p(e^t - 1)|} = (1 - 2(4p)(1 - 4p)(1 - \cos t))^{1/4} \quad \text{and by (D.6)}
\leq 1 + \frac{1}{4}(-2(4p)(1 - 4p)(1 - \cos t)) = 1 - 2p(1 - 4p)(1 - \cos t).
\]
Thus
\[
|z| \leq |z - \frac{1}{2}| + \frac{1}{2}
\leq \frac{1}{2}(1 - 2p(1 - 4p)(1 - \cos t)) + \frac{1}{2}
= 1 - p(1 - 4p)(1 - \cos t). \quad (D.10)
\]
Now
\[
|z|^n \leq (1 - p(1 - 4p)(1 - \cos t))^n
= e^{n \ln(1 - p(1 - 4p)(1 - \cos t))} \quad \text{and by (D.7)}
\leq e^{n(-p(1 - 4p)(1 - \cos t))}
= e^{-np(1 - 4p)(1 - \cos t)}.
\]
Thus
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |z(t)|^n dt \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-np(1 - 4p)(1 - \cos t)} dt \quad \text{and by (7.5)}
\leq \frac{1}{2}(np(1 - 4p))^{-1/2}. \quad (D.11)
\]
The bound on the right of (D.11) becomes large as \( p \) approaches zero and as \( p \) approaches 1/4. The behaviour as \( p \) approaches zero is not a cause for great concern since this bound will be part of a bound for \( d_{TV}(\mathcal{L}(W_n), \mathcal{L}(W_{n+1})) \) and it will be multiplied by \( p \) where it is used, for example in Theorem 4.
The bound for $|z|$ in (D.10) is not good when $p$ is near $1/4$ so we shall derive another bound specifically for $p$ near $1/4$. Write $p = \frac{1}{4} - \varepsilon$, for $\varepsilon$ a small non-negative number. The characteristic equation (D.2) becomes

$$z^2 - z - \left(\frac{1}{4} - \varepsilon\right)(e^{it} - 1) = 0.$$  \hspace{1cm} (D.12)

The roots are

$$z = \frac{1}{2} \pm \sqrt{\frac{1}{4} - \varepsilon}e^{it} + \varepsilon$$

$$= \frac{1}{2} \left(1 \pm e^{it} \sqrt{1 - 4\varepsilon + 4\varepsilon e^{-it}}\right)$$  \hspace{1cm} (D.13)

where $\sqrt{1 - 4\varepsilon + 4\varepsilon e^{-it}}$ denotes the square root with positive real part.

We write $\sqrt{1 - 4\varepsilon + 4\varepsilon e^{-it}}$ in polar coordinates. By straightforward calculations, the modulus

$$|1 - 4\varepsilon + 4\varepsilon e^{-it}| = \sqrt{1 - 8\varepsilon(1 - 4\varepsilon)(1 - \cos t)}$$

so

$$\sqrt{|1 - 4\varepsilon + 4\varepsilon e^{-it}|} = \left(1 - 8\varepsilon(1 - 4\varepsilon)(1 - \cos t)\right)^{1/4}.$$  \hspace{1cm} (D.14)

We also have

$$\arg \left(\sqrt{1 - 4\varepsilon + 4\varepsilon e^{-it}}\right) = \frac{1}{2} \tan^{-1}\left(\frac{-4\varepsilon \sin t}{1 - 4\varepsilon(1 - \cos t)}\right).$$  \hspace{1cm} (D.15)

Using equations (D.14) and (D.15) we have

$$\frac{1}{2} \left(1 + e^{it/2} \sqrt{1 - 4\varepsilon + 4\varepsilon e^{-it}}\right)$$

$$= \frac{1}{2} \left[1 + (1 - 8\varepsilon(1 - 4\varepsilon)(1 - \cos t))^{1/4} \exp \left\{i \left(\frac{t}{2} + \frac{1}{2} \tan^{-1}\left(\frac{-4\varepsilon \sin t}{1 - 4\varepsilon(1 - \cos t)}\right)\right)\right\} \right].$$  \hspace{1cm} (D.16)

We will show that

$$\left|1 + (1 - 8\varepsilon(1 - 4\varepsilon)(1 - \cos t))^{1/4} \exp \left\{i \left(\frac{t}{2} - \frac{1}{2} \tan^{-1}\left(\frac{4\varepsilon \sin t}{1 - 4\varepsilon(1 - \cos t)}\right)\right)\right\}\right| \leq \sqrt{2} \sqrt{1 + \cos(\frac{t}{2})}$$
when $\varepsilon$ lies in the interval $[0, \frac{3}{64}]$. Since the quantities on both sides of the inequality are even functions we need only consider $t \in [0, \pi]$.

Let $\varepsilon \in [0, \frac{3}{64}]$ and $t \in [0, \pi]$. Set

$$\theta = \frac{1}{2} \tan^{-1}\left(\frac{4\varepsilon \sin t}{1 - 4\varepsilon (1 - \cos t)}\right),$$

which lies in the interval $[0, \frac{\pi}{4}]$. Then

$$|1 + (1 - 8\varepsilon(1 - 4\varepsilon)(1 - \cos t))^{1/4} \exp\{i(\frac{t}{2} - \theta)\}|$$

$$= \sqrt{1 + 2(1 - 8\varepsilon(1 - 4\varepsilon)(1 - \cos t))^{1/4} \cos\left(\frac{t}{2} - \theta\right) + (1 - 8\varepsilon(1 - 4\varepsilon)(1 - \cos t))^{1/2}}$$

$$\leq \sqrt{1 + 2(1 - 2\varepsilon(1 - 4\varepsilon)(1 - \cos t)) \cos\left(\frac{t}{2} - \theta\right) + 1 - 4\varepsilon(1 - 4\varepsilon)(1 - \cos t)}, \text{ by (D.6)}$$

$$= \sqrt{2^2 + 2(1 - 2\varepsilon(1 - 4\varepsilon)(1 - \cos t))(\cos(\frac{t}{2}) \cos \theta + \sin(\frac{t}{2}) \sin \theta) - 4\varepsilon(1 - 4\varepsilon)(1 - \cos t)}$$

$$\leq \sqrt{2^2 + 2(1 - 2\varepsilon(1 - 4\varepsilon)(1 - \cos t))(\cos(\frac{t}{2}) + \sin(\frac{t}{2}) \theta) - 4\varepsilon(1 - 4\varepsilon)(1 - \cos t)}, \text{ using (D.9)}$$

$$= \sqrt{2 + 2 \cos\left(\frac{t}{2}\right) - 4\varepsilon(1 - 4\varepsilon)(1 - \cos t)\left(1 + \cos\left(\frac{t}{2}\right)\right) + \sin\left(\frac{t}{2}\right) \tan^{-1}\left(\frac{4\varepsilon \sin t}{1 - 4\varepsilon(1 - \cos t)}\right)(1 - 2\varepsilon(1 - 4\varepsilon)(1 - \cos t))}$$

$$\leq \sqrt{2 + 2 \cos\left(\frac{t}{2}\right) - 4\varepsilon(1 - 4\varepsilon)(2\sin^2\left(\frac{t}{2}\right))(1 + \cos\left(\frac{t}{2}\right)) + \frac{4\varepsilon}{1 - 4\varepsilon(1 - \cos t)}(2\sin^2\left(\frac{t}{2}\right)) \cos\left(\frac{t}{2}\right)(1 - 2\varepsilon(1 - 4\varepsilon)(1 - \cos t))}$$

$$\leq \sqrt{2 + 2 \cos\left(\frac{t}{2}\right) - 4\varepsilon(1 - 4\varepsilon)(2\cos^2\left(\frac{t}{2}\right))(1 + \cos\left(\frac{t}{2}\right))}$$

$$\leq \sqrt{2 + 2 \cos\left(\frac{t}{2}\right)}$$

(D.17)

where in the last line we have used (D.8). Now for $t \in [0, \pi]$ and $0 \leq \varepsilon \leq \frac{3}{64}$ we have

$$(1 - 4\varepsilon)(1 + \cos\left(\frac{t}{2}\right)) \geq \frac{13}{16}(1 + \cos\left(\frac{t}{2}\right)) \geq \frac{13}{8} \cos\left(\frac{t}{2}\right) \quad \text{and}$$

$$\frac{(1 - 2\varepsilon(1 - 4\varepsilon)(1 - \cos t))}{1 - 4\varepsilon(1 - \cos t)} \cos\left(\frac{t}{2}\right) \leq \frac{8}{5} \cos\left(\frac{t}{2}\right)$$

so the sum of the last two terms under the square root sign in (D.17) is always non-positive. Hence

$$|1 + (1 - 8\varepsilon(1 - 4\varepsilon)(1 - \cos t))^{1/4} \exp\{i(\frac{t}{2} - \theta)\}| \leq \sqrt{2 + 2 \cos\left(\frac{t}{2}\right)},$$

and from equation (D.16) we have

$$\frac{1}{2} \left(1 + e^{i\frac{t}{2}} \sqrt{1 - 4\varepsilon + 4\varepsilon e^{-ut}}\right) \leq \frac{1}{\sqrt{2}} \sqrt{1 + \cos\left(\frac{t}{2}\right)}, \quad \text{for} \ \varepsilon \in [0, \frac{3}{64}].$$
Again using equations (D.14) and (D.15), we have
\[
\left| \frac{1}{2} \left( 1 - e^{i\pi} \sqrt{1 - 4\varepsilon + 4\varepsilon e^{-it}} \right) \right|
\]
\[
= \frac{1}{2} \left| 1 + (1 - 8\varepsilon)(1 - 4\varepsilon)(1 - \cos t)^{1/4} \exp\{i(\pi + \frac{t}{2} - \frac{1}{2}\tan^{-1}(\frac{4\varepsilon \sin t}{1 - 4\varepsilon(1 - \cos t)}))\} \right| \quad \text{(D.18)}
\]
\[
\leq \frac{1}{2} \sqrt{2},
\]
since the argument of the second term inside the modulus signs on the right of equation (D.18),
\[
\pi + \frac{t}{2} - \frac{1}{2}\tan^{-1}(\frac{4\varepsilon \sin t}{1 - 4\varepsilon(1 - \cos t)}),
\]
lies in \([\frac{\pi}{2}, \frac{3\pi}{2}]\) for \(t \in (-\pi, \pi]\). Thus for \(\varepsilon \in [0, \frac{3}{64}]\), i.e. \(p \in [\frac{13}{64}, \frac{1}{4}]\), one of the roots of the characteristic equation (D.12) satisfies
\[
|z| \leq \frac{1}{\sqrt{2}} \sqrt{1 + \cos(\frac{t}{2})} \quad \text{(D.19)}
\]
and the other has modulus not greater than \(\frac{1}{\sqrt{2}}\) for all \(t\).

We have
\[
\left( \frac{1}{\sqrt{2}} \sqrt{1 + \cos(\frac{t}{2})} \right)^n = (\frac{1}{2}(1 + \cos(\frac{t}{2})))^{n/2}
\]
\[
= (1 - \frac{1}{2}(1 - \cos(\frac{t}{2})))^{n/2}
\]
\[
= e^{\frac{n}{2} \ln(1 - \frac{1}{2}(1 - \cos(\frac{t}{2}))}
\]
\[
\leq e^{\frac{n}{2}(-\frac{1}{2}(1 - \cos(\frac{t}{2})))} \quad \text{by (D.7)}
\]
\[
= e^{-\frac{n}{2}(1 - \cos(\frac{t}{2}))}.
\]

It follows that
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \frac{1}{\sqrt{2}} \sqrt{1 + \cos(\frac{t}{2})} \right)^n dt \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\frac{n}{2}(1 - \cos(\frac{t}{2}))} dt
\]
\[
= \frac{2}{\pi} \int_{0}^{\pi} e^{-\frac{n}{2}(1 - \cos t)} dt
\]
\[
\leq \frac{2}{\pi} \int_{0}^{\pi} e^{-\frac{n}{2}(1 - \cos t)} dt, \quad \text{and by (7.5)}
\]
\[
\leq \left( \frac{n}{4} \right)^{-1/2} = 2n^{-1/2}. \quad \text{(D.20)}
\]
We now have bounds for the moduli of $z_1$ and $z_2$ and $\frac{1}{\pi} \int_{-\pi}^{\pi} |z_i(t)|^n \, dt, \ i = 1, 2.$

The next step is to get constant bounds for the moduli of $c_1$ and $c_2$ in (D.5).

By (D.5), for $(p, t) \neq (\frac{1}{8}, \pi)$:

$$\phi_n(t) = c_1(t) \left(\frac{1}{2} \left(1 + \sqrt{1 + 4p(e^{it} - 1)}\right)\right)^n + c_2(t) \left(\frac{1}{2} \left(1 - \sqrt{1 + 4p(e^{it} - 1)}\right)\right)^n, \ for \ n \geq 0.$$  

In particular, for $n = 0$ and 1 respectively, we have

$$\phi_0(t) = 1 = c_1(t) + c_2(t) \quad \text{(D.21)}$$

and

$$\phi_1(t) = 1 + p(e^{it} - 1)$$

$$= c_1(t) \left(\frac{1}{2} \left(1 + \sqrt{4p(e^{it} - 1) + 1}\right)\right) + (1 - c_1(t)) \left(\frac{1}{2} \left(1 - \sqrt{4p(e^{it} - 1) + 1}\right)\right) \quad \text{(D.22)}$$

Rearranging equation (D.22) we get, for $(p, t) \neq (\frac{1}{8}, \pi)$:

$$c_1(t) = \frac{2p(e^{it} - 1) + 1}{2\sqrt{4p(e^{it} - 1) + 1}} + \frac{1}{2} \quad \text{(D.23)}$$

and also, from $c_2(t) = 1 - c_1(t)$,

$$c_2(t) = \frac{1}{2} - \frac{2p(e^{it} - 1) + 1}{2\sqrt{4p(e^{it} - 1) + 1}} \quad \text{(D.24)}$$

We shall bound $|\frac{2p(e^{it} - 1) + 1}{\sqrt{4p(e^{it} - 1) + 1}}|$. We have

$$\frac{|2p(e^{it} - 1) + 1|^2}{|4p(e^{it} - 1) + 1|} = \frac{1 - 2(2p)(1 - 2p)(1 - \cos t)}{\sqrt{1 - 2(4p)(1 - 4p)(1 - \cos t)}}.$$  

Set

$$y(t) = \frac{1 - 2(2p)(1 - 2p)(1 - \cos t)}{\sqrt{1 - 2(4p)(1 - 4p)(1 - \cos t)}}.$$  

This is an even function with period $2\pi$ so we need only consider $t \in [0, \pi]$. Notice that for $p = \frac{1}{8}$, $y(t)$ goes to $\infty$ as $t$ approaches $\pi$. The derivative is

$$\frac{dy}{dt} = \frac{\frac{1}{2} \sin t(-16p^2 + 32p^2(1 - 2p)(1 - 4p)(1 - \cos t))}{(1 - 8p(1 - 4p)(1 - \cos t))^{3/2}}.$$  

83
For $t \in (0, \pi)$, the factor $\sin t$ in $\frac{dy}{dt}$ is positive. Thus, as $t$ increases from 0 to $\pi$, $\frac{dy}{dt}$ goes from 0 and becomes negative and then, if $64p^2(1 - 2p)(1 - 4p) > 16p^2$, i.e. $0 < p < \frac{3 - \sqrt{3}}{8}$, it eventually becomes positive. If $p \geq \frac{3 - \sqrt{3}}{8}$ then $\frac{dy}{dt}$ never becomes positive for $t \in [0, \pi]$. See Figure D.1. Hence for $0 < p < \frac{3 - \sqrt{3}}{8}$, $y$ has relative maxima at $t = 0$ and $t = \pi$, except that when $p = \frac{1}{8}$ then $y$ approaches $\infty$ as $t$ approaches $\pi$. For $\frac{3 - \sqrt{3}}{8} \leq p \leq \frac{1}{4}$, $y$ has a relative maximum at $t = 0$. We conclude that for all $t \in [0, \pi]$:

$$\sqrt{y(t)} = \left| \frac{2p(e^{it} - 1) + 1}{\sqrt{4p(e^{it} - 1) + 1}} \right| \leq \max \left\{ 1, \frac{1 - 4p}{\sqrt{|1 - 8p|}} \right\}, \quad \text{for } 0 < p \leq \frac{1}{4}. \quad (D.25)$$

Now for $p$ near $\frac{1}{8}$, the number $\frac{1 - 4p}{\sqrt{|1 - 8p|}}$ in (D.25) becomes very big. Recall that an unusual thing happens at $(p, t) = \left( \frac{1}{8}, \pi \right)$: the two roots of the characteristic equation (D.2) coincide and then $\phi_n(t)$ is not given by the expression in (D.5). This will not bother us because later we will get a better bound for $|\phi_n(t)|$ for $(p, t)$ near $(\frac{1}{8}, \pi)$. 

Figure D.1: $\frac{dy}{dt}$ on $[-\pi, \pi]$ for four values of $p$. 

For $t \in (0, \pi)$, the factor $\sin t$ in $\frac{dy}{dt}$ is positive. Thus, as $t$ increases from 0 to $\pi$, $\frac{dy}{dt}$ goes from 0 and becomes negative and then, if $64p^2(1 - 2p)(1 - 4p) > 16p^2$, i.e. $0 < p < \frac{3 - \sqrt{3}}{8}$, it eventually becomes positive. If $p \geq \frac{3 - \sqrt{3}}{8}$ then $\frac{dy}{dt}$ never becomes positive for $t \in [0, \pi]$. See Figure D.1. Hence for $0 < p < \frac{3 - \sqrt{3}}{8}$, $y$ has relative maxima at $t = 0$ and $t = \pi$, except that when $p = \frac{1}{8}$ then $y$ approaches $\infty$ as $t$ approaches $\pi$. For $\frac{3 - \sqrt{3}}{8} \leq p \leq \frac{1}{4}$, $y$ has a relative maximum at $t = 0$. We conclude that for all $t \in [0, \pi]$:

$$\sqrt{y(t)} = \left| \frac{2p(e^{it} - 1) + 1}{\sqrt{4p(e^{it} - 1) + 1}} \right| \leq \max \left\{ 1, \frac{1 - 4p}{\sqrt{|1 - 8p|}} \right\}, \quad \text{for } 0 < p \leq \frac{1}{4}. \quad (D.25)$$

Now for $p$ near $\frac{1}{8}$, the number $\frac{1 - 4p}{\sqrt{|1 - 8p|}}$ in (D.25) becomes very big. Recall that an unusual thing happens at $(p, t) = \left( \frac{1}{8}, \pi \right)$: the two roots of the characteristic equation (D.2) coincide and then $\phi_n(t)$ is not given by the expression in (D.5). This will not bother us because later we will get a better bound for $|\phi_n(t)|$ for $(p, t)$ near $(\frac{1}{8}, \pi)$. 

For $t \in (0, \pi)$, the factor $\sin t$ in $\frac{dy}{dt}$ is positive. Thus, as $t$ increases from 0 to $\pi$, $\frac{dy}{dt}$ goes from 0 and becomes negative and then, if $64p^2(1 - 2p)(1 - 4p) > 16p^2$, i.e. $0 < p < \frac{3 - \sqrt{3}}{8}$, it eventually becomes positive. If $p \geq \frac{3 - \sqrt{3}}{8}$ then $\frac{dy}{dt}$ never becomes positive for $t \in [0, \pi]$. See Figure D.1. Hence for $0 < p < \frac{3 - \sqrt{3}}{8}$, $y$ has relative maxima at $t = 0$ and $t = \pi$, except that when $p = \frac{1}{8}$ then $y$ approaches $\infty$ as $t$ approaches $\pi$. For $\frac{3 - \sqrt{3}}{8} \leq p \leq \frac{1}{4}$, $y$ has a relative maximum at $t = 0$. We conclude that for all $t \in [0, \pi]$:

$$\sqrt{y(t)} = \left| \frac{2p(e^{it} - 1) + 1}{\sqrt{4p(e^{it} - 1) + 1}} \right| \leq \max \left\{ 1, \frac{1 - 4p}{\sqrt{|1 - 8p|}} \right\}, \quad \text{for } 0 < p \leq \frac{1}{4}. \quad (D.25)$$

Now for $p$ near $\frac{1}{8}$, the number $\frac{1 - 4p}{\sqrt{|1 - 8p|}}$ in (D.25) becomes very big. Recall that an unusual thing happens at $(p, t) = \left( \frac{1}{8}, \pi \right)$: the two roots of the characteristic equation (D.2) coincide and then $\phi_n(t)$ is not given by the expression in (D.5). This will not bother us because later we will get a better bound for $|\phi_n(t)|$ for $(p, t)$ near $(\frac{1}{8}, \pi)$.
We will, however, use the following fact.

\[
\left| \frac{2p(e^{it} - 1) + 1}{\sqrt{4p(e^{it} - 1) + 1}} \right| \leq \max \left\{ 1, \frac{\sqrt{1 - 7p(1 - 2p)}}{(1 - 14p(1 - 4p))^{1/4}} \right\} \quad \text{for } 0 \leq t \leq \cos^{-1}(-\frac{3}{4}).
\]  
(D.26)

One can show using calculus that \( \frac{1 - 4p}{\sqrt{|1 - 8p|}} \) does not exceed \( \frac{3}{4} \) for \( 0 < p < \frac{3}{32} \) and it does not exceed \( \frac{3}{4} \) for \( \frac{5}{32} < p \leq \frac{1}{4} \). Likewise one can show that \( \frac{\sqrt{1 - 7p(1 - 2p)}}{(1 - 14p(1 - 4p))^{1/4}} \leq \frac{21}{20} \) for \( \frac{3}{32} \leq p \leq \frac{5}{32} \). Putting this information together with inequalities (D.25) and (D.26) and the expressions for \( c_1 \) and \( c_2 \) in equations (D.23) and (D.24) we get

\[
|c_1| \quad \text{and} \quad |c_2| \leq \begin{cases} 
  \frac{1}{2} + \frac{1}{2} \left( \frac{5}{32} \right) = \frac{9}{32}, & p \in (0, \frac{3}{32}) \quad \text{and} \quad t \in (-\pi, \pi); \\
  \frac{1}{2} + \frac{1}{2} (1) = 1, & p \in (\frac{5}{32}, \frac{1}{4}] \quad \text{and} \quad t \in (-\pi, \pi]; \\
  \frac{1}{2} + \frac{1}{2} \left( \frac{21}{20} \right) = \frac{41}{40}, & p \in [\frac{3}{32}, \frac{5}{32}] \quad \text{and} \quad |t| \leq \cos^{-1}(-\frac{3}{4}). 
\end{cases}
\]  
(D.27)

Note that for \( p \in [\frac{3}{32}, \frac{5}{32}] \), our bound for \( |c_1| \) and \( |c_2| \) in (D.27) only holds for \( \cos t \geq -\frac{3}{4} \). We now look at bounding \( |\phi_n(t)| \) for \( p \in [\frac{3}{32}, \frac{5}{32}] \) and \( \cos t < -\frac{3}{4} \). It will be shown that for \( p \in [\frac{3}{32}, \frac{5}{32}] \) and \( \cos t < -\frac{3}{4} \) both roots of the characteristic equation, \( z^2 - z - p(e^{it} - 1) = 0 \), have moduli less than \( \frac{5}{6} \). How can we use this fact?

For \( (p, t) \neq (\frac{1}{8}, \pi) \), by (D.4) and (D.5) and (D.23) and (D.24) we have:

\[
\phi_n(t) = \left( \frac{2p(e^{it} - 1) + 1}{2\sqrt{4p(e^{it} - 1) + 1}} + \frac{1}{2} \right) (z_1(t))^n + \left( \frac{1}{2} - \frac{2p(e^{it} - 1) + 1}{2\sqrt{4p(e^{it} - 1) + 1}} \right) (z_2(t))^n
\]

\[
= \frac{1}{2} [(z_1(t))^n + (z_2(t))^n] + \frac{2p(e^{it} - 1) + 1}{2\sqrt{4p(e^{it} - 1) + 1}} [(z_1(t))^n - (z_2(t))^n]
\]

\[
= \frac{1}{2} [(z_1(t))^n + (z_2(t))^n] + \frac{2p(e^{it} - 1) + 1}{2} \sum_{k=0}^{n-1} (z_1(t))^{n-1-k}(z_2(t))^k. \quad \text{(D.28)}
\]

So for \( p \in [\frac{3}{32}, \frac{5}{32}] \) and \( \cos t < -\frac{3}{4} \), but \( (p, t) \neq (\frac{1}{8}, \pi) \), we have

\[
|\phi_n(t)| \leq \frac{1}{2} [|(z_1(t))^n| + |(z_2(t))^n|] + \frac{3}{5} n \left( \frac{5}{6} \right)^n. \quad \text{(D.29)}
\]

Now to prove both roots of the characteristic equation have moduli less than \( \frac{5}{6} \)
when \( p \in \left[ \frac{3}{32}, \frac{5}{32} \right] \) and \( \cos t < -\frac{3}{4} \). Define
\[
f(z) = z^2 - z, \quad z \in \mathbb{C}.
\]
We have \( f(z) + \frac{1}{4} = (z - \frac{1}{2})^2 \). Thus \( f(z) + \frac{1}{4} \) has \( \frac{1}{2} \) as a zero with multiplicity two.

Consider the circle in the \( z \)-plane centred at \( \frac{1}{2} \) and with positive radius \( r \):
\[
\gamma(t) = re^{it} + \frac{1}{2}.
\]
We have
\[
|f(\gamma(t)) + \frac{1}{4}| = |r^2e^{2it}| = r^2.
\]
By Lemma 12, for any complex number \( w \) such that \( |w + \frac{1}{4}| < r^2 \) the polynomials \( f(z) - w \) and \( f(z) + \frac{1}{4} \) have the same number of zeros, counting multiplicities, lying inside \( \gamma \). For \( f(z) + \frac{1}{4} \) this number is two so if \( |w + \frac{1}{4}| < r^2 \) then \( f(z) - w \) has two zeros, counting multiplicities, inside \( \gamma \).

If \( |p(e^{it} - 1) + \frac{1}{4}| < r^2 \) we have that both roots of \( z^2 - z - p(e^{it} - 1) = 0 \) lie inside the circle centred at \( \frac{1}{2} \) and with radius \( r \). Take \( r = \frac{1}{3} \). We claim that for \( p \in \left[ \frac{3}{32}, \frac{5}{32} \right] \) and \( \cos t \leq -\frac{3}{4} \) we have \( |p(e^{it} - 1) + \frac{1}{4}| < \frac{1}{9} \). After some algebraic manipulation we get
\[
|p(e^{it} - 1) + \frac{1}{4}| = \sqrt{2p(p - \frac{1}{4})(1 - \cos t) + \frac{1}{16}} \leq \sqrt{2(\frac{3}{32})(\frac{3}{32} - \frac{1}{4})(1 + \frac{3}{4}) + \frac{1}{16}} = \sqrt{\frac{23}{2048}} < \frac{1}{9}, \quad p \in \left[ \frac{3}{32}, \frac{5}{32} \right] \text{ and } \cos t \leq -\frac{3}{4}.
\]
So for \( p \in \left[ \frac{3}{32}, \frac{5}{32} \right] \) and \( \cos t \leq -\frac{3}{4} \), the two roots of \( z^2 - z - p(e^{it} - 1) = 0 \) lie inside the circle centred at \( \frac{1}{2} \) and with radius \( \frac{1}{3} \), which implies that each has modulus less than \( \frac{5}{6} \).

We can finally write down bounds for \( \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_n(t)| \, dt \).

For \( p \in (0, \frac{3}{32}) \), by (D.5) and inequalities (D.10) and (D.27) we have:
\[
|\phi_n(t)| \leq \frac{6}{5}(1 - p(1 - 4p)(1 - \cos t))^n + \left( \frac{1}{\sqrt{2}} \right)^n.
\]
So, by (D.11) we have
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_n(t)| \, dt \leq \frac{9}{16} (np(1 - 4p))^{-1/2} + \frac{9}{8} (\frac{1}{\sqrt{2}})^n
\]
which is the first result in the statement of Theorem 10.

For \( p \in [\frac{3}{32}, \frac{5}{32}] \) and \((p, t) \neq (\frac{1}{8}, \pi)\), equation (D.5) and inequalities (D.10), (D.27) and (D.29) give:
\[
|\phi_n(t)| \leq \frac{41}{40} ((1 - p(1 - 4p)(1 - \cos t))^n + (\frac{1}{\sqrt{2}})^n) + \frac{3}{5} n (\frac{3}{8})^n.
\]
So by (D.11) we have
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_n(t)| \, dt \leq \frac{41}{40} (np(1 - 4p))^{-1/2} + \frac{41}{40} (\frac{1}{\sqrt{2}})^n + \frac{3}{5} n (\frac{3}{8})^n
\]
which is the second result in the statement of Theorem 10.

For \( p \in (\frac{5}{32}, \frac{13}{64}) \) we have by equation (D.5) and inequalities (D.10) and (D.27):
\[
|\phi_n(t)| \leq (1 - p(1 - 4p)(1 - \cos t))^n + (\frac{1}{\sqrt{2}})^n.
\]
So by (D.11) we have
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_n(t)| \, dt \leq \frac{1}{2} (np(1 - 4p))^{-1/2} + (\frac{1}{\sqrt{2}})^n,
\]
which is the third result in the statement of Theorem 10.

Finally, for \( p \in [\frac{13}{64}, \frac{1}{4}] \) equation (D.5) and inequalities (D.19) and (D.27) give:
\[
|\phi_n(t)| \leq \left( \frac{1}{\sqrt{2}} \sqrt{1 + \cos(\frac{t}{2})} \right)^n + (\frac{1}{\sqrt{2}})^n
\]
so by (D.20) we have
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_n(t)| \, dt \leq 2n^{-1/2} + (\frac{1}{\sqrt{2}})^n
\]
which is the last result in the statement of Theorem 10. This completes the proof of Theorem 10.
Appendix E

Proofs of results for $l > 2$

In this chapter we prove some results culminating in the proof of Theorem 11.

Lemma 13
All roots of $z^l - z^{l-1} - p(e^{it} - 1) = 0$ lie inside the circle in the complex plane with centre $\frac{1}{l}$ and radius $\frac{l-1}{l}$ when $0 < p < \frac{l-1}{l}$ and $t \in (-\pi, \pi] - \{0\}$.

Proof We have

$$z^l - z^{l-1} - p(e^{it} - 1) = (z - \frac{1}{l})^l - \frac{\sum_{k=1}^{l-1} \binom{l-1}{k} \left(\frac{k}{l}\right)^k \left(\frac{1}{k+1}\right)}{(z - \frac{1}{l})^{l-1} - p(e^{it} - 1)}.$$ 

We must show all zeros of

$$h(w) := w^l - \sum_{k=1}^{l-1} \binom{l-1}{k} \left(\frac{1}{l}\right)^k \left(\frac{k}{k+1}\right) w^{l-1-k} - p(e^{it} - 1)$$

lie inside the circle centred at 0 and with radius $\frac{l-1}{l}$ when $0 < p < \frac{l-1}{l}$ and $t \in (-\pi, \pi] - \{0\}$.

Define the polynomial function

$$g(w) := w^l,$$

and the circle

$$\beta(\theta) = \frac{l-1}{l} e^{i\theta}, \quad -\pi < \theta \leq \pi.$$ 

g has $l$ zeros, counting multiplicity, inside $\beta$. 

88
We have
\[ |g(\beta(\theta))| = \left(\frac{l-1}{l}\right)^l, \quad \text{for all } \theta \in (-\pi, \pi]. \]

We shall show
\[ |h(\beta(\theta)) - g(\beta(\theta))| < \left(\frac{l-1}{l}\right)^l, \quad \text{for all } \theta \in (-\pi, \pi] \quad \text{(E.2)} \]
when \(0 < p < \frac{l-1}{l}\) and \(t \in (-\pi, \pi] - \{0\}\). Then Rouché’s theorem in Lemma 12 tells us \(h\) has \(l\) zeros, counting multiplicities, inside \(\beta\).

We have
\[ |h(\beta(\theta)) - g(\beta(\theta))| = \left| \sum_{k=1}^{l-1} \left( \begin{array}{c} l-1 \\ k \end{array} \right) \left( \frac{1}{l} \right)^k \left( \frac{k}{k+1} \right) \left[ \left( \frac{l-1}{l} \right) e^{i\theta} \right]^{l-1-k} + p(e^{it} - 1) \right| \]
\[ = \left| \sum_{k=1}^{l-1} a_k e^{i(l-1-k)\theta} + p(e^{it} - 1) \right| \quad \text{(E.3)} \]
where \(a_k = \left( l - 1 \right) \left( \frac{1}{l} \right)^k \left( \frac{k}{k+1} \right) \left( \frac{l-1}{l} \right)^{l-1-k} \).

Note that \(\sum_{k=1}^{l-1} a_k = \left( \frac{l-1}{l} \right)^l\). Expanding the right hand side of (E.3) and rearranging we get:

\[ |h(\beta(\theta)) - g(\beta(\theta))| \]
\[ = \sqrt{\left( \sum_{k=1}^{l-1} a_k \cos((l-1-k)\theta) + p\cos t - 1 \right)^2 + \left( \sum_{k=1}^{l-1} a_k \sin((l-1-k)\theta) + p\sin t \right)^2} \]
\[ = \sqrt{\left( \frac{l-1}{l} \right)^{2l} - 2p(1 - \cos t) \left( \frac{l-1}{l} \right) - p \sum_{1 \leq k < j \leq l-1} a_ka_j(1 - \cos((j-k)\theta)) \]
\[ + 2p \sum_{k=1}^{l-2} a_k(\cos((l-1-k)\theta) - t - \cos((l-1-k)\theta)). \]

For \(t \in (-\pi, \pi] - \{0\}\) and \(0 < p < \frac{l-1}{l}\) the second term under the square root is
negative. To establish inequality (E.2) it suffices to show

\[
\sum_{1 \leq k < j \leq l-1} a_k a_j (1 - \cos((j - k)\theta)) \geq p \sum_{k=1}^{l-2} a_k (\cos((l - 1 - k)\theta - t) - \cos((l - 1 - k)\theta)).
\]

(E.4)

We have

\[
p \sum_{k=1}^{l-2} a_k (\cos((l - 1 - k)\theta - t) - \cos((l - 1 - k)\theta))
\]

\[
\leq p \sum_{k=1}^{l-2} a_k (1 - \cos((l - 1 - k)\theta))
\]

\[
\leq a_{l-1} \sum_{k=1}^{l-2} a_k (1 - \cos((l - 1 - k)\theta)), \quad \text{if } p < a_{l-1} = \frac{l-1}{l}
\]

\[
\leq \sum_{1 \leq k < j \leq l-1} a_j a_k (1 - \cos((j - k)\theta))
\]

which is inequality (E.4). Thus (E.2) holds when \(0 < p < \frac{l-1}{l}\) and \(t \in (-\pi, \pi) - \{0\}\). This completes the proof of Lemma 13.

**Lemma 14**

All roots of \(z^l - z^{l-1} - p(e^{it} - 1) = 0\) satisfy

\[
|z(t)| \leq 1 - \frac{1}{l} \left( \frac{l}{l-1} \right)^{l-1} p \left( 1 - p/(l-1) \right) (1 - \cos t)
\]

when \(0 \leq p < \frac{l-1}{l}\).

**Proof** The equation \(z^l - z^{l-1} - p(e^{it} - 1) = 0\) can be written as

\[
\left( z - \frac{1}{l} \right)^l - \sum_{k=1}^{l-1} \binom{l-1}{k} \left( \frac{1}{l} \right)^k \left( \frac{k}{k+1} \right) \left( z - \frac{1}{l} \right)^{l-1-k} - p(e^{it} - 1) = 0. \quad (E.5)
\]

Set \(w = z - \frac{1}{l}\). Then equation (E.5) becomes

\[
w^l - \sum_{k=1}^{l-1} \binom{l-1}{k} \left( \frac{1}{l} \right)^k \left( \frac{k}{k+1} \right) w^{l-1-k} - p(e^{it} - 1) = 0. \quad (E.6)
\]
Lemma 13 established that all roots of equation (E.6) have modulus not greater than \( \frac{l-1}{l} \) when \( 0 \leq p < \frac{l-1}{l} \). The first step is to get a better bound for roots of equation (E.6). Let \( 0 \leq p < \frac{l-1}{l} \). We can rearrange equation (E.6):

\[
w^l = \sum_{k=1}^{l-1} \left( \frac{l-1}{k} \right) \left( \frac{1}{l} \right)^k \left( \frac{k}{k+1} \right)^{l-1-k} w^{l-1-k} + p(e^{it} - 1).
\]

Thus roots satisfy:

\[
|w|^l \leq \sum_{k=1}^{l-2} \left( \frac{l-1}{k} \right) \left( \frac{1}{l} \right)^k \left( \frac{k}{k+1} \right) \left( \frac{l-1}{l} \right)^{l-1-k} + \frac{l-1}{p^l} + p(e^{it} - 1),
\]

which implies

\[
|w| \leq \left( \sum_{k=1}^{l-2} \left( \frac{l-1}{k} \right) \left( \frac{1}{l} \right)^k \left( \frac{k}{k+1} \right) \left( \frac{l-1}{l} \right)^{l-1-k} + \frac{l-1}{p^l} - p \left( 1 - p/(l-1) \right) (1 - \cos t) \right)^{1/l}
\]

\[
= \left( \left( \frac{l-1}{l} \right)^l - p \left( 1 - p/(l-1) \right) (1 - \cos t) \right)^{1/l}
\]

\[
\leq \frac{l-1}{l} \left( 1 - \frac{1}{l} \left( \frac{l}{l-1} \right)^l p \left( 1 - p/(l-1) \right) (1 - \cos t) \right). \tag{E.7}
\]

In deriving this inequality we twice used: \( (1 + x)^{1/r} \leq 1 + \frac{1}{r} x \) for \( x \geq -1 \) and \( r > 1 \).

Now to get a bound for roots, \( z \), of equation (E.5). Using (E.7) we get:

\[
|z| \leq |w| + \frac{1}{l} \leq 1 - \frac{1}{l} \left( \frac{l}{l-1} \right)^l p \left( 1 - p/(l-1) \right) (1 - \cos t).
\]

This completes the proof of Lemma 14.

**Lemma 15**

i) Let \( 0 < s \leq \frac{l-1}{l} \). If \( 0 < p < \frac{1}{2}s^{l-1}(1 - s) \) and \( t \in (-\pi, \pi) - \{0\} \) then \( z^l - z^{l-1} - p(e^{it} - 1) = 0 \) has \( l \) distinct roots and \( l-1 \) of them have modulus less than \( s \).

ii) If \( p = \frac{r}{l} \left( \frac{l-1}{l} \right)^{l-1}(\frac{1}{t}) \) where \( 0 \leq r < 1 \), then there is a root \( z_1 \) of \( z^l - z^{l-1} - p(e^{it} - 1) = 0 \) that satisfies

\[
|z_1 - 1| \leq \frac{r}{t} \quad \text{and} \quad |z_1 - (1 + p(e^{it} - 1))| \leq \frac{r}{1-t} |p(e^{it} - 1)|.
\]

91
Proof  i) Define \( f(z) = z^l - z^{l-1} \). For any given complex number \( w \), repeated zeros of \( f(z) - w \) must also be zeros of \( f'(z) = z^{l-2}(lz - (l - 1)) \); there are only the two possibilities: \( z = 0 \) or \( \frac{l-1}{l} \). Hence \( z^l - z^{l-1} - w = 0 \) has no repeated roots if \( w \neq 0 \) and \( w \neq -\frac{1}{l}(\frac{l-1}{l})^{l-1} \).

Let \( 0 < s \leq \frac{l-1}{l} \) and \( 0 < p < \frac{1}{2} s^{l-1}(1-s) \) and \( t \in (-\pi, \pi) - \{0\} \).

Define the circle
\[
\gamma(\theta) = se^{i\theta}, \quad \theta \in (-\pi, \pi].
\]

We have
\[
|f(\gamma(\theta))| = s^{l-1}|se^{i\theta} - 1| \geq s^{l-1}(1 - s),
\]

and
\[
|p(e^{it} - 1)| \leq 2p < s^{l-1}(1 - s).
\]

Hence, by Rouché’s theorem (Lemma 12), \( f(z) - p(e^{it} - 1) \) and \( f(z) \) have the same number of zeros, counting multiplicities, inside \( \gamma \). Thus \( z^l - z^{l-1} - p(e^{it} - 1) = 0 \) has \( l - 1 \) roots with modulus less than \( s \) and since \( p(e^{it} - 1) \) is not 0 or \( -\frac{1}{l}(\frac{l-1}{l})^{l-1} \), these roots are distinct.

ii) Let \( p = \frac{r}{2} \left( \frac{l-1}{l} \right)^{l-1} \left( \frac{1}{l} \right) \) where \( 0 \leq r < 1 \).

Define
\[
X := \{ z \in \mathbb{C} : |z - 1| \leq \frac{r}{l} \}.
\]

Define \( g : X \to X \) by:
\[
g(z) = 1 + \frac{p(e^{it} - 1)}{z^{l-1}}, \quad z \in X.
\]

\( g \) is well defined because if \( z \in X \) then the modulus of \( z \) is greater than \( \frac{l-1}{l} \) and
\[
\left| \frac{p(e^{it} - 1)}{z^{l-1}} \right| < \frac{r}{2} \left( \frac{l-1}{l} \right)^{l-1} \left( \frac{1}{l} \right) (2) \left( \frac{l-1}{l} \right)^{l-1} = \frac{r}{l}.
\]
Let $X$ be a complete metric space. We will show $g$ is a contraction map. From this it will follow that $g$ has a unique fixed point $z_1$ which satisfies

$$z_1 = 1 + \frac{p(e^{it} - 1)}{z_{l-1}};$$

that is, there is a unique root of $z^l - z^{l-1} - p(e^{it} - 1) = 0$ in $X$. Let $z$ and $w$ be elements of $X$. We have

$$|g(z) - g(w)| = \left| p(e^{it} - 1) \left( \frac{1}{z^{l-1}} - \frac{1}{w^{l-1}} \right) \right|$$

$$= \left| p(e^{it} - 1) \frac{w^{l-1} - z^{l-1}}{z^{l-1}w^{l-1}} \right| |z - w|$$

$$\leq \frac{r}{2} \left( \frac{l-1}{l} \right)^{l-1} \left( \frac{1}{7} \right) (2)(l-1) \frac{1}{(l-1)!} |z - w| = r|z - w|.$$

Since $r < 1$, $g$ is a contraction map. Let $g^k$ denote the composition of $k$ $g$’s for $k = 1, 2, 3, \ldots$. By standard results for contraction maps, for any $z \in X$ the sequence $(g^k(z))$ converges to the fixed point $z_1$ and

$$|z_1 - g^k(z)| \leq \frac{r^k}{1 - r}|g(z) - z|, \quad k = 1, 2, 3, \ldots \quad (E.8)$$

For $z = 1$ and $k = 1$ inequality (E.8) becomes

$$|z_1 - (1 + p(e^{it} - 1))| \leq \frac{r}{1 - r}|p(e^{it} - 1)|.$$

This completes the proof of Lemma 15.

**Theorem 11**

For patterns of length $l = 3$ or $4$ we have

$$d_{TV}(\mathcal{L}(W_k), \mathcal{L}(W_k + 1)) \leq \begin{cases} 0.77(p(1 - \frac{27}{2}p))^{-1/2}k^{-1/2} + (0.33 + 0.61k)(\frac{1}{3})^k, \quad l = 3 \text{ and } 0 < p \leq \frac{1}{27}; \\ 0.68(p(1 - \frac{256}{3}p))^{-1/2}k^{-1/2} + (0.04 + 0.05k + 0.1k^2)(\frac{1}{3})^k, \quad l = 4 \text{ and } 0 < p < \frac{3}{256}. \end{cases}$$
Proof We prove the result for \( l = 4 \); the proof for \( l = 3 \) is very similar.

Let \( l = 4 \) and \( 0 < p < \frac{3}{256} \). We firstly use Lemma 15 to get some estimates.

Since \( p < \frac{3}{256} < \frac{3}{243} = \frac{1}{2}(\frac{3}{4})^3(\frac{2}{3}) \), we have

\[
|z_j| < \frac{1}{3} \quad \text{for } j = 2, 3 \text{ and } 4. \tag{E.9}
\]

Also \( p < \frac{21}{64}(\frac{3}{4})^3(\frac{1}{4}) \) so

\[
|z_1 - (1 + p(e^{it} - 1))| \leq \frac{2}{7}|p(e^{it} - 1)| \leq \frac{4}{7}p. \tag{E.10}
\]

It follows that

\[
|z_1 - 1| \leq |z_1 - (1 + p(e^{it} - 1))| + |p(e^{it} - 1)| \leq \frac{2}{7}p + 2p < \frac{1}{32}
\]

and

\[
|z_1 - z_j| > \frac{61}{96} \quad \text{for } j = 2, 3 \text{ and } 4. \tag{E.11}
\]

We use equation (7.18) to get an estimate for \( |c_1(t)| \):

\[
c_1(t) = \prod_{j=2}^{4} \left( 1 + \frac{z_1 - (1 + p(e^{it} - 1))}{z_j - z_1} \right) + p^2(e^{it} - 1)^2(3 + p(e^{it} - 1) - (z_2 + z_3 + z_4)) \frac{(z_2 - z_1)(z_3 - z_1)(z_4 - z_1)}{(z_2 - z_1)(z_3 - z_1)(z_4 - z_1)} \tag{E.12}
\]

so

\[
|c_1(t)| \leq \prod_{j=2}^{4} \left( 1 + \frac{|z_1 - (1 + p(e^{it} - 1))|}{|z_j - z_1|} \right) + \frac{|p^2(e^{it} - 1)^2||2 + z_1 + p(e^{it} - 1)||}{|z_2 - z_1||z_3 - z_1||z_4 - z_1|} \leq (1 + \frac{4}{7}p(\frac{96}{61})^3 + 4p^2(3)(\frac{96}{61})^3), \quad \text{by (E.10) and (E.11)} \leq 1.04. \tag{E.13}
\]

Set \( c = c_2 + c_3 + c_4 \). We have

\[
c(t) = 1 - c_1(t), \quad \text{by the first row of equation (7.15)}
\]

94
\[
\begin{align*}
&= - \left( (z_1 - (1 + p(e^{it} - 1))) \sum_{j=2}^{4} \frac{1}{z_j - z_1} 
+ (z_1 - (1 + p(e^{it} - 1)))^2 \right. \\
&\quad \times \left( \frac{1}{(z_2 - z_1)(z_3 - z_1)} + \frac{1}{(z_2 - z_1)(z_4 - z_1)} + \frac{1}{(z_3 - z_1)(z_4 - z_1)} \right) \\
&\quad + \frac{(z_1 - (1 + p(e^{it} - 1)))^3}{(z_2 - z_1)(z_3 - z_1)(z_4 - z_1)} + p^2(e^{it} - 1)^2(2 + z_1 + p(e^{it} - 1)) \\
&\left. + \frac{(z_1 - (1 + p(e^{it} - 1)))^3}{(z_2 - z_1)(z_3 - z_1)(z_4 - z_1)} + p^2(e^{it} - 1)^2(2 + z_1 + p(e^{it} - 1)) \right),
\end{align*}
\]

by equation (E.12). Thus

\[
|c(t)| < 3(\frac{4}{7}p)^3(\frac{61}{96})^3 + 3(\frac{4}{7}p)^2(\frac{61}{96})^2 + 4(\frac{4}{7}p)^3(\frac{61}{96})^3 + 4p^2(3)(\frac{61}{96})^3, \text{ by (E.10) and (E.11)}
\]

\[
< 0.04. \tag{E.14}
\]

We write (7.11) as:

\[
\phi_k(t) = c_1(t)(z_1(t))^k + c(t)(z_2(t))^k + c_3(t)((z_3(t))^k - (z_2(t))^k) + c_4(t)((z_4(t))^k - (z_2(t))^k),
\]

and we have bounds for \(c_1\) and \(c\) so the next step is to bound

\[
|c_3(z_3^k - z_2^k) + c_4(z_4^k - z_2^k)|.
\]

By equation (7.17) we have:

\[
c_3 = \frac{1}{(z_4 - z_3)(z_3 - z_2)(z_4 - z_1)} \left[ z_4 z_1 z_2 - (1 + p(e^{it} - 1))(z_4 z_2 + z_4 z_1 + z_1 z_2) \\
+ (1 + 2p(e^{it} - 1))(z_4 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \right]
\]

and

\[
c_4 = \frac{-1}{(z_4 - z_3)(z_4 - z_2)(z_4 - z_1)} \left[ z_1 z_2 z_3 - (1 + p(e^{it} - 1))(z_3 z_2 + z_3 z_1 + z_1 z_2) \\
+ (1 + 2p(e^{it} - 1))(z_3 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \right].
\]

Thus
\[ c_3 (z_3^k - z_2^k) + c_4 (z_4^k - z_2^k) \]

\[ = \frac{1}{z_4 - z_3} \left\{ \frac{1}{z_3 - z_1} \left[ z_4 z_1 z_2 - (1 + p(e^{it} - 1))(z_4 z_2 + z_4 z_1 + z_1 z_2) \right. \\
+ (1 + 2p(e^{it} - 1))(z_4 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \left. \right] \sum_{j=0}^{k-1} (z_3^{k-1-j} z_2^j) \right\} - \frac{1}{z_4 - z_1} \left[ z_1 z_2 z_3 - (1 + p(e^{it} - 1))(z_3 z_2 + z_3 z_1 + z_1 z_2) \right. \\
+ (1 + 2p(e^{it} - 1))(z_3 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \left. \right] \sum_{j=0}^{k-1} (z_4^{k-1-j} z_2^j) \right\} \\
+ \left( \frac{1}{z_3 - z_1} \left[ z_4 z_1 z_2 - (1 + p(e^{it} - 1))(z_4 z_2 + z_4 z_1 + z_1 z_2) \right. \\
+ (1 + 2p(e^{it} - 1))(z_4 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \left. \right] \right) - \frac{1}{z_4 - z_1} \left[ z_1 z_2 z_3 - (1 + p(e^{it} - 1))(z_3 z_2 + z_3 z_1 + z_1 z_2) \right. \\
+ (1 + 2p(e^{it} - 1))(z_3 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \left. \right] \sum_{j=0}^{k-1} (z_3^{k-1-j} z_2^j) \right\} \right. \\
\times \left. \sum_{j=0}^{k-1} (z_3^{k-1-j} z_2^j) \right\}. \tag{E.16} \]

Our aim now is to rearrange the expression inside large parentheses in equation (E.16) so that a factor of \( z_4 - z_3 \) comes out. Using the equation

\[ \frac{1}{z_3 - z_1} = \frac{1}{z_4 - z_1} + \frac{z_4 - z_3}{(z_4 - z_1)(z_3 - z_1)} \]

we get:

\[ \frac{1}{z_3 - z_1} \left[ z_4 z_1 z_2 - (1 + p(e^{it} - 1))(z_4 z_2 + z_4 z_1 + z_1 z_2) \right. \\
+ (1 + 2p(e^{it} - 1))(z_4 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \left. \right] \\
- \frac{1}{z_4 - z_1} \left[ z_1 z_2 z_3 - (1 + p(e^{it} - 1))(z_3 z_2 + z_3 z_1 + z_1 z_2) \right. \\
+ (1 + 2p(e^{it} - 1))(z_3 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \left. \right] \\
= \frac{z_4 - z_3}{(z_4 - z_1)(z_3 - z_1)} \left[ z_4 z_1 z_2 - (1 + p(e^{it} - 1))(z_4 z_2 + z_4 z_1 + z_1 z_2) \right. \\
+ (1 + 2p(e^{it} - 1))(z_4 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \left. \right] \\
+ \frac{z_4 - z_3}{(z_4 - z_1)(z_3 - z_1)} \left[ z_1 z_2 z_3 - (1 + p(e^{it} - 1))(z_3 z_2 + z_3 z_1 + z_1 z_2) \right. \\
+ (1 + 2p(e^{it} - 1))(z_3 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \left. \right] \right. \\
\times \left. \sum_{j=0}^{k-1} (z_3^{k-1-j} z_2^j) \right\}. \]
Substituting this in equation (E.16) and cancelling the factor $z_4 - z_3$ from numerator and denominator, we arrive at:

$$c_3(z_3^k - z_2^k) + c_4(z_4^k - z_2^k)$$

$$= -\frac{1}{z_4 - z_1} \left[ z_1 z_2 z_3 - (1 + p(e^{it} - 1))(z_3 z_2 + z_3 z_1 + z_1 z_2) ight. + (1 + 2p(e^{it} - 1))(z_3 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \right]$$

$$\times \sum_{j=0}^{k-2} \sum_{m=0}^{k-2-j} z_3^{k-2-j-m} z_4^m z_2^j$$

$$+ \frac{1}{z_4 - z_1} \left[ z_1 z_2 - (1 + p(e^{it} - 1))(z_1 + z_2) + (1 + 2p(e^{it} - 1)) \right] \sum_{j=0}^{k-1} z_3^{k-1-j} z_2^j$$

$$+ \frac{1}{(z_4 - z_1)(z_3 - z_1)} \left[ z_4 z_1 z_2 - (1 + p(e^{it} - 1))(z_4 z_2 + z_4 z_1 + z_1 z_2) ight. + (1 + 2p(e^{it} - 1))(z_4 + z_2 + z_1) - (1 + 3p(e^{it} - 1)) \right]$$

$$\times \sum_{j=0}^{k-1} z_3^{k-1-j} z_2^j.$$

(E.17)

The expressions in square brackets in equation (E.17) can be bounded more easily when written as follows:

$$z_1 z_2 z_3 - (1 + p(e^{it} - 1))(z_3 z_2 + z_3 z_1 + z_1 z_2)$$

$$+ (1 + 2p(e^{it} - 1))(z_3 + z_2 + z_1) - (1 + 3p(e^{it} - 1))$$

$$= (z_1 - (1 + p(e^{it} - 1)))(z_2 - (1 + p(e^{it} - 1)))(z_3 - (1 + p(e^{it} - 1)))$$

$$+ p^2(e^{it} - 1)^2(3 + p(e^{it} - 1) - (z_1 + z_2 + z_3)),$$

and

$$z_1 z_2 - (1 + p(e^{it} - 1))(z_2 + z_1) + (1 + 2p(e^{it} - 1))$$

$$= (z_1 - (1 + p(e^{it} - 1)))(z_2 - (1 + p(e^{it} - 1))) - p^2(e^{it} - 1)^2.$$

From equation (E.17) we thus get:
\[ |c_3(z_3^k - z_2^k) + c_4(z_4^k - z_2^k)| \]
\[ \leq \frac{1}{|z_4 - z_1|} \left( |z_1 - (1 + p(e^{it} - 1))||z_2 - (1 + p(e^{it} - 1))||z_3 - (1 + p(e^{it} - 1))| \right) \]
\[ + \frac{1}{|z_4 - z_1|} |z_2 - (1 + p(e^{it} - 1))| + |p^2(e^{it} - 1)^2| \left( \sum_{j=0}^{k-1} z_3^{k-1-j} z_2^j \right) \]
\[ = \frac{96}{61} (\frac{1}{4} p)^2 + 4p^2(2\frac{1}{3}) \frac{k-1}{2} (\frac{1}{3})^{k-2} + \frac{96}{61} (\frac{1}{4} p)^2 (\frac{1}{3})^{k-1} \]
\[ + (\frac{96}{61})^2 (\frac{1}{4} p)^2 + 4p^2(2\frac{1}{3}) \frac{1}{k} (\frac{1}{3})^{k-1}, \quad \text{by (E.9), (E.10) and (E.11)} \]
\[ \leq 0.1k(k + \frac{1}{2})(\frac{1}{3})^k. \quad \text{(E.18)} \]

Using equation (E.15) and inequalities (E.9), (E.13), (E.14), (E.18), (7.13) and (7.14) we have

\[ \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi_k(t)| \, dt \]
\[ \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( |c_1(t)| |z_1(t)|^k + |c(t)| |z_2(t)|^k + |c_3(z_3^k - z_2^k) + c_4(z_4^k - z_2^k)| \right) \, dt \]
\[ \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( 1.04 \left( 1 - \frac{16}{27} p(1 - \frac{256}{3} p)(1 - \cos t) \right)^k + 0.04(\frac{1}{3})^k + 0.1k(k + \frac{1}{2})(\frac{1}{3})^k \right) \, dt \]
\[ \leq 0.68 \left( p(1 - \frac{256}{3} p) \right)^{-1/2} k^{-1/2} + (0.04 + 0.05k + 0.1k^2)(\frac{1}{3})^k. \]

This completes the proof of Theorem 11 for \( l = 4 \).
Bibliography


