On the Locality of the Prüfer Code

Craig Lennon

Department of Mathematics
United States Military Academy
218 Thayer Hall
West Point, NY 10996
craigtlennon@gmail.com

Submitted: Feb 21, 2008; Accepted: Dec 22, 2008; Published: Jan 23, 2009
Mathematics Subject Classification: 05D40

Abstract

The Prüfer code is a bijection between trees on the vertex set $[n]$ and strings on the set $[n]$ of length $n-2$ (Prüfer strings of order n). In this paper we examine the 'locality' properties of the Prüfer code, i.e. the effect of changing an element of the Prüfer string on the structure of the corresponding tree. Our measure for the distance between two trees T, T^* is $\Delta(T, T^*) = n - 1 - |E(T) \cap E(T^*)|$. We randomly mutate the μth element of the Prüfer string of the tree T, changing it to the tree T^*, and we asymptotically estimate the probability that this results in a change of ℓ edges, i.e. $P(\Delta = \ell \,|\, \mu)$. We find that $P(\Delta = \ell \,|\, \mu)$ is on the order of $n^{-1/3+o(1)}$ for any integer $\ell > 1$, and that $P(\Delta = 1 \,|\, \mu) = (1 - \mu/n)^2 + o(1)$. This result implies that the probability of a 'perfect' mutation in the Prüfer code (one for which $\Delta(T, T^*) = 1$) is 1/3.

1 Introduction

The Prüfer code is a bijection between trees on the vertex set $[n] := \{1, \ldots, n\}$ and strings on the set $[n]$ of length $n-2$ (which we will refer to as P-strings). If we are given a tree T, we encode T as a P-string as follows: at step i $(1 \le i \le n-2)$ of the encoding process the lowest number leaf is removed, and its neighbor is recorded as p_i, the ith element of the P-string

$$P = (p_1, \ldots, p_{n-2}), \quad p_i \in [n], \quad (1 \le i \le n-2).$$

We will describe a decoding algorithm in a moment.

First we observe that the Prüfer code is one of many methods of representing trees as numeric strings, [4], [7], [8]. A representation with the property that small changes in the representation lead to small changes in the represented object is said to have high locality, a desirable property when the representation is used in a genetic algorithm [2], [7]. The distance between two numeric string tree representations is the number of elements in the string which differ, and the distance between two trees T, T^* is measured by the number of edges in one tree which are not in the other:

$$\Delta = \Delta^{(n)} = \Delta^{(n)}(T, T^*) := n - 1 - |E(T) \cap E(T^*)|,$$

where $E(T)$ is the edge set of tree T.

By a mutation in the P-string we mean the change of exactly one element of the P-string. Thus we denote the set of all ordered pairs of P-strings differing in exactly one coordinate (the mutation space) by $\mathcal{M}$, and by $\mathcal{M}_\mu$ we mean the subset of the mutation space in which the P-strings differ in the μ th coordinate:

$$\mathcal{M} = \bigcup_{1=\mu}^{n-2} \mathcal{M}_\mu, \quad \mathcal{M}_\mu := \left\{(P, P^*) \,:\, p_i = p_i^* \text{ for } i \neq \mu, \text{ and } p_\mu \neq p_\mu^*\right\},$$

where

$$P = (p_1, \ldots, p_{n-2}), \quad P^* = (p_1^*, \ldots, p_{n-2}^*),$$

so $|\mathcal{M}| = n^{n-2}(n-2)(n-1)$, and $|\mathcal{M}_\mu| = n^{n-2}(n-1)$. We choose a pair $(P, P^*) \in \mathcal{M}$ uniformly at random, and the random variable Δ measures the distance between the trees corresponding to (P, P^*). Using $\mathrm{P}\left(\{\text{event}\}|\circ\right)$ to denote conditional probability, we have

$$\begin{aligned}
\mathrm{P}\left(\Delta = \ell\right) &= \sum_{\mu=1}^{n-2} \mathrm{P}\left(\Delta = \ell \,|\, (P, P^*) \in \mathcal{M}_\mu\right) \mathrm{P}\left((P, P^*) \in \mathcal{M}_\mu\right) \\
&= \sum_{\mu=1}^{n-2} \mathrm{P}\left(\Delta = \ell \,|\, (P, P^*) \in \mathcal{M}_\mu\right) \frac{1}{n-2}.
\end{aligned}$$

Hereafter we will represent the event $(P, P^*) \in \mathcal{M}_\mu$ by μ, as in

$$\mathrm{P}\left(\{\text{event}\} \,|\, \mu\right) := \mathrm{P}\left(\{\text{event}\} \,|\, (P, P^*) \in \mathcal{M}_\mu\right).$$

Computer assisted experiments conducted by Thompson (see [8] page 195-196) for trees with a vertex size as large as $n = 100$ led him to conjecture that:

$$\lim_{n\to\infty} \mathrm{P}\left(\Delta^{(n)} = 1\right) = \frac{1}{3}, \tag{1.1}$$

and that if $\mu/n \to \alpha$, then

$$\lim_{n\to\infty} \mathrm{P}\left(\Delta^{(n)} = 1 \,\middle|\, \mu\right) = (1-\alpha)^2. \tag{1.2}$$

In a recent paper [6], Paulden and Smith use combinatorial and numerical methods to develop conjectures about the exact value of $\mathrm{P}\left(\Delta = \ell \,|\, \mu\right)$ for $\ell = 1, 2$, and about the generic form that $\mathrm{P}\left(\Delta = \ell \,|\, \mu\right)$ would take for $\ell > 2$. These conjectures, if true, would prove (1.1)-(1.2). Unfortunately, the formulas representing the exact value of $\mathrm{P}\left(\Delta = \ell \,|\, \mu\right)$

are complicated, even for $\ell = 1, 2$, and the proof of their correctness may be difficult. In this paper we will show by a probabilistic method that (1.1)-(1.2) are indeed correct, proving that

$$\mathrm{P}\left(\Delta^{(n)} = 1 \,\middle|\, \mu\right) = (1 - \mu/n)^2 + O\left(n^{-1/3} \ln^2 n\right), \tag{1.3}$$

and showing in the process that

$$\mathrm{P}\left(\Delta^{(n)} = \ell \,\middle|\, \mu\right) = O\left(n^{-1/3} \ln^2 n\right), \qquad (\ell > 1). \tag{1.4}$$

Of course (1.3) implies (1.1), because $\int_0^1 (1-\alpha)^2 \, d\alpha = 1/3$. In order to prove these results we will need to analyze the following P-string decoding algorithm, which we learned of from [1], [6].

1.1 A Decoding Algorithm

In the decoding algorithm, the P-string $P = (p_1, \ldots, p_{n-2})$ is read from right to left, so we begin the algorithm at step $n-2$ and count down to step 0. We begin a generic step i with a tree T_{i+1} which is a subgraph of the tree T which was encoded as P. This tree has vertex set V_{i+1} of cardinality $n-i-1$ and edge set E_{i+1} of cardinality $n-i-2$. We will add to T_{i+1} a vertex from $X_{i+1} := [n] \setminus V_{i+1}$, and an edge, and the resulting tree T_i will contain T_{i+1} as a subgraph. The vertex added at step i of the decoding algorithm is the vertex which was removed at step $i+1$ of the encoding algorithm, and will be denoted by y_i. A formal description of the decoding algorithm is given below.

Decoding Algorithm

Input: $P = (p_1, \ldots, p_{n-2})$ and $X_{n-1} = [n-1]$, $V_{n-1} = \{n\}$, $E_{n-1} = \emptyset$.
Step i ($1 \le i \le n-2$): We begin with the set X_{i+1} and a tree T_{i+1} having vertex set V_{i+1} and edge set E_{i+1}. We examine entry p_i of P.

1. If $p_i \in X_{i+1}$, then set $y_i = p_i$.

2. If $p_i \notin X_{i+1}$, then let $y_i = \max X_{i+1}$ (the largest element of X_{i+1}).

In either case we add y_i to the tree T_{i+1}, joining it by an edge to the vertex p_{i+1} (which must already be a vertex of T_{i+1}), with $p_{n-1} := n$. So $X_i = X_{i+1} \setminus \{y_i\}, V_i = V_{i+1} \cup \{y_i\}$, and $E_i = E_{i+1} \cup \{\, \{y_i, p_{i+1}\} \,\}$.

Step 0: We add y_0, the only vertex in X_1, and the edge $\{y_0, p_1\}$ to the tree T_1 to form the tree $T_0 = T$.

In this algorithm, we do not need to know the values of $p_1, \ldots, p_i$ until after step $i+1$. We will take advantage of this by using the principle of deferred decisions. With μ fixed, we will begin with $p_{\mu+1}, \ldots, p_{n-2}$ determined, but with $p_1, \ldots, p_\mu$, as yet undetermined.

We will then choose the values of the p_i for $1 \le i \le \mu$ when the algorithm requires those values and no sooner.

This will mean that the composition of the sets X_i, V_i, E_i will only be determined once we have conditioned on $p_i, \ldots, p_{n-2}$. When we compute the probability that p_{i-1} is in a set $\mathcal{A}_i$ whose elements are determined by p_j, $j > i$, (for example X_i or V_i) we are implicitly using the law of total probability:

$$\mathrm{P}\left(p_{i-1} \in \mathcal{A}_i \mid \mu\right) = \sum_{P_i} \mathrm{P}\left(p_{i-1} \in \mathcal{A}_i \mid P_i\,;\, \mu\right) \mathrm{P}\left(P_i \mid \mu\right),$$

where the sum above is over all P-sub-strings $P_i = (p_i, \ldots, p_{n-2})$ of the appropriate length, and $\mathrm{P}\left(P_i \mid \mu\right)$ is the probability of entries i through $n-2$ of the P-string taking the values $(p_i, \ldots, p_{n-2})$. We will leave such conditioning as implicit when estimating probabilities of the type $\mathrm{P}\left(p_{i-1} \in \mathcal{A}_i \mid \mu\right)$.

In the next section, we will use the principle of deferred decisions to easily find a lower bound for $\mathrm{P}\left(\Delta = 1 \mid \mu\right)$, and in later sections we will use similar techniques to establish asymptotically sharp upper bounds for $\mathrm{P}\left(\Delta = 1 \mid \mu\right)$, as well as for $\mathrm{P}\left(\Delta = \ell \mid \mu\right)$ $(\ell > 1)$. The combination of these bounds will prove (1.3)-(1.4).

2 The lower bound

For a fixed value of μ, we will construct a pair of strings from $\mathcal{M}_\mu$, starting our construction with two partial strings

$$P_{\mu+1} = \left(p_{\mu+1}, \ldots, p_{n-2}\right), \qquad P^*_{\mu+1} = \left(p^*_{\mu+1}, \ldots, p^*_{n-2}\right), \quad p_j = p^*_j,$$

where p_j has been selected uniformly at random from $[n]$ for $\mu + 1 \le j \le n - 2$. We have not yet chosen p_j, p^*_j for $j \le \mu$. We run the decoding algorithm from step $n - 2$ down through step $\mu + 1$, and at this point we have two trees $T_{\mu+1} = T^*_{\mu+1}$ as which $P_{\mu+1} = P^*_{\mu+1}$ have been partially decoded. Of course we also have the sets $V_{\mu+1} = V^*_{\mu+1}$ and $X_{\mu+1} = X^*_{\mu+1}$, where

$$V_i := \{j \,:\, j \text{ is a vertex of } T_i\}, \quad V^*_i := \{j \,:\, j \text{ is a vertex of } T^*_i\},$$

and $X_i = [n] \setminus V_i, X^*_i = [n] \setminus V^*_i$. We let E_i, E^*_i represent the edge sets of T_i, T^*_i.

Now we choose p_μ and $p^*_\mu \ne p_\mu$, and execute step μ of the decoding algorithm. There are two possibilities:

1. If both $p_\mu, p^*_\mu \in V_{\mu+1} \cup \{\max X_{\mu+1}\}$, then $y_i = y^*_i = \max X_{\mu+1}$. We have added the same vertex and the same edge (y_i and $\{y_i, p_{\mu+1}\}$) to both $T_{\mu+1}$ and $T^*_{\mu+1}$. We have $V_\mu = V^*_\mu$ and $E_\mu = E^*_\mu$.

2. One of p_μ, p^*_μ is not an element of the set $V_{\mu+1} \cup \{\max X_{\mu+1}\}$.

We will denote the first of these two events by

$$\mathcal{E} := \{\text{both } p_\mu, p_\mu^* \in V_{\mu+1} \cup \{\max X_{\mu+1}\}\}, \tag{2.1}$$

and we will show that on this event, $\Delta = 1$ no matter what values of $p_j = p_j^*$ $(1 \le j \le \mu-1)$ we choose to complete the strings P, P^*. Thus

$$\mathcal{E} \subseteq \{\Delta = 1\} \Longrightarrow \mathrm{P}(\mathcal{E} \mid \mu) \le \mathrm{P}(\Delta = 1 \mid \mu).$$

Let us prove the set containment shown in the previous line.

Proof. Suppose that event $\mathcal{E}$ occurs, so that $V_\mu = V_\mu^*$ and $X_\mu = X_\mu^*$, and $T_\mu = T_\mu^*$. Now choose $p_1, \dots, p_{\mu-1}$ uniformly at random from $[n]$, with $p_i^* = p_i$ for $1 \le i \le \mu - 1$.

At steps $\mu-1, \mu-2, \dots, 0$ of the algorithm, we will, at every step, read the same entry $p_i = p_i^*$ from the strings P, P^*. Because $X_\mu = X_\mu^*$ and $p_{\mu-1} = p_{\mu-1}^*$, the algorithm demands that we add to T_μ, T_μ^* the same vertex $y_{\mu-1} = y_{\mu-1}^*$. This in turn means that $X_{\mu-1} = X_{\mu-1}^*$. In a similar fashion, for $0 \le i \le \mu - 2$ we have

$$X_{i+1} = X_{i+1}^* \Longrightarrow y_i = y_i^*.$$

Thus at every step $i \le \mu$ of the algorithm we add the same vertex to V_{i+1}, V_{i+1}^*. Furthermore, at every step we are adding the edge $\{y_i, p_{i+1}\}$ to E_{i+1} and the edge $\{y_i, p_{i+1}^*\}$ to E_{i+1}^*. Since $p_i = p_i^*$ for $i \ne \mu$ and $p_\mu \ne p_\mu^*$, we add the same edge to T_{i+1} and T_{i+1}^* at every step except at step $\mu-1$ at which we add $\{y_{\mu-1}, p_\mu\}$ to T_μ and $\{y_{\mu-1}, p_\mu^*\}$ $(\ne \{y_{\mu-1}, p_\mu\})$ to T_μ^*. Of course the same edge cannot be added to a tree twice, so at no point could we have added $\{y_{\mu-1}, p_\mu^*\}$ to T or $\{y_{\mu-1}, p_\mu\}$ to T^*. Thus T and T^* must have exactly $n-2$ edges in common, and

$$\Delta = \Delta^{(n)}(T, T^*) := n - 1 - |E(T) \cap E(T^*)| = 1.$$

$\square$

Note: We have proved that if $X_k = X_k^*$ for $k \le \mu$ then $X_j = X_j^*$ for all $j < k$, that the same vertex is added at every step $j < k$, and that the same edge is added at every step $j < \min\{k,\, \mu - 1\}$. We will need this result later.

Now we bound the conditional probability of event $\mathcal{E}$. Because there are $n - \mu - 1$ elements in the set $V_{\mu+1} \cup \{\max X_{\mu+1}\}$, we have

$$\begin{aligned}
\mathrm{P}\left(\Delta = 1 \mid \mu\right) \ge \mathrm{P}\left(\mathcal{E} \mid \mu\right) &= \frac{n-\mu-1}{n} \cdot \frac{n-\mu-2}{n-1} \\
&= 1 - \frac{2\mu}{n} + \frac{\mu^2}{n^2} + O\left(n^{-1}\right) \\
&= (1 - \mu/n)^2 + O\left(n^{-1}\right).
\end{aligned}$$

Of course $\mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{E} \mid \mu\right) = 0$ for $\ell > 1$, so in order to prove (1.3)-(1.4) it remains to show that

$$\mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{E}^c \mid \mu\right) = O\left(n^{-1/3} \ln^2 n\right), \quad (\ell \ge 1). \tag{2.2}$$

This endeavor will prove more complicated than the upper bounds, so we will need to establish some preliminary results and make some observations which will prove useful later.

3 Observations and preliminary results

Recall that after step j of the decoding algorithm we have two sets X_j, X_j^* of vertices which have not been placed in T_j, T_j^*. For $j \geq \mu+1$, we know that $X_j = X_j^*$, but we may have $X_j \neq X_j^*$ for $j \leq \mu$. So let us consider then the set $\mathcal{X}_j := X_j \cup X_j^*$.

Our goal is to show that either $\mathcal{X}_j = X_j$, or $\mathcal{X}_j$ consists of $X_j \cap X_j^*$ and of two additional vertices, one in $V_j \setminus V_j^*$ and one in $V_j^* \setminus V_j$. This means $\mathcal{X}_j$ has the following form:

$$\begin{aligned}\mathcal{X}_j :=& \{x_1 < \cdots < x_a < \min\{z_j, z_j^*\} < x_{a+1} < \cdots < x_{a+b} < \\ & \qquad \max\{z_j, z_j^*\} < x_{a+b+1} < \cdots < x_{a+b+c}\},\end{aligned} \tag{3.1}$$

where

$$z_j \in V_j \setminus V_j^*, \quad z_j^* \in V_j^* \setminus V_j, \quad x_i \in X_j \cap X_j^*, \ (1 \leq i \leq a+b+c),$$

and $a, b, c \geq 0$, with $a+b+c = j-1$. Let us also take the opportunity to define

$$\mathcal{V}_j := V_j \cap V_j^*,$$

and note that

$$|\mathcal{V}_j| = n-j \quad (\text{if } \{z_j, z_j^*\} = \emptyset), \qquad |\mathcal{V}_j| = n-j-1 \quad (\text{if } |\{z_j, z_j^*\}| = 2). \tag{3.2}$$

We will consider a set $\mathcal{X}_j = X_j$ to also have the form shown in (3.1), but with $\{z_j, z_j^*\} = \emptyset$ and $b(j) = c(j) = 0$, $a(j) = j$. Thus when showing that $\mathcal{X}_j$ must be of the form (3.1), our concern is to show that there is at most one vertex $z_j \in V_j \setminus V_j^*$, and that there can be such a vertex if and only if there is exactly one vertex $z_j^* \in V_j^* \setminus V_j$, so $|\{z_j, z_j^*\}|$ is 0 or 2.

Now, for $j \geq \mu+1$, the set $\mathcal{X}_j = X_j = X_j^*$, so $\mathcal{X}_\mu$ is of the form (3.1). Also, we showed in the previous section that if $X_k = X_k^*$ for $k \leq \mu$ then $X_j = X_j^*$ for all $j < k$. Thus it is enough to show that if $\mathcal{X}_j$ $(j \leq \mu)$ is of the form (3.1) with $\{z_j, z_j^*\} \neq \emptyset$, then $\mathcal{X}_{j-1}$ is also of the form (3.1).

This will be shown in the process of examining what happens to a set $\mathcal{X}_j$ of the form (3.1) (with $\{z_j, z_j^*\} \neq \emptyset$) at step $j-1$ of the decoding algorithm, an examination which will take most of this section. In this examination, we present notation and develop results upon which our later probabilistic analysis will depend. We begin by considering the parameters a, b, c.

Of course,

$$a = a(j), \quad b = b(j), \quad c = c(j),$$

depend on j, (and on p_μ^* and p_i, $i \geq j$), but we will use the letters a, b, c when j is clear. We let

$$A_j := \{x_1 < \cdots < x_a\}, \quad B_j := \{x_{a+1} < \cdots < x_{a+b}\},$$

and

$$C_j := \{x_{a+b+1} < \cdots < x_{a+b+c}\},$$

so $\mathcal{X}_j = A_j \cup B_j \cup C_j \cup \{z_j,\ z_j^*\}$.

Ultimately, we are interested not just in the set $\mathcal{X}_j$, but in the distance between two trees, i.e. Δ. We will find it useful to examine how this distance changes with each step of the decoding algorithm, so we define

$$\Delta_j = \Delta_j^{(n)}\left(T_j, T_j^*, T_{j+1}, T_{j+1}^*\right) := 1 - |E_j \cap E_j^*| + |E_{j+1} \cap E_{j+1}^*|, \quad (0 \le j \le n-2),$$

and observe that

$$\begin{aligned} \Delta^{(n)} &= n - 1 - |E_0 \cap E_0^*| + |E_{n-1} \cap E_{n-1}^*| \\ &= \Delta_0 + \cdots + \Delta_{n-2} \end{aligned} \tag{3.3}$$

(recall that T_{n-1} is the single vertex n and $T = T_0$). We add exactly one edge to each tree at each step of the algorithm, so the function Δ_j has range $\{-1, 0, 1\}$. Of course $\Delta_j = 0$ for $j > \mu$, and it is easy to check that $\Delta_\mu = 1$ as long as $\min\{p_\mu,\, p_\mu^*\} \notin V_{\mu+1} \cup \{\max X_{\mu+1}\}$ (so on $\mathcal{E}^c$). Further, if $X_j = X_j^*$ and $j < \mu$, then we will add the same edge at every step $i < j$, so $\Delta_i = 0$ for all $i < j$.

Finally, we will need some notation to keep track of what neighbor a given vertex had when it was first added to the tree. Thus for $v \in \{1, \ldots, n-1\}$ we denote by $h(v)$ the neighbor of v in T_j, where j is the highest number such that v is a vertex of T_j. Formally,

$$\text{for } v = y_j, \quad h(v) = h_P(v) := p_{j+1}, \quad (P = (p_1, \ldots, p_{n-2})). \tag{3.4}$$

For example, if our string is $(4, 3, 2, 2, 7)$, then

$$h(1) = 4,\ h(2) = 7,\ h(3) = 2,\ h(4) = 3,\ h(5) = 2,\ h(6) = 7.$$

Now we are prepared to examine the behavior of the parameters a, b, c, and to make some crucial observations about the behavior of Δ_j. In the process we will show that if $\mathcal{X}_j$ is of the form (3.1) with $\{z_j, z_j^*\} \neq \emptyset$ then $\mathcal{X}_{j-1}$ is of the same form (but possibly with $\{z_{j-1}, z_{j-1}^*\} = \emptyset$, meaning $\mathcal{X}_{j-1} = X_{j-1}$). The observations below apply to all $1 \le j \le \mu$.

1. If $p_{j-1} \in A_j \cup B_j \cup C_j$, then $y_{j-1} = y_{j-1}^* = p_{j-1}$, while $z_{j-1} = z_j$, $z_{j-1}^* = z_j^*$, and $\Delta_{j-1} = 0$ because we add the edge $\{p_{j-1}, p_j\}$ to both of T_j, T_j^*, (unless $j = \mu$ in which case $\Delta_{\mu-1} = 1$).

   (a) If $p_{j-1} \in A_j$ then $a(j-1) = a(j) - 1$, while $b(j-1) = b(j)$ and $c(j-1) = c(j)$.

   (b) If $p_{j-1} \in B_j$ then $b(j-1) = b(j) - 1$ while $a(j-1) = a(j)$ and $c(j-1) = c(j)$.

   (c) If $p_{j-1} \in C_j$ then $c(j-1) = c(j) - 1$ while $a(j-1) = a(j)$ and $b(j-1) = b(j)$.

   Thus in every case, one of the parameters a, b, c decreases by 1 while the others remain unchanged.

2. Suppose that $p_{j-1} \in \mathcal{V}_j := V_j \cap V_j^*$. Then

   (a) If $b(j) = c(j) = 0$ then $y_{j-1} = z_j^*$ and $y_{j-1}^* = z_j$, so $X_{j-1} = X_{j-1}^*$. While Δ_{j-1} could assume any of the values $-1, 0, 1$, we have $\Delta_i = 0$ for all $i < j - 1$.

(b) First suppose that $z_j < z_j^*$ and $b(j) > 0, c(j) = 0$. Then $y_{j-1}^* = x_{a+b}$ and $y_{j-1} = z_j^*$, making $z_{j-1}^* = x_{a+b}$, $z_{j-1} = z_j$. We have $B_{j-1} = B_j \setminus \{x_{a+b}\}$, so $a(j-1) = a(j)$, $b(j-1) = b(j) - 1$, $c(j-1) = 0$. Further, $\Delta_{j-1} = 0$ if and only if the event

$$\mathcal{H}_{j-1}^* := \{p_j = h_{P^*}(z_j^*)\} \tag{3.5}$$

occurs, and otherwise $\Delta_{j-1} = 1$.

Similarly, if $z_j > z_j^*$ and $b(j) > 0, c(j) = 0$, then $y_{j-1} = x_{a+b}$ and $y_{j-1}^* = z_j$ with $z_{j-1} = x_{a+b}$, $z_{j-1}^* = z_j^*$. The change in the values of a, b, c are the same as in the case of $z_j < z_j^*$. We also have $\Delta_{j-1} = 0$ if and only if the event

$$\mathcal{H}_{j-1} := \{p_j^* = h_P(z_j)\} \tag{3.6}$$

occurs, and otherwise $\Delta_{j-1} = 1$. In summary, if $b(j) > 0, c(j) = 0$ and $p_{j-1} \in \mathcal{V}_j$, then $\Delta_{j-1} = 1$ unless $\mathcal{H}_{j-1} \cup \mathcal{H}_{j-1}^*$ occurs.

(c) If $b(j) \geq 0, c(j) > 0$ and $p_{j-1} \in \mathcal{V}_j$ then $y_{j-1}^* = y_{j-1} = x_{a+b+c}$, $z_{j-1} = z_j$, $z_{j-1}^* = z_j^*$, and we have $a(j-1) = a(j)$, $b(j-1) = b(j)$, $c(j-1) = c(j) - 1$. Since we add the edge $\{x_{a+b+c}, p_j\}$ to both of T_j, T_j^* we have $\Delta_{j-1} = 0$ (unless $j = \mu$ in which case $\Delta_{\mu-1} = 1$).

3. Suppose that $p_{j-1} = \max\{z_j, z_j^*\}$.

   (a) If $b(j) = c(j) = 0$ then the results are the same as in the case 2a.

   (b) If $b(j) > 0, c(j) = 0$ then the results are the same as in the case 2b.

   (c) Suppose $b(j) \geq 0, c(j) > 0$. If $z_j < z_j^*$ and $p_{j-1} = z_j^*$ then $y_{j-1}^* = x_{a+b+c}$ and $y_{j-1} = z_j^*$, making $z_{j-1}^* = x_{a+b+c}$, $z_{j-1} = z_j$. If $z_j > z_j^*$ and $p_{j-1} = z_j$ then $y_{j-1} = x_{a+b+c}$ and $y_{j-1}^* = z_j$, making $z_{j-1} = x_{a+b+c}$, $z_{j-1}^* = z_j^*$. In both cases, $a(j-1) = a(j)$, but $B_{j-1} = B_j \cup C_j \setminus \{x_{a+b+c}\}$, so $c(j-1) = 0$, $b(j-1) = b(j) + c(j) - 1$. In this case we have $\Delta_{j-1} \geq 0$.

4. The last remaining possibility is that $p_{j-1} = \min\{z_j, z_j^*\}$.

   (a) If $c(j) = 0$ then $y_{j-1} = z_j^*$ and $y_{j-1}^* = z_j$ so $X_{j-1} = X_{j-1}^*$. We have $\Delta_{j-1} \in \{-1, 0, 1\}$ and $\Delta_i = 0$ for all $i < j - 1$.

   (b) If $c(j) > 0$ and $z_j < z_j^*$ then $y_{j-1} = x_{a+b+c}$ and $y_{j-1}^* = z_j$, making $z_{j-1} = x_{a+b+c}$, $z_{j-1}^* = z_j^*$. If $z_j > z_j^*$ then $y_{j-1}^* = x_{a+b+c}$ and $y_{j-1} = z_j^*$, making $z_{j-1}^* = x_{a+b+c}$, $z_{j-1} = z_j$. In both cases $a(j-1) = a(j) + b(j)$ because the set $A_{j-1} = A_j \cup B_j$, and $B_{j-1} = C_j \setminus \{x_{a+b+c}\}$, so $c(j-1) = 0$, $b(j-1) = c(j) - 1$. In this case we have $\Delta_{j-1} \geq 0$.

We have shown that if $\mathcal{X}_j$ is of the form shown in (3.1) then $\mathcal{X}_{j-1}$ will be of the same form. Furthermore, if $\{z_j, z_j^*\} \neq \emptyset$, then $\{z_{j-1}, z_{j-1}^*\} = \emptyset$ (i.e. $X_{j-1} = X_{j-1}^*$) can only occur if $c(j) = 0$, see cases 2a, 3a, and 4a. We have also seen that as j *decreases*: 1) the parameter $c(j)$ never gets larger, and 2) the parameter $b(j)$ decreases by 1 if $p_{j-1} \in B_j$

and otherwise can only decrease if $p_{j-1} \in \{z_j, z_j^*\}$. We end our analysis of the decoding algorithm with one last observation, which is that $\Delta_j = -1$ for at most one value of j, which is clear from an examination of cases 2a, 3a, and 4a, since only in these cases can $\Delta_j = -1$, and in every case $\Delta_i = 0$ for all $i < j$.

In light of the knowledge that $\Delta_j = -1$ at most once, and of (3.3), we now see that (on $\mathcal{E}^c$) if there are $\ell+2$ indices $j_1, \dots j_{\ell+2} \le \mu$ such that $\Delta_i = 1$ (for all $i \in \{j_1, \dots j_{\ell+2}\}$), then $\Delta > \ell$. Thus in order to show that $\Delta(T, T^*) > \ell$ it suffices to show that there are $\ell + 2$ such indices. So we have reduced the 'global' problem of bounding (from below) $\Delta = \Delta_0 + \cdots + \Delta_{n-2}$ to the 'local' problem of showing that it is likely (on $\mathcal{E}^c$) that for at least $\ell + 2$ indices $i \le \mu$ we have $\Delta_i = 1$. We will begin this process in the next section.

4 The upper bound

4.1 Dividing the set $\mathcal{E}^c$

We now begin the process of showing that for any positive integer ℓ,

$$\mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{E}^c \,|\, \mu\right) = O\left(n^{-1/3} \ln^2 n\right). \tag{4.1}$$

The event $\mathcal{E}$ is the event that $p_\mu, p_\mu^* \in V_{\mu+1} \cup \{\max X_{\mu+1}\}$, which means that $\{z_\mu, z_\mu^*\} = \emptyset$ (equivalently $\mathcal{X}_\mu = X_\mu$). So on $\mathcal{E}^c$ we have $|\{z_\mu, z_\mu^*\}| = 2$, and $\mathcal{E}^c$ is the union of the following events:

1. $\mathcal{E}_1 := \{b(\mu) < \delta_n\} \cap \{|\{z_\mu, z_\mu^*\}| = 2\}, \qquad \delta_n := n^{1/3}$,
2. $\mathcal{E}_2 := \{b(\mu) \ge \delta_n\}$.

This means that

$$\mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{E}^c \,|\, \mu\right) \le \mathrm{P}\left(\mathcal{E}_1 \,|\, \mu\right) + \mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{E}_2 \,|\, \mu\right).$$

Let us show now that

$$\mathrm{P}\left(\mathcal{E}_1 \,|\, \mu\right) = O(\delta_n/n). \tag{4.2}$$

Proof. From the definitions of $\mathcal{X}, \mathcal{V}, b(j)$, see (3.1), it is clear that on $\mathcal{E}_1$ either:

1. $\max\{p_\mu, p_\mu^*\} \in \mathcal{V}_{\mu+1}$ and $\min\{p_\mu, p_\mu^*\}$ is one of the $\lceil \delta_n \rceil$ largest elements of $\mathcal{X}_{\mu+1}$, or
2. $p_\mu \in \mathcal{X}_{\mu+1}$ and p_μ^* is separated from p_μ by at most $\lceil \delta_n \rceil$ elements of $\mathcal{X}_{\mu+1}$.

So $\mathcal{E}_1$ is contained in the union of the two events $\mathcal{U}_1, \mathcal{U}_2$ defined as follows:

$$\begin{aligned}
&\mathcal{U}_1 := \{\text{at least one of } p_\mu, p_\mu^* \text{ is one of the } \lceil \delta_n \rceil \text{ largest elements of } \mathcal{X}_{\mu+1}\} \\
&\mathcal{U}_2 := \left\{p_\mu = x_j \in X_{\mu+1}\,;\, p_\mu^* \in \mathcal{Y}(x_j)\right\}, \\
&\mathcal{Y}(x_j) = \mathcal{Y}(p_\mu, \dots, p_{n-2}) := \left\{x_{\min\{1, j - \lceil \delta_n \rceil\}}, \dots, x_{\max\{\mu+1, j + \lceil \delta_n \rceil\}}\right\} \setminus \{x_j\} \subseteq X_{\mu+1}^*
\end{aligned}$$

(note that $|\mathcal{Y}(x_j)| \leq 2\lceil \delta_n \rceil$). Because p_μ is chosen uniformly at random from $[n]$ and p^*_μ is chosen uniformly at random from $[n] \setminus \{p_\mu\}$, a union bound gives us

$$\mathrm{P}\left(\mathcal{U}_1 \mid \mu\right) \leq \frac{\lceil \delta_n \rceil}{n} + \frac{\lceil \delta_n \rceil}{n-1} = O(\delta_n/n).$$

As for $\mathcal{U}_2$, we have

$$\begin{aligned}
\mathrm{P}\left(\mathcal{U}_2 \mid \mu\right) &= \sum_{j=1}^{\mu+1} \mathrm{P}\left(p^*_\mu \in \mathcal{Y}(x_j) \mid p_\mu = x_j \,;\, \mu\right) \mathrm{P}\left(p_\mu = x_j \in X_{\mu+1} \mid \mu\right) \\
&\leq \sum_{j=1}^{\mu+1} \frac{2\lceil \delta_n \rceil}{n-1} \frac{1}{n} = O(\delta_n/n).
\end{aligned}$$

Thus

$$\mathrm{P}\left(\mathcal{E}_1 \mid \mu\right) \leq \mathrm{P}\left(\mathcal{U}_1 \mid \mu\right) + \mathrm{P}\left(\mathcal{U}_2 \mid \mu\right) = O(\delta_n/n).$$

□

So we have proved (4.2), and from now on, we may assume that $b(\mu) = |B_\mu|$ is at least $\lceil \delta_n \rceil$. Further, $B_\mu \subseteq X_j \setminus \{z_j\}$, and $|X_\mu| = \mu$, so we must have $\mu \geq \lceil \delta_n \rceil + 1$ on the event $\mathcal{E}_2$. So from here on, we will also be restricting our attention to $\mu \geq \lceil \delta_n \rceil + 1$. We will end this section with an overview of how we plan to deal with the event $\mathcal{E}_2$.

In order to show that $\mathcal{E}_2$ is negligible, we will start at step $\mu - 1$, with $p^*_\mu, p_\mu, \ldots, p_{n-2}$ already chosen (so that $(P, P^*) \in \mathcal{E}_2$), and we will begin choosing values for a number of positions $p_j = p^*_j$ $(j < \mu)$ of our P-strings. We must eventually reach a step $\tau = \tau(P, P^*)$ at which $c(\tau) = 0$, and we will find that at this step it is unlikely that $b(\tau) << \delta_n$. Then with $b(\tau)$ (on the order of δ_n) values of p_j $(j < \mu)$ left to choose, it is unlikely that fewer than $\ell + 1$ of those choices we will have $p_j \in \mathcal{V}_{j+1}$. From case 2b of section 3, we know that each time $p_j \in \mathcal{V}_{j+1}$ there are three possibilities:

1. the event $\mathcal{H}_j := \{p^*_{j+1} = h_P(z_{j+1})\}$ occurs,
2. the event $\mathcal{H}^*_j := \{p_{j+1} = h_{P^*}(z^*_{j+1})\}$ occurs, or
3. $\Delta_j = 1$

(recall that $h_P(z) = y$ means that y was the neighbor of z when z was added to the tree T corresponding to P). So conditioning on the event that $\Delta_j = 1$ for $\ell + 1$ values of j, we will prove that the event $\mathcal{H}_j \cup \mathcal{H}^*_j$ is unlikely to occur, which makes it likely that we have $\Delta_j = 1$ for $\ell + 1$ values of $j < \mu$. This in turn implies $\Delta > \ell$. Thus we show that $\mathcal{E}_2$ is the union of several unlikely events, and an event on which the conditional probability that $\Delta > \ell$ is high.

In the next section, after introducing some definitions and explaining some technical details, we will elaborate on the plan outlined above. We will end this section by observing that the problem we are trying to solve is conceptually similar to a Pólya urn model with four colors $A, B, C, \mathcal{V}$ (the balls are the vertices in each set) in which the drawing of any ball results in the removal of that ball and its replacement by a ball of color $\mathcal{V}$ (see [3]). The added difficulty we face is that the sizes of our sets change radically if we choose either of two distinguished balls z_j, z_j^* (which may happen with positive probability for μ of order n).

4.2 Definitions and details

We will begin with some definitions we require to carry out the steps outlined at the end of the previous section. Let us start by defining the random variable

$$\tau(z) = \tau(z)(P, P^*) := \max_{j \leq \mu}\{j \,:\, c(j) \leq z\} \qquad (\mu \geq \lceil \delta_n \rceil + 1),$$

and the events

$$\begin{aligned} \mathcal{S} &:= \{b(\tau(0)) \geq \delta_n/5\}, \qquad \delta = \delta_n := n^{1/3}, \\ \mathcal{T} &:= \{\tau(\delta) - \tau(0) \leq 2\beta_n\}, \quad \beta_n := n^{2/3}\ln^2 n. \end{aligned} \tag{4.3}$$

We observe that for $u \leq v$ we have $\tau(u) \leq \tau(v)$ because $c(j)$ is a non-decreasing function of j $(j \leq \mu)$. Further, we note that if $\tau(z) < \mu$, then $|C_{\tau(z)+1}| \geq \lfloor z \rfloor + 1$, and because $C_j \subseteq X_j \setminus \{z_j\}$, we have $|X_{\tau(z)+1}| \geq \lfloor z \rfloor + 2$. Since $|X_j| = j$, it must be true that $\tau(z) \geq \lfloor z \rfloor + 1$, and in particular we have $\tau(\delta) \geq \lfloor \delta \rfloor + 1$, $\tau(0) \geq 1$. These bounds also hold if $\tau(\delta), \tau(0) = \mu$, because we are considering only $\mu \geq \lceil \delta \rceil + 1$. By a similar argument we can see that if $b(\tau(0)) \geq \delta/5$ (as on the event $\mathcal{S}$) then we must have $\tau(0) \geq \lfloor \delta/5 \rfloor + 1$.

Next we note that the following set containment holds for any sets $\mathcal{S}, \mathcal{T}$:

$$\{\Delta = \ell\} \cap \mathcal{E}_2 \subseteq \mathcal{T}^c \cup (\mathcal{S}^c \cap \mathcal{T} \cap \mathcal{E}_2) \cup (\{\Delta = \ell\} \cap \mathcal{S}). \tag{4.4}$$

This containment, along with a union bound, means that

$$\mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{E}_2 \,|\, \mu\right) \leq \mathrm{P}\left(\mathcal{T}^c \,|\, \mu\right) + \mathrm{P}\left(\mathcal{S}^c \cap \mathcal{T} \cap \mathcal{E}_2 \,|\, \mu\right) + \mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{S} \,|\, \mu\right), \tag{4.5}$$

and in the next two sections we will bound each of the terms on the right side of the previous line.

Our discussion at the end of the last section explains our interest in the event $\{\Delta = \ell\} \cap \mathcal{S}$, which depends on $\tau(0)$. But why must we concern ourselves with $\tau(\delta)$ and $\mathcal{T}$? The reason is the complications caused by the possibility of choosing $p_j \in \{z_{j+1}, z_{j+1}^*\}$. To explain fully, we must introduce the events

$$\begin{aligned} &\mathcal{Z}_i := \{p_j \notin \{z_{j+1}, z_{j+1}^*\} \text{ for } i \leq j < \mu\}, \quad (1 \leq i < \mu), \\ &\mathcal{Z}_\delta := \{p_j \notin \{z_{j+1}, z_{j+1}^*\} \text{ for } \tau(\delta) \leq j < \mu\}, \quad \mathcal{Z}_0 := \{p_j \notin \{z_{j+1}, z_{j+1}^*\} \text{ for } \tau(0) \leq j < \mu\}. \end{aligned} \tag{4.6}$$

For a fixed integer $i \geq 1$, we know if the event $\mathcal{Z}_i$ occurred after examining $p_i, \ldots, p_{n-2}, p_\mu^*$, while the events $\mathcal{Z}_\delta$, $\mathcal{Z}_0$ require knowledge of all $p_1, \ldots, p_{n-2}, p_\mu^*$. Of course if we condition on $\tau(0)$ or $\tau(\delta)$ then these last two events require knowledge of only $p_\tau, \ldots, p_{n-2}, p_\mu^*$, for

$\tau = \tau(0)$, $\tau(\delta)$. Also, if $\tau(\delta) = \mu$ (respectively if $\tau(0) = \mu$) then the event $\mathcal{Z}_\delta$ (respectively $\mathcal{Z}_0$) trivially occurred.

To conclude our remarks on the events $\mathcal{Z}_\delta$, $\mathcal{Z}_0$, we note that an examination of their definitions shows that on $\mathcal{Z}_\delta$ (respectively on $\mathcal{Z}_0$) we cannot have chosen $p_j \in \{z_{j+1}, z^*_{j+1}\}$ for $\tau(\delta) < j < \mu$ (resp. for $\tau(0) < j < \mu$), which in turn implies that the parameter $c(j) \geq c(j+1) - 1$ and $b(j) \geq b(j+1) - 1$ for $j > \tau(\delta)$ (resp. $\tau(0)$). On the other hand, on the set $\mathcal{Z}^c_\delta$ we have $\tau(\delta) = \tau(0)$.

To see why we must consider $\tau(\delta)$, note that on the event

$$\{p_j = \min\{z_{j+1}, z^*_{j+1}\} \text{ for } \tau(0) < j < \tau(\delta)\} \subseteq \mathcal{Z}^c_0,$$

we could have

$$c(j+1) << \delta_n \Longrightarrow b(j) = c(j+1) - 1 << \delta_n, \quad c(j) = 0,$$

see case 4b of section 3. This is a problem because we want $b(\tau(0))$ to be at least on the order of δ_n. But if the event $\mathcal{Z}^c_\delta$ occurs, then for some $j \geq \tau(\delta)$ either:
$p_j = \min\{z_{j+1}, z^*_{j+1}\}$ and

$$c(j+1) \geq \delta_n \Longrightarrow b(j) = c(j+1) - 1 \geq \delta_n - 1, \quad c(j) = 0,$$

(see case 4b), or $p_j = \max\{z_{j+1}, z^*_{j+1}\}$ and

$$c(j+1) \geq \delta_n \Longrightarrow b(j) = b(j+1) + c(j+1) - 1 \geq \delta_n - 1, \quad c(j) = 0,$$

(see case 3c).

Thus (for n large enough that $\delta_n - 1 \geq \delta_n/5$) we have

$$\mathcal{Z}^c_\delta \subseteq \mathcal{S} \Longrightarrow \mathcal{S}^c \cap \mathcal{T} \cap \mathcal{E}_2 \subseteq (\mathcal{S}^c \cap \mathcal{Z}_0 \cap \mathcal{E}_2) \cup (\mathcal{Z}^c_0 \cap \mathcal{Z}_\delta \cap \mathcal{T})$$

(the set containment in the previous line is true for any sets $\mathcal{S}, \mathcal{T}, \mathcal{E}_2, \mathcal{Z}_0, \mathcal{Z}_\delta$). This means that

$$\mathrm{P}\left(\mathcal{S}^c \cap \mathcal{T} \cap \mathcal{E}_2 \,|\, \mu\right) \leq \mathrm{P}\left(\mathcal{S}^c \cap \mathcal{Z}_0 \cap \mathcal{E}_2 \,|\, \mu\right) + \mathrm{P}\left(\mathcal{Z}^c_0 \cap \mathcal{Z}_\delta \cap \mathcal{T} \,|\, \mu\right). \tag{4.7}$$

In the next section, we start by showing that

$$\mathrm{P}\left(\mathcal{T}^c \,|\, \mu\right) + \mathrm{P}\left(\mathcal{Z}^c_0 \cap \mathcal{Z}_\delta \cap \mathcal{T} \,|\, \mu\right) = O(\beta_n/n), \tag{4.8}$$

and then we will prove that

$$\mathrm{P}\left(\mathcal{S}^c \cap \mathcal{Z}_0 \cap \mathcal{E}_2 \,|\, \mu\right) = O(\beta_n/n). \tag{4.9}$$

Finally, in section 4.4 we will prove that

$$\mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{S} \,|\, \mu\right) = O(\delta_n/n). \tag{4.10}$$

The combination of (4.7)-(4.10), along with (4.5), implies

$$\mathrm{P}\left(\{\Delta=\ell\}\cap\mathcal{E}_2\,|\,\mu\right)=O(\beta_n/n)=O\left(n^{-1/3}\ln^2 n\right).$$

We end this section with one final detail: we comment on a method of proof we will use in the next two sections. We will occasionally show (for some events $\mathcal{A},\mathcal{B}$ that depend on n) that $\mathrm{P}\left(\mathcal{B}\,|\,\mu\right)\to 0$ by first showing that for some event $\mathcal{A}$ we have $\mathrm{P}\left(\mathcal{A}^c\,|\,\mu\right)\to 0$, and then showing that

$$\mathrm{P}\left(\mathcal{B}\,|\,\mathcal{A}\,;\,\mu\right):=\frac{\mathrm{P}\left(\mathcal{B}\cap\mathcal{A}\,|\,\mu\right)}{\mathrm{P}\left(\mathcal{A}\,|\,\mu\right)}\to 0,\quad n\to\infty.$$

Obviously the combination of these results proves that $\mathrm{P}\left(\mathcal{B}\,|\,\mu\right)\to 0$ as $n\to\infty$. A conditional probability like the one above is only defined as long as $\mathrm{P}\left(\mathcal{A}\,|\,\mu\right)>0$, but of course if $\mathrm{P}\left(\mathcal{A}\,|\,\mu\right)=0$ then because $\mathcal{B}\subset\mathcal{A}\cap\mathcal{A}^c$ we must have $\mathrm{P}\left(\mathcal{B}\,|\,\mu\right)\to 0$ anyway. Thus whenever we discuss conditional probabilities we will assume (and not prove) that the event we condition on has positive probability.

4.3 Bounding some unlikely events

Let us begin by proving the result (4.8).

Lemma 4.1 *Let $\mathcal{T}=\{\tau(\delta)-\tau(0)\le 2\beta_n\}$, and let $\mathcal{Z}_0,\mathcal{Z}_\delta$ be defined as in* (4.6)*. Then*

$$P\left(\mathcal{T}^c\,|\,\mu\right)=O\left(n^{-1}\right),\qquad\qquad P\left(\mathcal{Z}_0^c\cap\mathcal{Z}_\delta\cap\mathcal{T}\,|\,\mu\right)=O(\beta_n/n).$$

Proof. We will start with the second of the results above. We will condition on the value of $\tau(\delta)$, and introduce notation for events conditioned on that value:

$$\mathrm{P}\left(\mathcal{W}\,|\,\tau\,;\,\mu\right):=\mathrm{P}\left(\mathcal{W}\,|\tau=\tau(\delta)\,;\,\mu\right).$$

With $\mathcal{Z}_i$ defined as in (4.6), we observe that $\mathcal{Z}_i\subseteq\mathcal{Z}_{i+1}$. If the set$\{z_{i+1},z^*_{i+1}\}$ is empty, then the (conditional) probability that $p_i\in\{z_{i+1},z^*_{i+1}\}$ is 0, and if the set $\{z_{i+1},z^*_{i+1}\}$ is non-empty, then the (conditional) probability that $p_i\in\{z_{i+1},z^*_{i+1}\}$ is $2/n$. Thus we have

$$\mathrm{P}\left(\mathcal{Z}_i^c\cap\mathcal{Z}_{i+1}|\,\tau\,;\,\mu\right)\le 2/n,\quad(1\le i<\mu-1).\tag{4.11}$$

To avoid having to condition also on the value of $\tau(0)$, we introduce $\mathcal{Z}_\phi$, where $\phi=\max\{\tau(\delta)-2\lfloor\beta_n\rfloor,1\}$. On $\mathcal{T}$, the event $\mathcal{Z}_\phi$ implies the event $\mathcal{Z}_0$, so $\mathcal{Z}_\phi\cap\mathcal{T}\subseteq\mathcal{Z}_0\cap\mathcal{T}$. Also, consideration of the definitions of $\mathcal{Z}_i$ and $\mathcal{T}$ shows that $\mathcal{Z}_\phi^c\cap\mathcal{Z}_\delta\subseteq\mathcal{T}$.

From law of total probability we have

$$\mathrm{P}\left(\mathcal{Z}_\phi^c\cap\mathcal{Z}_\delta\,|\,\mu\right)=\sum_{\tau=\lfloor\delta\rfloor+1}^{\mu}\mathrm{P}\left(\mathcal{Z}_\phi^c\cap\mathcal{Z}_\delta\,|\,\tau\,;\,\mu\right)\mathrm{P}\left(\tau=\tau(\delta)\,|\,\mu\right)\tag{4.12}$$

(the discussion after (4.3) explains why $\tau \geq \lfloor \delta \rfloor + 1$). Since $\tau - \phi \leq 2\beta_n$, we obtain from (4.11) the bound

$$\begin{aligned} \mathrm{P}\left(\mathcal{Z}_\phi^c \cap \mathcal{Z}_\delta \,|\, \tau\,;\, \mu\right) &= \sum_{i=\phi}^{\tau-1} \mathrm{P}\left(\mathcal{Z}_i^c \cap \mathcal{Z}_{i+1} \,|\, \tau\,;\, \mu\right) \\ &\leq 2\beta_n(2/n) = 4\beta_n/n. \end{aligned} \tag{4.13}$$

This bound is independent of τ, so (4.13), combined with (4.12) shows that

$$\mathrm{P}\left(\mathcal{Z}_\phi^c \cap \mathcal{Z}_\delta \,|\, \mu\right) = O(\beta_n/n). \tag{4.14}$$

We noted (above (4.12)) that $\mathcal{Z}_\phi \cap \mathcal{T} \subseteq \mathcal{Z}_0 \cap \mathcal{T}$. For any sets $\mathcal{Z}_0, \mathcal{Z}_\phi, \mathcal{Z}_\delta, \mathcal{T}$, we have

$$\mathcal{Z}_\phi \cap \mathcal{T} \subseteq \mathcal{Z}_0 \cap \mathcal{T} \Longrightarrow \mathcal{Z}_0^c \cap \mathcal{T} \subseteq \mathcal{Z}_\phi^c \cap \mathcal{T},$$

with the right-most set containment implying that

$$\mathcal{Z}_0^c \cap \mathcal{Z}_\delta \cap \mathcal{T} \subseteq \mathcal{Z}_\phi^c \cap \mathcal{Z}_\delta \cap \mathcal{T} = \mathcal{Z}_\delta \cap \mathcal{Z}_\phi^c$$

(the last equality follows from the observation that $\mathcal{Z}_\delta \cap \mathcal{Z}_\phi^c \subseteq \mathcal{T}$, as was noted above (4.12)). The discussion above, combined with (4.14) shows that

$$\mathrm{P}\left(\mathcal{Z}_0^c \cap \mathcal{Z}_\delta \cap \mathcal{T} \,|\, \mu\right) = O(\beta_n/n).$$

Further, on the event $\mathcal{Z}_\delta^c$ we have $\tau(0) = \tau(\delta)$ (see the second paragraph after (4.6)), so $\mathcal{Z}_\delta^c \subseteq \mathcal{T}$ and

$$\mathcal{Z}_\delta^c \subseteq \mathcal{T},\ \mathcal{Z}_\delta \cap \mathcal{Z}_\phi^c \subseteq \mathcal{T} \Longrightarrow \mathcal{T}^c = \mathcal{T}^c \cap \mathcal{Z}_\phi,$$

(the implication above is true for any sets $\mathcal{Z}_\delta, \mathcal{Z}_\phi, \mathcal{T}$). Thus we conclude that

$$\mathrm{P}\left(\mathcal{T}^c \,|\, \mu\right) = \mathrm{P}\left(\mathcal{T}^c \cap \mathcal{Z}_\phi \,|\, \mu\right),$$

and we will proceed to bound the probability above. Now $\{\tau(\delta) \leq 2\lfloor \beta_n \rfloor\} \subseteq \mathcal{T}$, so we may restrict our attention to $\tau(\delta) > 2\lfloor \beta_n \rfloor$. Hence

$$\mathrm{P}\left(\mathcal{T}^c \cap \mathcal{Z}_\phi \,|\, \mu\right) = \sum_{\tau=2\lfloor \beta_n \rfloor+1}^{\mu} \mathrm{P}\left(\mathcal{T}^c \cap \mathcal{Z}_\phi \,|\, \tau\,;\, \mu\right) \mathrm{P}\left(\tau(\delta) = \tau \,|\, \mu\right).$$

To complete the proof of this lemma it is sufficient to show that

$$\mathrm{P}\left(\mathcal{T}^c \cap \mathcal{Z}_\phi \,|\, \tau\,;\, \mu\right) = O\left(n^{-1}\right).$$

Toward this end, we define

$$\epsilon_n := \frac{1}{\ln n}, \quad \nu = \nu_n := \lfloor \epsilon_n \beta_n / \delta_n \rfloor, \quad k = k_n := \lfloor \delta_n / \epsilon_n \rfloor,$$

observing that

$$k\nu \le \lfloor \beta_n \rfloor, \qquad k >> \delta_n, \qquad k\nu^2 >> n \ln n.$$

Then we consider the sub-string $(p_{\tau-2\nu k}, \ldots, p_{\tau-1})$, which can be divided into $2k$ segments of length ν, leading us to introduce the notation

$$P(i) := (p_{m(i)}, \ldots, p_{m(i-1)-1}), \quad m(i) := \tau - i\nu, \quad (1 \le i \le 2k),$$

and

$$\mathcal{D}_i := \{p_j \in \mathcal{V}_{j+1} \text{ for at least one } p_j \in P(i)\}.$$

At step $\tau(\delta)$, we have $c(\tau(\delta)) \le \delta_n$. Each choice of $p_j \in \mathcal{V}_{j+1}$ forces us to add an element of $C_{\tau(\delta)}$ as a vertex of the pair of trees we are constructing (see section 3 case 2c). If we choose $p_j \in \mathcal{V}_{j+1}$ at least δ_n times in steps $\tau - 1$ through $\tau - 2\nu k$, then $\tau(0) \ge \tau - 2\nu k$ (which means $\mathcal{T}$ occurred). Consequently, the event $\mathcal{T}^c \cap \mathcal{Z}_\phi$ is the event that in steps $\tau - 1$ through $\tau - 2\nu k$ we add fewer than δ_n elements of $C_{\tau(\delta)}$ as vertices of the pair of trees we are building. Because $k >> \delta_n$, this means

$$\mathcal{T}^c \cap \mathcal{Z}_\phi \subseteq \bigcup_{i=k+1}^{2k} \mathcal{D}_i^c. \tag{4.15}$$

This will prove useful, because we can easily bound $\mathrm{P}\left(\mathcal{D}_i^c \,|\, \mathcal{Z}_\phi \,;\, \tau \,;\, \mu\right)$ (and we will now do so).

On the event $\mathcal{Z}_\phi$, we have $|\mathcal{V}_{j+1}| = n - (j+1) - 1$ for $\tau - 2k\nu \le j \le \tau - 1$, see (3.2). Thus

$$\mathrm{P}\left(p_j \notin \mathcal{V}_{j+1} \,|\, \mathcal{Z}_\phi \,;\, \tau \,;\, \mu\right) = 1 - \frac{n-j-2}{n-2},$$

and the events $p_j \notin \mathcal{V}_{j+1}$ are conditionally independent for $\tau - 2k\nu \le j \le \tau - 1$. Also for $m(i) \le j \le m(i-1) - 1$ we have

$$\begin{aligned} |\mathcal{V}_{j+1}| = n - j - 2 &\ge n - \big(\tau - (i-1)\nu - 1\big) - 2 \\ &\ge n - \big(n - 2 - (i-1)\nu - 1\big) - 2 \\ &\ge (i-1)\nu. \end{aligned}$$

Using this bound, along with the inequality $1 - x \le e^{-x}$, we find that

$$\begin{aligned} \mathrm{P}\left(\mathcal{D}_i^c \,|\, \mathcal{Z}_\phi \,;\, \tau \,;\, \mu\right) &= \prod_{j=m(i)}^{m(i-1)-1} \mathrm{P}\left(p_j \notin \mathcal{V}_{j+1} \,|\, \mathcal{Z}_\phi \,;\, \tau \,;\, \mu\right) \\ &= \prod_{j=m(i)}^{m(i-1)-1} \left(1 - \frac{n-j-2}{n-2}\right) \\ &\le \left(1 - \frac{(i-1)\nu}{n-2}\right)^{\nu} \le e^{-(i-1)\nu^2/(n-2)}. \end{aligned} \tag{4.16}$$

Hence, combining (4.16)-(4.15), we have

$$\begin{aligned}\mathrm{P}\left(\mathcal{T}^c \,|\, \mathcal{Z}_\phi \,;\, \tau \,;\, \mu\right) &\leq \sum_{i=k+1}^{2k} \mathrm{P}\left(\mathcal{D}_i^c \,|\, \mathcal{Z}_\phi \,;\, \tau \,;\, \mu\right) \\ &\leq ke^{-k\nu^2/(n-2)} = O\left(n^{-1}\right),\end{aligned}$$

and we find that

$$\mathrm{P}\left(\mathcal{T}^c \cap \mathcal{Z}_\phi \,|\, \tau \,;\, \mu\right) \leq \mathrm{P}\left(\mathcal{T}^c \,|\, \mathcal{Z}_\phi \,;\, \tau \,;\, \mu\right) = O\left(n^{-1}\right).$$

□

Next we prove (4.9).

Lemma 4.2 *Let $\mathcal{S} = \{b(\tau(0)) \geq \delta_n/5\}$, and let $\mathcal{Z}_0$ be defined as in* (4.6). *Then*

$$P\left(\mathcal{S}^c \cap \mathcal{Z}_0 \cap \mathcal{E}_2 \,|\, \mu\right) = O\left(1/\lfloor \delta_n \rfloor\right) = O\left(\beta_n/n\right).$$

Proof. We will condition of the composition of the set B_μ, $(b(\mu) \geq \lceil \delta_n \rceil)$ and prove that

$$\mathrm{P}\left(\mathcal{S}^c \cap \mathcal{Z}_0 \cap \mathcal{E}_2 \,|\, B_\mu \,;\, \mu\right) = O\left(1/\lfloor \delta_n \rfloor\right).$$

With the law of total probability, this is enough to prove the lemma.

Given a fixed set B_μ, we enumerate its lowest $d := \lfloor \delta_n \rfloor$ elements from least to greatest as $S = \{q_1, \ldots, q_d\}$. Denoting by $\mathcal{Q}$ the event that more than $4/5$ of these q_i are chosen as values for p_j (for $j < \mu$), we have (conditioned on B_μ) $\mathcal{S}^c \cap \mathcal{Z}_0 \cap \mathcal{E}_2 \subseteq \mathcal{Q}$. Thus it is enough to show that the conditional probability of $\mathcal{Q}$ occurring is on the order of $1/d$. To do so, we will treat $\mu - 1 > n/5$ and $\mu - 1 \leq n/5$ as separate cases. Let us begin with the first of these cases.

We denote by $\mathcal{Q}(i)$ the event that element q_i is chosen as some p_j for $j < \mu$. Then we count the number of times this happens with the random variable

$$Q_n = \sum_{i=1}^{d} \mathbf{I}_{\mathcal{Q}(i)}$$

where $\mathbf{I}_{\mathcal{A}}$ is the indicator of the event $\mathcal{A}$. We note that on $\mathcal{Q}$ we must have $Q_n \geq 4d/5$. So our goal is to show that

$$\mathrm{P}\left(Q_n \geq 4d/5 \,|\, B_\mu \,;\, \mu\right) = O(1/d). \tag{4.17}$$

Observing that

$$\mathrm{P}\left(\mathcal{Q}(i) \,|\, B_\mu \,;\, \mu\right) = 1 - \left(1 - \frac{1}{n}\right)^{\mu-1} \geq 1 - e^{-1/5} \geq 1/10, \tag{4.18}$$

$$\mathrm{P}\left(\mathcal{Q}(i) \cap \mathcal{Q}(j) \,|\, B_\mu \,;\, \mu\right) = 1 - 2\left(1 - \frac{1}{n}\right)^{\mu-1} + \left(1 - \frac{2}{n}\right)^{\mu-1} \qquad \text{(for } i \neq j\text{)},$$

(so for fixed μ, n, the random variable Q_n is a sum of Bernoulli random variables) we expect that we can use Chebyshev's inequality to bound the probability that Q_n deviates from its expectation by more than a fraction of that expectation. The first and second (conditional) factorial moments of Q_n are easy to find, because

$$\begin{aligned} E_{1,n} &:= \mathrm{E}\left[Q_n \mid B_\mu\,;\,\mu\right] = d\mathrm{P}\left(\mathcal{Q}(i) \mid B_\mu\,;\,\mu\right) \geq d/10, \\ E_{2,n} &:= \mathrm{E}\left[Q_n(Q_n-1) \mid B_\mu\,;\,\mu\right] = d(d-1)\mathrm{P}\left(\mathcal{Q}(i)\cap\mathcal{Q}(j) \mid B_\mu\,;\,\mu\right). \end{aligned} \tag{4.19}$$

Now for large enough n,

$$E_{1,n} \leq d\big(1 - e^{-1}(1+o(1))\big) < 7d/10.$$

Consequently, we obtain

$$\mathrm{P}\left(Q_n \geq 4d/5 \mid B_\mu\,;\,\mu\right) \leq \mathrm{P}\left(Q_n - E_{1,n} > E_{1,n}/7 \mid B_\mu\,;\,\mu\right). \tag{4.20}$$

Using (4.18)-(4.19), we can bound the variance of Q_n,

$$\begin{aligned} \mathrm{VAR}\left[Q_n \mid B_\mu\,;\,\mu\right] &= E_{2,n} + E_{1,n} - E_{1,n}^2 \\ &= d^2\left(\left(1-\frac{2}{n}\right)^\mu - \left(1-\frac{1}{n}\right)^{2\mu}\right) + d\left(\left(1-\frac{1}{n}\right)^\mu - \left(1-\frac{2}{n}\right)^\mu\right) \\ &= d^2\left(1-\frac{2}{n}\right)^\mu\left(1-\left(1-\frac{1}{n(n-2)}\right)^\mu\right) + O(d) \\ &= O\left(\frac{d^2\mu}{n^2}\right) + O(d) = O(d). \end{aligned} \tag{4.21}$$

Combining (4.19) and (4.21), and using Chebyshev's inequality, we find that

$$\mathrm{P}\left(|Q_n - E_{1,n}| > E_{1,n}/7 \mid B_\mu\,;\,\mu\right) \leq \frac{49\mathrm{VAR}[Q_n \mid B_\mu\,;\,\mu]}{E_{1,n}^2} = O(1/d). \tag{4.22}$$

This proves (4.20) and thus (4.17), completing the case of $\mu - 1 > n/5$.

In order to bound the probability of $\mathcal{Q}$ when $\mu - 1 \leq n/5$, we will count the number strings of length $\mu-1$ which use at least $4/5$ of the elements of S, and denote this number by $N(S)$. Then we will divide $N(S)$ by the total number of strings of length $\mu - 1$. So the probability of event $\mathcal{Q}$ (conditioned on μ and the composition of B_μ, with $b(\mu) \geq \delta_n$) is $N(S)/n^{\mu-1}$. Before we begin counting, let us also introduce the notation

$$(z)_j := z(z-1)\cdots(z-j+1), \quad k := \lceil 4d/5 \rceil.$$

To calculate (and then bound) $N(S)$, we

1. choose k out of $\mu - 1$ positions,
2. choose k distinct elements of S for those positions, and
3. then we choose any value of p_j for the remaining $\mu - 1 - k$ positions.

Thus we have

$$\frac{N(S)}{n^k} \leq \binom{\mu-1}{k} \frac{(d)_k\, n^{\mu-1-k}}{n^{\mu-1}}.$$

Using the bounds

$$\frac{k^k}{e^k} \leq k! \qquad\qquad (d)_k \leq d^k \quad ,$$

we obtain

$$\begin{aligned}\binom{\mu-1}{k} \frac{(d)_k\, n^{\mu-1-k}}{n^{\mu-1}} &= O\left(\frac{\mu^k d^k e^k}{k^k n^k}\right) \\ &= O\left(5^{-k}(5/4)^k e^k\right) = O\left(n^{-1}\right).\end{aligned}$$

□

In this section we have shown that

$$\mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{E}_2 \,|\, \mu\right) = \mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{S} \,|\, \mu\right) + O(\beta_n/n).$$

In the next section we will consider the event $\{\Delta = \ell\} \cap \mathcal{S}$.

4.4 The event $\{\Delta = \ell\} \cap \mathcal{S}$

At the end of section 4.1 we outlined our plan for showing that it is unlikely that $\Delta \leq \ell$ on $\mathcal{E}_2$. We have now reached the point of having proved it unlikely that $b(\tau(0)) < \lfloor \delta_n/5 \rfloor$, and it remains for us to show that:

1. it is unlikely (on $\mathcal{S}$) that $\Delta_j = 1$ fewer than $\ell + 2$ times,
2. for $\ell + 2$ values of j such that $\Delta_j = 1$, it is unlikely that $\mathcal{H}_j \cup \mathcal{H}_j^*$ occurs.

In this fashion we will show that

$$\mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{S} \,|\, \mu\right) = O\left(\delta_n/n\right). \tag{4.23}$$

We will begin by conditioning on the value of $\tau(0)$ ($\tau(0) = \tau \geq \lfloor \delta_n/5 \rfloor$), and let

$$\nu = \nu_n := \lfloor \tau/k \rfloor, \quad k := \ell + 3.$$

We will then divide the substring $(p_{\tau-k\nu}, \ldots, p_{\tau-1})$, into k segments of length ν. The event $\mathcal{H}_j \cup \mathcal{H}_j^*$ depends on the value of p_{j+1}. Thus to preserve the (conditional) independence of the segments, we need to leave the right-most element of each segment as a buffer between adjacent sub-segments. This leads us to introduce the notation

$$P^-(i) := (p_{m(i)}, \ldots, p_{m(i-1)-2}), \quad m(i) := \tau - i\nu, \quad (1 \leq i \leq k),$$

to denote the last $\nu - 1$ elements of the i th segment.

Let us now introduce a familiar event:

$$\mathcal{Z}_* := \{p_j \notin \{z_{j+1}, z^*_{j+1}\} \text{ for } \tau - k\nu \le j < \tau\}.$$

The event ${\mathcal{Z}_*}^c$ is unlikely to occur (conditioned on $\mathcal{S}$); in fact

$$\mathrm{P}\left({\mathcal{Z}_*}^c \,|\, \mathcal{S}\,;\,\tau\,;\,\mu\right) = O(\delta_n/n). \tag{4.24}$$

The proof is similar to the proof of (4.14) in lemma 4.1, so we omit it.

We will show that conditioned on $\mathcal{S} \cap \mathcal{Z}_*$, it is likely that the event

$$\mathcal{C} := \{\text{we choose at least one } p_j \in \mathcal{V}_{j+1} \text{ in each segment } P^-(i) \text{ for } 2 \le i \le k\}$$

occurs. Then, conditioned on $\mathcal{C} \cap \mathcal{S} \cap \mathcal{Z}_*$, we will show that it is unlikely that $\Delta \le \ell$. Notice that in the definition of $\mathcal{C}$, we are ignoring the rightmost segment $P^-(0) := (p_{\tau-\nu}, \ldots, p_{\tau-2})$. The reason is that we want to make sure that the set $\mathcal{V}_{j+1}$ is large enough that we can find a good lower bound for $\mathrm{P}\left(p_j \in \mathcal{V}_{j+1} \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\,\tau\,;\,\mu\right)$, which we will now proceed to do.

Lemma 4.3 *For $j \le \tau - \nu$,*

$$P\left(p_j \in \mathcal{V}_{j+1} \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\,\tau\,;\,\mu\right) \ge \frac{1}{k} + O\left(n^{-1}\right)$$

Proof. Because $|\mathcal{V}_{j+1}| = n - j - 2$ $(1 \le j \le \tau - 1)$, we can obtain the bound

$$\mathrm{P}\left(p_j \in \mathcal{V}_{j+1} \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\,\tau\,;\,\mu\right) = \frac{n-j-2}{n-2}, \tag{4.25}$$

and that the events $p_j \in \mathcal{V}_{j+1}$ are conditionally independent for $\tau - k\nu \le j < \tau$. Now, if $n/2 \le \tau \le n$, then $\nu \ge \tau/k - 1$, so

$$\frac{n-j-2}{n-2} \ge \frac{n-(\tau-\nu)-2}{n-2} \ge \frac{n/k-3}{n-2} \ge \frac{1}{k} + O\left(n^{-1}\right).$$

Meanwhile, if $\tau \le n/2$, then

$$\frac{n-j-2}{n-2} \ge \frac{n-\tau-2}{n-2} \ge \frac{n/2-2}{n-2} \ge \frac{1}{2} + O\left(n^{-1}\right).$$

Thus we obtain

$$\mathrm{P}\left(p_j \in \mathcal{V}_{j+1} \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\,\tau\,;\,\mu\right) \ge \frac{1}{k} + O\left(n^{-1}\right).$$

$\square$

In order to proceed further, we must define

$$\mathcal{C}_i = \{p_j \in \mathcal{V}_{j+1} \text{ for at least one } p_j \in P(i)^-\}, \qquad \mathcal{C} = \cap_{i=2}^k \mathcal{C}_i.$$

Lemma 4.4 *Let $\mathcal{C}$ and $\mathcal{C}_i$ be defined as above. Then*

$$P\left(\mathcal{C}^c \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\, \tau\,;\, \mu\right) = O\left(n^{-1}\right).$$

Proof. First, we denote $\rho(i) := m(i-1) - 2$. Then, repeating the arguments of (4.16) (and using lemma 4.3), we have

$$\begin{aligned} \mathrm{P}\left(\mathcal{C}_i^c \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\, \tau\,;\, \mu\right) &= \prod_{j=m(i)}^{\rho(i)} \mathrm{P}\left(p_j \notin \mathcal{V}_{j+1} \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\, \tau\,;\, \mu\right) \\ &\leq \left(1 - \frac{1}{k} + O\left(n^{-1}\right)\right)^{\nu-1} = O\left(n^{-1}\right), \end{aligned}$$

where the last bound follows from the fact that ν is of order $n^{1/3}$. Since $\mathcal{C}^c = \cup_{i=2}^{k} \mathcal{C}_i^c$, we use a union bound to obtain

$$\mathrm{P}(\mathcal{C}^c \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\, \tau\,;\, \mu) \leq \sum_{i=2}^{k} \mathrm{P}\left(\mathcal{C}_i^c \,|\, \mathcal{Z}_* \cap \mathcal{S}\,;\, \tau\,;\, \mu\right) = O\left(n^{-1}\right).$$

□

In order to complete our proof of (4.23), we introduce the notation

$$\mathcal{G} = \mathcal{C} \cap \mathcal{Z}_* \cap \mathcal{S}, \tag{4.26}$$

and note that (4.24) and lemma 4.4 imply that

$$\begin{aligned} \mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{S} \,|\, \mu\right) &\leq \mathrm{P}\left(\{\Delta = \ell\} \cap \mathcal{G} \,|\, \mu\right) + \mathrm{P}\left(\mathcal{C}^c \cap \mathcal{S} \,|\, \mu\right) + \mathrm{P}\left({\mathcal{Z}_*}^c \cap \mathcal{S} \,|\, \mu\right) \\ &\leq \mathrm{P}\left(\{\Delta = \ell\} \,|\, \mathcal{G}\,;\, \mu\right) + O(\delta_n/n). \end{aligned}$$

On the event $\mathcal{G}$, we will choose at least one $p_j \in \mathcal{V}_{j+1}$ from each segment $P^-(i)$ for the $\ell + 2$ segments $2 \leq i \leq k$. Thus we can consider the (random) subset of indices

$$\Gamma = \{\gamma(k) < \cdots < \gamma(2)\}, \tag{4.27}$$

where $\gamma(i)$ is the largest element of $\{m(i), \ldots, \rho(i)\}$ such that $p_{\gamma(i)} \in \mathcal{V}_{\gamma(i)+1}$. This makes $p_{\gamma(i)}$ the right-most element of the segment such that $p_j \in \mathcal{V}_{j+1}$. Now, unless the event $\mathcal{H}_{\gamma(i)} \cup \mathcal{H}^*_{\gamma(i)}$ occurs, we will have $\Delta_{\gamma(i)} = 1$. Determining whether this event occurs should require conditioning on the set Γ, but we will be able to avoid this by proving that (on $\mathcal{G}$) the event $\gamma(i) < \rho(i)$ implies that $\mathcal{H}_{\gamma(i)} \cup \mathcal{H}^*_{\gamma(i)}$ did not occur.

Lemma 4.5 *Let $\rho(i) := m(i-1)-2$, and let $\gamma(i)$ be defined as above. Then for $2 \leq i \leq k$,*

$$\{\gamma(i) < \rho(i)\} \cap \mathcal{G} \subseteq \mathcal{G} \cap \left(\mathcal{H}_{\gamma(i)} \cup \mathcal{H}^*_{\gamma(i)}\right)^c.$$

Proof. We begin by noting that on $\mathcal{Z}_*$,

$$p_j \in \mathcal{V}_{j+1}^c = A_{j+1} \cup B_{j+1} \cup \{z_{j+1}, z_{j+1}^*\} \Longrightarrow p_j \in A_{j+1} \cup B_{j+1},$$

for $\tau - k\nu \leq j < \tau$. Thus if $\gamma(i) = j < \rho(i)$, then

$$p_{j+1} \in A_{j+2} \cup B_{j+2}.$$

Now, recall that the elements of $A_{j+2} \cup B_{j+2}$ have *not* appeared as any entry p_i or p_i^* $(i \geq j+2)$, but both $h_P(z_{j+1}^*), h_{P^*}(z_{j+1})$ *have* appeared as some p_i or p_i^* $(i \geq j+2)$ because $h_P(x)$ is defined as the neighbor of x when x was added to the tree corresponding to P. Thus

$$h_P(z_{j+1}^*), h_{P^*}(z_{j+1}) \notin A_{j+2} \cup B_{j+2} \Longrightarrow p_{j+1} \neq h_P(z_{j+1}^*), h_{P^*}(z_{j+1}).$$

But this means that (on $\mathcal{G}$) the event $\gamma(i) < \rho(i)$ implies that $\mathcal{H}_{\gamma(i)} \cup \mathcal{H}_{\gamma(i)}^*$ did not occur. □

In order to use the lemma we have just proved, let us define

$$H = H(P, P^*) := \sum_{i=2}^{k} \mathbf{I}_{\mathcal{G} \cap (\mathcal{H}_{\rho(i)} \cup \mathcal{H}_{\rho(i)}^*)}.$$

So (on the event $\mathcal{G}$), the random variable H counts the number of i for which $\mathcal{H}_{\rho(i)} \cup \mathcal{H}_{\rho(i)}^*$ occurs. Lemma 4.5 implies that $\mathcal{H}_{\gamma(i)} \cup \mathcal{H}_{\gamma(i)}^*$ can only occur if both: 1) $\gamma(i) = \rho(i)$, and 2) $\mathcal{H}_{\rho(i)} \cup \mathcal{H}_{\rho(i)}^*$ occur. In terms of indicator variables, this means that for every i $(2 \leq i \leq k)$,

$$\mathbf{I}_{\mathcal{G} \cap (\mathcal{H}_{\gamma(i)} \cup \mathcal{H}_{\gamma(i)}^*)} \leq \mathbf{I}_{\mathcal{G} \cap (\mathcal{H}_{\rho(i)} \cup \mathcal{H}_{\rho(i)}^*)}.$$

Thus H is an upper bound for the number of times that $\mathcal{H}_{\gamma(i)} \cup \mathcal{H}_{\gamma(i)}^*$ occurred (conditioned on the event $\mathcal{G}$). In light of our discussions at the beginning of this section and at the end of section 4.1, this means that

$$\mathrm{P}\left(\{\Delta = \ell\} \,|\, \mathcal{G}\,;\, \mu\right) \leq \mathrm{P}\left(\{H > 0\} \,|\, \mathcal{G}\,;\, \mu\right).$$

Thus it remains to prove the following lemma.

Lemma 4.6

$$P\left(\{H > 0\} \,|\, \mathcal{G}\,;\, \mu\right) = O\left(n^{-1}\right).$$

Proof. Appealing to the law of total probability, we will show that

$$\mathrm{P}\left(\{H > 0\} \,|\, \tau\,;\, \mathcal{G}\,;\, \mu\right) = O\left(n^{-1}\right).$$

Conditioned on the event $\mathcal{G} = \mathcal{C} \cap \mathcal{Z}_* \cap \mathcal{S}$, we have $|\{z_{j+1}, z_{j+1}^*\}| = 2$ (for $\tau - k\nu \leq j < \tau$), so

$$\mathrm{P}\left(\mathcal{H}_{\rho(i)} \,|\, \mathcal{G}\,;\, \tau\,;\, \mu\right) = \begin{cases} 1/(n-2), & h_P(z_{\rho(i)+1}) \notin \{z_{\rho(i)+2}, z_{\rho(i)+2}^*\}, \\ 0, & \text{otherwise.} \end{cases}$$

This means that

$$\mathrm{P}\left(\mathcal{H}_{\rho(i)}\,|\,\mathcal{G}\,;\,\tau\,;\,\mu\right) = \mathrm{P}\left(\mathcal{H}^*_{\rho(i)}\,|\,\mathcal{G}\,;\,\tau\,;\,\mu\right) \leq \frac{1}{n-2},$$

and a union bound gives us

$$\mathrm{P}\left(\mathcal{H}_{\rho(i)} \cup \mathcal{H}^*_{\rho(i)}\,|\,\mathcal{G}\,;\,\tau\,;\,\mu\right) \leq \frac{2}{n-2}.$$

Hence

$$\begin{aligned}\mathrm{P}\left(\{H > 0\}\,|\,\mathcal{G}\,;\,\tau\,;\,\mu\right) \leq \mathrm{E}\left[H\,|\,\mathcal{G}\,;\,\tau\,;\,\mu\right] \\ &= \sum_{i=2}^{k} \mathrm{P}\left(\mathcal{H}_{\rho(i)} \cup \mathcal{H}^*_{\rho(i)}\,|\,\mathcal{G}\,;\,\tau\,;\,\mu\right) \\ &\leq \frac{2k-2}{n-2} = O\left(n^{-1}\right).\end{aligned}$$

□

5 Conclusion

In [6], Paulden and Smith conjectured that $\mathrm{P}\left(\Delta = \ell\,|\,\mu\right)$ was on the order of n^{-1} for $\ell > 1$ (conjecture 3 on page 16). We agree with this conjecture, even though we have only proved that this probability is on the order of $n^{-1/3+o(1)}$. But even our bound implies that

$$\lim_{n\to\infty} \mathrm{P}\left(\Delta^{(n)} \geq n^{1/3-o(1)}\,\middle|\,\mu\right) = \frac{2}{3},$$

which means that the Prüfer code does have low locality.

Acknowledgements: I would like to thank Boris Pittel for suggesting this problem to me, and David Smith for sending me a copy of Thompson's dissertation.

References

[1] M. Cho, Kim, S. Seo, and H Shin, "Colored Prüfer codes for k-edge colored trees," *The Electronic Journal of Combinatorics*, vol. 10, 2004.

[2] J. Gottlieb, B. Julstrom, G. Raidl, F. Rothlauf. "Prüfer numbers and genetic algorithms: A lesson how the low locality of an encoding can harm the performance of GAs." *Lecture Notes in Computer Science* vol. 1917, Proc. PPSN VI Paris, France, pp. 395-404, September 2000.

[3] H. Mahmoud, *Pólya Urn Models* , New York, New York: CRC Press, 2008.

[4] T. Paulden and D. K. Smith, "From the Dandelion Code to the Rainbow Code: A class of bijective spanning tree representations with linear complexity and bounded locality," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 2, pp. 108-122, April 2006.

[5] T. Paulden and D. K. Smith, "Some Novel Locality Results for the Blob Code Spanning Tree Representation," *Genetic and Evolutionary Computation Conference: Proceedings of the 9th annual conference on genetic and evolutionary computation*, pp. 1320-1327, 2007.

[6] T. Paulden and D. K. Smith, "Developing new locality results for the Prüfer Code using a remarkable linear-time decoding algorithm," *The Electronic Journal of Combinatorics*, vol. 14, August 2007.

[7] F. Rothlauf, *Representations for Genetic and Evolutionary Algorithms*, Second edition. Heidelberg, Germany: Physica-Verlag, 2006.

[8] E. B. Thompson, "The application of evolutionary algorithms to spanning tree problems," Ph.D. dissertation, University of Exeter, U.K., 2003.

[9] E. Thompson, T. Paulden, and D. K. Smith, "The Dandelion Code: A new coding of spanning trees for genetic algorithms," *IEEE Transactions on Evolutionary Computing*, vol. 1,no. 1, 1 pp. 91-100, February 2007.