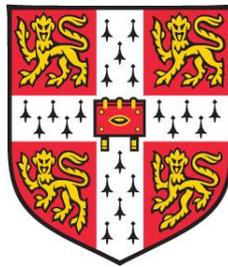


Problems of optimal choice on posets and generalizations of acyclic colourings



Bryn Garrod

Department of Pure Mathematics and Mathematical Statistics

Trinity College

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

3rd May 2011

For Raphaële

Acknowledgements

I am very grateful to my supervisor, Béla Bollobás, for everything that he has done for me over the last three and a half years, from offering me professional assistance, support and advice to organizing research visits to the Rényi Institute in Budapest and to the University of Memphis. He has kept me going in the right direction when I have lost my way or my motivation, asked crucial questions when I have been stuck, and of course taught me and introduced me to a wealth of mathematical knowledge and understanding. At the same time, he has become an integral part of my social life, whether sharing coffee and biscuits in the Pavilion E common room, or inviting me to parties or a Super Bowl in his home. I have enjoyed the many non-mathematical conversations that we have had, even if my lack of cultural knowledge is a constant disappointment!

I should like to thank my collaborators Grzegorz Kubicki, Michał Morayne and Rob Morris. I am particularly grateful to Michał and to Jan Tkocz for their warm hospitality when I visited them in Wrocław and to Rob for the help and advice that he has given me with my work far beyond his role as my departmental mentor. Thank you also to Graham Brightwell, whose discussions with Béla and Rob led to an earlier proof of Theorem 3.2, for his helpful comments on the paper that I wrote with Rob, which became Chapter 3. Thank you to Victor Falgas-Ravry for a useful conversation about Lemma 2.19. I hugely appreciate the constructive comments and corrections provided by my examiners, Andrew Thomason and Mark Walters.

I am grateful to the Engineering and Physical Sciences Research Council for funding me for the duration of my Ph.D. and to the Department of Pure Mathematics and Mathematical Statistics at the University of Cambridge for providing me with a very pleasant and productive working environment. Thank you to the organizers of the following conferences, all of whom supported my attendance financially, and also to the department, Trinity College and the London Mathematical Society for contributing to these costs:

the Bristol Summer School on Probabilistic Techniques in Computer Science, Building Bridges, the Fete of Combinatorics and Computer Science, the LMS-EPSRC Short Course on Probabilistic Combinatorics, the 22nd British Combinatorial Conference, the 14th International Conference on Random Structures and Algorithms, a conference in honour of the 70th birthday of Endre Szemerédi, and the International Congress of Mathematicians 2010.

I have been very lucky with the people whom I have had to guide me through my mathematical studies. I am extremely grateful to my secondary school maths teacher, Peter Jack, for the encouragement and enthusiasm that he provided for more than seven years, and to Tim Gowers and Imre Leader for inspiring me in supervisions and lectures for the next seven.

Thank you to the rest of Team Bollobás for the coffee times, mini-seminars and conference social life, particularly my good friends Tom Coker, Michał Przykucki and Misha Tyomkyn. Thank you to all my other friends for their emotional support, particularly Jen Gold, Alex Holyoake, Phil Horler, Céline Miani, Paul Smith, David Taylor, Charles Vial and Ioanna Vlahou, with whom I have spent most of my spare time over the last three and a half years.

I am hugely indebted to my parents, Martin and Dilys, and my sister and brother, Tania and Ross, for the happy, supportive family background in which I have grown up, and the encouragement that they have always given me.

My greatest thanks go to my wife, Raphaële, whom I met at the start of my Ph.D. and who is as much a part of it as the combinatorics. Thank you to her for marrying me and for looking after me so well. I am thankful for everything that she has done for me, which is more than I can write here.

Declaration of joint work

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text. The first five sections of Chapter 2 are based on joint work with Grzegorz Kubicki¹ of the University of Louisville, USA, and Michał Morayne² of Politechnika Wroclawska, Poland, and about 50% of this is attributable to me. The work that forms the basis of Chapter 3 was joint with Robert Morris,³ then of the University of Cambridge, UK, but now of IMPA, Brazil, and my contribution was about 75%.

¹gkubicki@louisville.edu

²michal.morayne@pwr.wroc.pl

³rob@impa.br

Abstract

This dissertation is in two parts, each of three chapters. In Part 1, I shall prove some results concerning variants of the ‘secretary problem’. In Part 2, I shall bound several generalizations of the acyclic chromatic number of a graph as functions of its maximum degree.

I shall begin Chapter 1 by describing the classical secretary problem, in which the aim is to select the best candidate for the post of a secretary, and its solution. I shall then summarize some of its many generalizations that have been studied up to now, provide some basic theory, and briefly outline the results that I shall prove.

In Chapter 2, I shall suppose that the candidates come as m pairs of equally qualified identical twins. I shall describe an optimal strategy, a formula for its probability of success and the asymptotic behaviour of this strategy and its probability of success as $m \rightarrow \infty$. I shall also find an optimal strategy and its probability of success for the analagous version with c -tuplets.

I shall move away from known posets in Chapter 3, assuming instead that the candidates come from a poset about which the only information known is its size and number of maximal elements. I shall show that, given this information, there is an algorithm that is successful with probability at least $\frac{1}{e}$. For posets with $k \geq 2$ maximal elements, I shall prove that if their width is also k then this can be improved to ${}^{k-1}\sqrt{\frac{1}{k}}$, and show that no better bound of this type is possible.

In Chapter 4, I shall describe the history of acyclic colourings, in which a graph must be properly coloured with no two-coloured cycle, and state some results known about them and their variants. In particular, I shall highlight a result of Alon, McDiarmid and Reed, which bounds the acyclic chromatic number of a graph by a function of its maximum degree. My results in the next two chapters are of this form.

I shall consider two natural generalizations in Chapter 5. In the first, only cycles of length at least l must receive at least three colours. In the second, every cycle must receive at least c colours, except those of length less than c , which must be multicoloured.

My results in Chapter 6 generalize the concept of a cycle; it is now subgraphs with minimum degree r that must receive at least three colours, rather than subgraphs with minimum degree two (which contain cycles). I shall also consider a natural version of this problem for hypergraphs.

Contents

Acknowledgements	v
Declaration of joint work	vii
Abstract	ix
Introduction	1
Part 1: Problems of optimal choice on posets	1
Part 2: Generalizations of acyclic colourings	2
Part 1. Problems of optimal choice on posets	5
Chapter 1. The classical secretary problem	7
1.1. The problem	7
1.2. Outline solution	7
1.3. Variants	9
1.4. Formal model and notation	18
1.5. Useful theorems	21
Chapter 2. How to choose the best twins	27
2.1. Introduction	27
2.2. Notation and basic definitions	29
2.3. The optimal stopping time	30
2.4. The probability of success	36
2.5. Asymptotics	40
2.6. How to choose the best c -tuplets	44
2.7. Open problems	52
Chapter 3. The secretary problem on an unknown poset	55
3.1. Introduction	55

3.2. Lower bounds	56
3.3. Upper bound	66
3.4. Open problems	80
Part 2. Generalizations of acyclic colourings	83
Chapter 4. Acyclic colourings of graphs	85
4.1. Colouring problems	85
4.2. From Nash-Williams to acyclic colourings	87
4.3. Variants	88
4.4. Generalized acyclic colourings	92
4.5. Definitions and a useful tool	93
Chapter 5. Long acyclic colourings and acyclic colourings with many colours	97
5.1. Introduction	97
5.2. The length- l -acyclic chromatic number	97
5.3. The c -colour acyclic chromatic number	103
5.4. Open problems	113
Chapter 6. Small minimum degree colourings of graphs and hypergraphs	115
6.1. Introduction	115
6.2. The degree- r chromatic number of a graph	115
6.3. The degree- r chromatic number of a hypergraph	121
6.4. Open problems	126
Conclusion	129
Part 1: Problems of optimal choice on posets	129
Part 2: Generalizations of acyclic colourings	130
Bibliography	133

Introduction

This dissertation is split into two unrelated parts. In Part 1, I shall consider several problems of optimal choice on posets, which are generalizations of a problem popularly known as the ‘secretary problem’. In Part 2, I shall consider generalizations of acyclic colourings of graphs, a concept whose origins can be traced back to Nash-Williams’s theorem concerning the decomposition of the edge set of a graph into forests.

Part 1: Problems of optimal choice on posets

The classical secretary problem is as follows. There are n candidates to be interviewed for a position as a secretary. They are interviewed one by one and, after each interview, the interviewer must decide whether or not to accept that candidate. If the candidate is accepted then the process stops, and if the candidate is rejected then the interviewer moves on to the next candidate. The interviewer may only accept the most recently interviewed candidate. At each stage, the interviewer knows the complete ranking of the candidates interviewed so far, all of whom are comparable, but has no other measure of their ability. The interviewer is only interested in finding the very best candidate; selecting any other for the job is considered a failure. The aim is to find a strategy that maximizes the probability that the interviewer chooses the best candidate, under the assumption that the candidates are seen in a uniformly random ordering. In Chapter 1, I shall describe the solution to this problem, that there is a strategy that is successful with probability at least $\frac{1}{e}$ and that this is asymptotically best possible. I shall also provide more of the historical background of this problem and some of its generalizations up to now.

In the rest of Part 1, I shall consider two generalizations of this problem. In Chapter 2, I shall first assume that there are $2m$ candidates who are in fact m pairs of identical twins, each pair of twins being equally well-qualified for the job. As in the classical problem, the interviewer knows at each stage how the candidates interviewed so far compare with each other, but has no other measure of their ability. I shall describe an optimal strategy, a

formula for its probability of success and the asymptotic behaviour of this strategy and its probability of success as $m \rightarrow \infty$. I shall also find an optimal strategy and its probability of success for the analogous version on m sets of c -tuplets, and provide bounds that give some indication of their asymptotic behaviour.

I shall move from these known posets to unknown posets in Chapter 3. Specifically, I shall assume that the candidates come from a poset whose size n and number of maximal elements k is known, but whose structure is unknown. Here, the interviewer knows the poset induced by the candidates interviewed so far. I shall describe a strategy that is successful on all posets with given n and k with probability at least $\frac{1}{e}$, which is the best possible bound when $k = 1$, but probably not for $k > 1$. I shall also find a strategy that is successful with probability at least $\sqrt[k-1]{\frac{1}{k}}$ when the width of the poset is known to be the same as the number of maximal elements k and $k \geq 2$. By considering the poset consisting of k disjoint chains, I shall show that no greater probability of success can be guaranteed.

Part 2: Generalizations of acyclic colourings

An acyclic colouring of a graph is a proper vertex-colouring such that every cycle contains vertices of at least three colours. To put it another way, it is an assignment of colours to the vertices such that the graph induced by the vertices in any colour class must be an independent set, and the graph induced by the vertices in any two colour classes must be a forest. The acyclic chromatic number of a graph is the minimum number of colours needed to colour it acyclically. In Chapter 4, I shall describe how this concept came into being as a generalization of the arboricity of a graph, which is the minimum number of forests into which its edge set can be decomposed, and I shall state some of the many results proved about acyclic colourings up to this point. In particular, I shall highlight a result of Alon, McDiarmid and Reed, which bounds the acyclic chromatic number of a graph by a function of its maximum degree. My results in Part 2 will be of this form.

In Chapter 5, I shall consider two generalizations in which the subgraphs under scrutiny are still cycles. In the first, the extra condition that a proper colouring must satisfy is relaxed so that only cycles of length at least l must receive three colours, that is, the

graph induced by the vertices in any two colour classes does not contain a cycle of length at least l . In the second, I shall strengthen the condition so that every cycle must receive at least c colours, with the obvious exception of cycles of length less than c , which must be multicoloured. In this case, the graph induced by the vertices in any x colour classes with $x < c$ does not contain any cycles of length greater than x .

The definitions given so far would work just as well if ‘cycle’ were replaced by ‘2-regular subgraph’ or even ‘subgraph with minimum degree at least 2,’ since cycles fall into both of these categories and any subgraph of either type must contain a cycle. I shall focus my attention in Chapter 6 on subgraphs with minimum degree r ; a graph contains at least as many of these as r -regular subgraphs, and indeed is unlikely to have any r -regular subgraphs for large r , so this is more restrictive. In a proper colouring, it is possible for any bipartite subgraph with minimum degree r to receive only two colours; I shall insist that it receive three. I shall also consider the same problem for u -uniform hypergraphs, under the assumption that a proper colouring is one in which every edge is multicoloured.

Part 1

Problems of optimal choice on posets

CHAPTER 1

The classical secretary problem

1.1. The problem

The exact origins of the classical secretary problem are complicated (and the subject of Ferguson's history of the problem [26]), but the problem was popularized by Martin Gardner [32, 33] in his *Scientific American* column in February 1960, as the game *googol*. The problem itself is simple to state, and its 'secretary problem' formulation is as follows. There are n candidates to be interviewed for a position as a secretary. They are interviewed one by one and, after each interview, the interviewer must decide whether or not to accept that candidate. If the candidate is accepted then the process stops, and if the candidate is rejected then the interviewer moves on to the next candidate. The interviewer may only accept the most recently interviewed candidate. At each stage, the interviewer knows the complete ranking of the candidates interviewed so far, all of whom are comparable, but has no other measure of their ability. The interviewer is only interested in finding the very best candidate; selecting any other for the job is considered a failure. The aim is to find a strategy that maximizes the probability that the interviewer chooses the best candidate, under the assumption that the candidates are seen in a uniformly random ordering.

1.2. Outline solution

The solution to the classical secretary problem is now folklore but was first published by Lindley [57], and I shall give an outline of it here.

It is obvious that the interviewer should only consider accepting a candidate who is the best seen so far. It is intuitively clear that a candidate should not be accepted very early on even if he or she is the best seen so far, since there is a reasonable probability that a small number of candidates all come from near the bottom of the ranking. Conversely, the interviewer should not wait too long, or the best candidate will probably be missed and

the interviewer will not have the opportunity to select anyone. Furthermore, if a strategy dictates that the i^{th} candidate should be accepted if he or she is the best candidate seen so far, it seems reasonable that the $(i+1)^{\text{th}}$ should be accepted in the same circumstances, since more candidates have been seen and that candidate's credentials are stronger. From these observations, it might be expected that some sort of threshold should be passed before the interviewer considers choosing a candidate.

Using backward induction, one can prove exactly that. This will be described in more detail in Section 1.5; for now, I shall assume the following consequence of it without proof. For some k , the strategy “reject the first k candidates, and accept the next who is the best seen so far” is optimal. As an aside, it is worth noting that there might be more than one optimal strategy: for example, when there are exactly two candidates, the two possible deterministic strategies are equivalent, and both are equivalent to tossing a coin to choose between the two candidates.

Let W be the event that, using this strategy, the interviewer chooses the best candidate and let B_i be the event that the i^{th} candidate interviewed is the best candidate. Let A_i be the event that the interviewer is still interviewing by the time the i^{th} candidate arrives, that is, that the best of the first $i-1$ candidates interviewed is in the first k interviewed. Then B_i and A_i are independent, and the probability of winning is given by

$$\begin{aligned} \mathbb{P}(W) &= \sum_{i=k+1}^n \mathbb{P}(B_i)\mathbb{P}(W|B_i) \\ &= \sum_{i=k+1}^n \mathbb{P}(B_i)\mathbb{P}(A_i) \\ &= \sum_{i=k+1}^n \frac{1}{n} \cdot \frac{k}{i-1} \\ &= \frac{k}{n} \sum_{j=k}^{n-1} \frac{1}{j}. \end{aligned}$$

A value of k that maximizes this satisfies

$$\frac{k}{n} \sum_{j=k}^{n-1} \frac{1}{j} \geq \frac{k-1}{n} \sum_{j=k-1}^{n-1} \frac{1}{j} \quad \text{and} \quad \frac{k}{n} \sum_{j=k}^{n-1} \frac{1}{j} \geq \frac{k+1}{n} \sum_{j=k+1}^{n-1} \frac{1}{j},$$

that is,

$$\sum_{j=k}^{n-1} \frac{1}{j} \geq \frac{k-1}{k-1} = 1 \quad \text{and} \quad \sum_{j=k+1}^{n-1} \frac{1}{j} \leq \frac{k}{k} = 1,$$

from which simple integration arguments give

$$\frac{n}{e} - 1 \leq k \leq \frac{n}{e} + \frac{e-1}{e}.$$

From this, it is clear that for such k

$$\lim_{n \rightarrow \infty} \frac{k}{n} = \frac{1}{e}$$

and that the probability of winning tends to the same limit. In fact, in Chapter 3, it will become evident that this is a lower bound.

1.3. Variants

A problem posed by Cayley [18] may have inspired the classical secretary problem. This is what is now known as the ‘full information’ case, where the candidates’ abilities are represented by real random variables from a known distribution, and the aim is to maximize the expected ability of the chosen candidate. The uniform distribution $U[0, 1]$ was considered by Moser [62]; Guttman [46] found an optimal strategy for a general distribution and also gave an explicit optimal strategy for the normal distribution $N(0, 1)$. His general optimal strategy is to accept a candidate if there are at least m candidates remaining after that one and his or her ability is at least E_m , for some $(E_m)_{m \in \mathbb{N}}$. For $U(0, 1)$, the first few values of E_m are 0.5, 0.625, 0.6953, 0.7417 and 0.775, and for $N(0, 1)$ they are 0, 0.3992, 0.6298, 0.7904 and 0.9127. The expected ability is the first threshold for acceptance, E_1 .

Since 1960, many generalizations of the classical secretary problem have been considered. Freeman [28] wrote an extensive review of the area in 1983, which shows how many versions had already been considered by then. I shall describe only some of them, and some more recent results.

Besides the full information case, the most obvious generalization might be to be more flexible over what constitutes success, and to try to minimize the expected rank (viewing the best candidate as being from rank 1 and the worst from rank n) rather than insisting

on choosing the best candidate. This version was solved by Chow, Moriguti, Robbins and Samuels [21]. Perhaps surprisingly, in the limit as $n \rightarrow \infty$ the optimal expected rank tends to a constant rather than a multiple of n , namely,

$$\prod_{j=1}^{\infty} \left(\frac{j+2}{j} \right)^{\frac{1}{j+1}} \approx 3.8695.$$

They showed that an optimal strategy is to accept a candidate who is in the top k seen so far as long as at least i_k candidates have been seen in total, for some thresholds i_k . They also showed that the i_k satisfy

$$\lim_{n \rightarrow \infty} \frac{i_k}{n} = \prod_{j=k}^{\infty} \left(\frac{j}{j+2} \right)^{\frac{1}{j+1}}.$$

This means that, for large n , once we have seen approximately 26% of candidates we should be prepared to accept the next one who is the best so far, after 45% we should accept anyone who is one of the top two seen so far, after 56% one of the top three, after 64% one of the top four, after 69% one of the top five and so on. At the other end of the scale, once we have seen 99% of candidates we should accept anyone who is in the top 200 and after 99.9% anyone in the top 2000.

Yang [80] considered the situation where, as well as being allowed to offer the job to the most recently interviewed candidate, who would accept it, the interviewer can offer the job to any of the other candidates seen so far, who is still available with probability $q(r)$, where q is a known non-increasing function of the number r of candidates interviewed since that one. If a candidate is unavailable, he or she never becomes available again. The aim is to choose the best possible secretary, as in the classical secretary problem. Smith [74] studied a version where the job can only be offered to the currently interviewed candidate, but there is some fixed probability that the candidate will refuse the offer. Petruccelli [67] worked on these two problems simultaneously, that is, Yang's problem with $q(r)$ still non-increasing but with $q(0)$ no longer forced to be 1.

In particular, Petruccelli considered the case where the probabilities form a geometric progression, that is, where $q(r) = qp^r$ for some p and q . He proved that there are two cases, depending on p , q and the number n of candidates. If $\sum_{r=0}^{\infty} q(r) = \frac{q}{1-p} > n - 1$ then an optimal strategy is to observe all n candidates and then to offer the job to the

best one. If $\frac{q}{1-p} \leq n-1$, then an optimal strategy is to wait until s_n candidates have been seen, for some s_n , and then to offer the job to the best one seen so far. If this one is unavailable, then the interviewer should continue interviewing and offer the job to the next candidate who is the best seen so far. If this one is unavailable, then the interviewer should continue as before, and so on. He gave an explicit formula for s_n , namely, the smallest value of s for which

$$\prod_{k=s+1}^{n-1} \left(1 + \frac{1-q}{k}\right) \leq \left[q \left(1 + \frac{1-q}{s(1-p)}\right) \right]^{-1}.$$

He also showed that both $\frac{s_n}{n}$ and the probability of success tend to $q^{\frac{1}{1-q}}$ as $n \rightarrow \infty$. This is independent of p , which means that for large n there is effectively no benefit to being allowed to recall a previous candidate, since p can be arbitrarily small. However, this is not surprising, since if p is fixed then only a constant number of candidates are likely to be available at any one point, even as $n \rightarrow \infty$.

Gusein-Zade [45] allowed selection of any of the top r candidates to count as success, and showed that an optimal strategy is of the same form as in the expected rank case, that is, an optimal strategy is to accept a candidate who is in the top k seen so far as long as at least i_k candidates have been seen in total, for some thresholds i_k . He showed that for $r=2$ the limiting probability of success as $n \rightarrow \infty$ is about 0.5736. Frank and Samuels [27] proved that the optimal probability of success $p(n, r)$ satisfies

$$\lim_{r \rightarrow \infty} \lim_{n \rightarrow \infty} (1 - p(n, k))^{\frac{1}{r}} = 1 - t^*, \quad \text{where } t^* = \lim_{n \rightarrow \infty} \frac{i_1}{n} \approx 0.2834.$$

Gilbert and Mosteller [40] considered what could be called the inverse of this problem: the interviewer is allowed to pick up to r candidates and wins if any of them is the best. (Many other variations of the secretary problem are included in the same paper.) They showed that an optimal strategy is to wait until $t(n, r)$ candidates have been seen, for some function $t(n, r)$, then to pick the next who is the best seen so far, and then to play an optimal strategy for the remaining candidates and $r-1$ choices. They found an iterative method to calculate the limits

$$u_r = \lim_{n \rightarrow \infty} \frac{t(n, r)}{n},$$

showed that the first few values of u_r are e^{-1} , $e^{-\frac{3}{2}}$, $e^{-\frac{47}{24}}$ and $e^{-\frac{2761}{1152}}$, and showed that as $n \rightarrow \infty$ the optimal probability of success tends to

$$\sum_{i=1}^r u_i.$$

Some authors wondered what would happen if the number of candidates were unknown, but the distribution of that number, the random variable N , were known. Presman and Sonin [69] found an explicit optimal strategy for a general distribution, where the aim is to choose the best candidate. They also showed that if the number of candidates is uniform in $[n]$, then an optimal strategy is of the same form as in the classical secretary problem, but where the threshold is asymptotically equivalent to $\frac{n}{e^2}$ and the probability of success tends to $\frac{2}{e^2} \approx 0.2707$ as $n \rightarrow \infty$.

Gianini-Pettitt [39] considered the minimal expected rank version of this problem, and restricted her attention to distributions of the form

$$\mathbb{P}(N = x | N \geq x) = (n - x + 1)^{-\alpha},$$

for some α . She proved that, as when the number of candidates is known, an optimal strategy is of the form ‘accept the i^{th} candidate if it is one of the best $k(i)$ seen so far,’ but that $k(i)$ is not necessarily an increasing function. She also proved the surprising fact that if N_1 and N_2 are possible distributions of the number of candidates and N_1 is stochastically smaller than N_2 , that is, $\mathbb{P}(N_1 \leq x) \geq \mathbb{P}(N_2 \leq x)$ for all x , this does not imply that the minimum expected rank decreases. One example of this is that if N_1 is uniformly distributed on $[n]$, that is, if $\alpha = 1$, then the optimal expected rank tends to infinity as $n \rightarrow \infty$, whereas if $N_2 = n$ with probability 1 then, as shown by Chow, Moriguti, Robbins and Samuels [21], the optimal expected rank tends to about 3.8695. In fact, she showed that the optimal expected rank tends to infinity if $\alpha < 2$ and to the Chow, Moriguti, Robbins and Samuels limit of approximately 3.8695 if $\alpha > 2$, and if $\alpha = 2$ then the lim inf of the optimal expected rank is greater than 3.8695 and the lim sup is finite.

More generally, one could consider problems of optimal choice on more complicated systems; up to this point the assumption has always been that there are n rankable

candidates. Kuchta and Morayne [56] considered a version of the classical secretary problem where the interviewer has k ‘lives’: if all n candidates are interviewed without any of them being chosen, then a new set of n candidates is interviewed, and the aim is to choose the best of these, and so on. At most k sets of candidates are allowed to be interviewed in total. They showed that, for some function $t(n, k)$, an optimal strategy is to ignore the first $t(n, k)$ candidates from the first set, and accept the next candidate who is the best seen so far; if no such candidate appears, then ignore the first $t(n, k - 1)$ candidates from the next set and accept the next candidate who is the best seen so far, and so on. They showed that

$$\lim_{n \rightarrow \infty} \frac{t(n, k)}{n}$$

exists, and denoting it by a_k , that $a_{k+1} = e^{a_k - 1}$, where of course $a_1 = \frac{1}{e}$.

Stadje [75] introduced the idea that the candidates could be ranked separately in each of $k > 1$ different criteria, with the interviewer wishing to select a candidate who is maximal in at least one of them, and Gnedin [41] solved the version where these rankings are random and independent of each other. He proved that an optimal strategy is to wait until a certain number of candidates have been seen and then to select the next who is best according to at least one criterion, and that the limiting values of this threshold and the probability of success are both $\sqrt[k-1]{\frac{1}{k}}$. Gnedin has also produced a more general survey of multicriteria problems [42]. In fact, these last two problems could both be viewed as versions of the secretary problem on k disjoint chains, which I shall solve in Chapter 3, with an extra restriction, about which I shall say more at the time.

Moving slightly further away from the classical secretary problem, Kubicki and Morayne [55] considered the problem on a directed path, where at each stage the selector knows the directed graph induced by the vertices seen so far and wishes to choose the end-vertex with no edge going out of it. This is similar to the classical secretary problem, but each candidate can only be compared with the one immediately above it or below it in the ranking. They showed that an optimal strategy is to wait until the first time t when the induced graph has $n - t + 1$ connected components and to pick the t^{th} vertex. Note that this is the first time when the selector can be sure that the sought after end-vertex has been seen: if at time t the induced graph has $n - t + 1$ connected components, then

the remaining $n - t$ vertices must be used to join components together, and so none of them is the end-vertex. Note also that this strategy is independent of whether or not the t^{th} vertex is an end-vertex of its component; if it is not, then it does not make any difference which vertex is chosen. They showed that the probability of success p_n satisfies

$$\lim_{n \rightarrow \infty} p_n \sqrt{n} = \frac{\sqrt{\pi}}{2}.$$

Przykucki and Sulkowska [71] adapted this problem in a similar way to Gusein-Zade's version of the classical secretary problem, so that choosing the end-vertex or its neighbour counts as success. In this case, the optimal stopping time and its analysis are more complicated, but their numerical analysis shows that the probability of success behaves approximately like $\frac{1.26}{\sqrt{n}}$. For comparison with the previous result, $\frac{\sqrt{\pi}}{2} \approx 0.8862$.

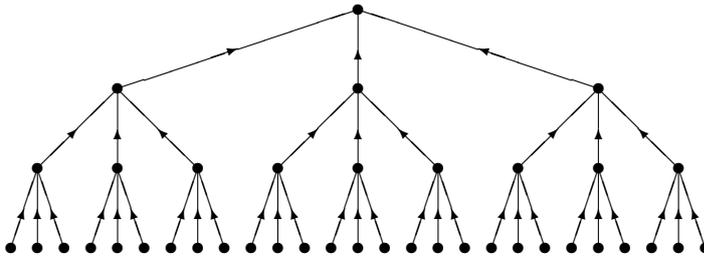


FIGURE 1.1. Directed ternary tree of depth 3.

Morayne and Sulkowska [61] studied a variation of the directed path version, in this case working with the complete k -ary directed rooted tree of depth n (see Figure 1.1) and again assuming that the selector knows the induced directed graph at each stage of the process. They found a lower bound for the probability of success of an optimal strategy by considering a natural (but not necessarily optimal) strategy: select the currently examined element if there is a directed path of length n terminating at it. If this ever happens, then the strategy must be successful. In this way, they showed that on a binary tree the limit of the optimal probability of success is at least $2 \log 2 - 1 \approx 0.3863$ and that on a ternary tree it is at least $\frac{3}{2} \log 3 - 2 + \frac{\pi}{2\sqrt{3}} \approx 0.5548$, and that it tends to 1 as $k \rightarrow \infty$.

Przykucki [70] posed a problem concerning the random graph $G_{n,p}$ with n vertices and any two connected by an edge with probability p independently of the others. Again, the selector knows the graphs induced by the vertices seen so far, and wishes to find a vertex of full degree, that is, degree $n - 1$. He showed that an optimal strategy is to wait

until $k(n, p)$ vertices have been seen and then to pick the next vertex connected to every vertex seen so far, for some $k(n, p)$. He showed that, for fixed $p \in (0, 1)$,

$$k(n, p) = \log_{\frac{1}{p}} n + O(1)$$

as $n \rightarrow \infty$, and found a formula for the probability of its success, which of course tends to zero more quickly than np^{n-1} , which is an upper bound for the probability that a vertex of full degree exists.

In this dissertation, the type of generalization that I shall consider is to put partial orders other than a total order on the candidates. Here, the selector knows the poset from which the elements are taken and the poset induced by the elements observed so far, and wishes to choose an element that is maximal in the ground poset.

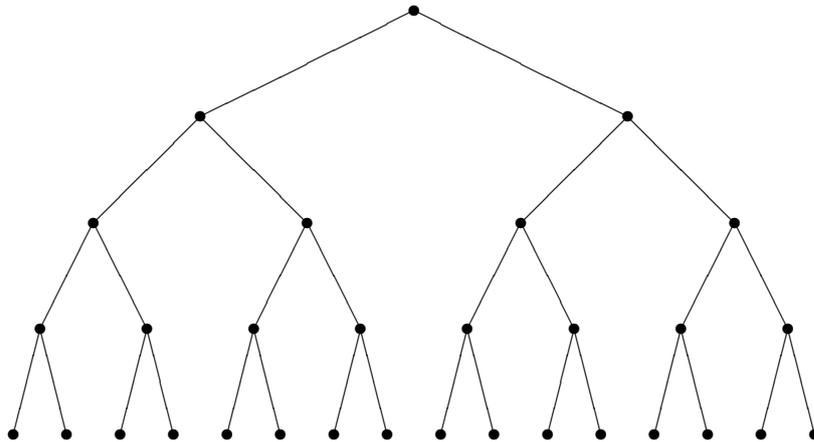


FIGURE 1.2. Binary tree of depth 4.

A generalization due to Morayne [60] is to consider the case of a binary tree of depth n (see Figure 1.2). Intuitively, it seems unlikely that a random selection of nodes would come from a subtree with a maximum other than the global maximum unless they are linearly ordered, and he showed that this is indeed the case. An optimal strategy here is to select the maximum out of the elements seen so far when the poset induced by these elements is either linear of length greater than $\frac{n}{2}$ or non-linear with a unique maximum. He showed further that as $n \rightarrow \infty$, the probability of success tends to 1.

Kaźmierczak [49] added a ‘witness’ to the classical secretary problem, an extra element w in the poset that lies immediately below the maximal element but cannot be compared with any other element. Tkocz [77] extended this concept to put the witness below the

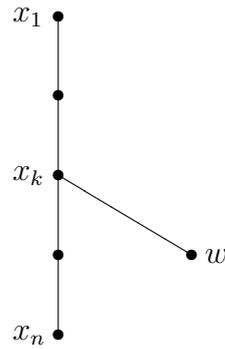


FIGURE 1.3. Tkocz's poset.

k^{th} highest element (see Figure 1.3). He gave an explicit optimal strategy, which uses three different thresholds, essentially depending on the size of k relative to n . If the poset induced by the elements seen so far is linear and of length greater than the threshold then a maximal element should be accepted; if it is not linear then an optimal strategy for the classical secretary problem on k elements should be followed. He also calculated the asymptotic probabilities of success for $k = 2$ and 3 , approximately 0.415 and 0.384 respectively, compared with the figure of approximately 0.573 for $k = 1$ obtained by Kaźmierczak.

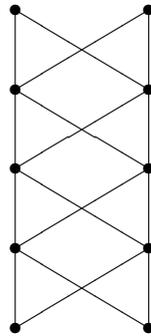


FIGURE 1.4. Five pairs of identical twins.

Michał Morayne, Grzegorz Kubicki and I [34] considered the case of m pairs of ‘twins’, where there are m levels with two incomparable elements on each level (see Figure 1.4); I shall present these results in Chapter 2. I shall show that an optimal strategy is to wait until elements from a certain threshold number of levels have been seen and then to select the next element that is maximal and whose twin has already been seen. I shall further show that as $m \rightarrow \infty$, this threshold behaves roughly like $0.4709m$ and the probability of success tends to approximately 0.7680 . Calculating these asymptotic values for the

natural extension to ‘ c -tuplets’ for $c > 2$, is a harder problem, and I shall provide some bounds.

A further interesting generalization, due to Preater [68], was an attempt to find an algorithm that was successful on *all* posets of a given size with positive probability. Surprisingly, he proved that there is such a ‘universal’ algorithm (depending only on the size of the poset), which is successful on every poset with probability at least $\frac{1}{8}$. In this algorithm, an initial random number of elements are rejected and a subsequent element is accepted according to randomized criteria. A slightly modified version of the algorithm, also suggested by Preater, was analysed by Georgiou, Kuchta, Morayne and Niemiec [36], and gave an improved lower bound of $\frac{1}{4}$ for the probability of success. More recently, Kozik [54] introduced a ‘dynamic threshold strategy’ and showed that it was successful with probability at least $\frac{1}{4} + \varepsilon$, for some $\varepsilon > 0$ and for all sufficiently large posets. When I was about to submit this dissertation, Michał Morayne drew my attention to a very recent paper of Freij and Wästlund [29]. In it, they describe a strategy that is successful with probability at least $\frac{1}{e}$. This cannot be improved, since the best possible probability of success in the classical secretary problem, on a totally ordered set, is $\frac{1}{e}$. I shall say more about this in Section 3.4.

Before Kozik, Freij and Wästlund had published their results, Robert Morris and I [35] showed that, given any poset, there is an algorithm that is successful with probability at least $\frac{1}{e}$, so, in this sense, the total order is the hardest possible partial order. I shall present these results in Chapter 3. In fact, this algorithm depends only on the size of the poset and its number of maximal elements, so it is universal for any family where these are given. It is therefore natural to ask which is the hardest partial order with a given number of maximal elements. The most obvious choice is the poset consisting of k disjoint chains. I shall give an asymptotically sharp lower bound on the probability of success in the problem of optimal choice on k disjoint chains, and show that it is at least as hard as on any poset with k maximal elements and of width k , that is, whose largest antichain has size k .

1.4. Formal model and notation

In this section, I shall define formally the probability space in which I shall work in Chapters 2 and 3.

This probability space will depend on a poset (P, \prec) with $P = \{x_1, \dots, x_n\}$. In Chapter 2, this will be a known poset, the poset of m pairs of identical twins or, later, m sets of identical c -tuplets. In Chapter 3, this will be a fixed but unknown poset. Let $\max_{\prec}(P)$ denote the set of its maximal elements, that is,

$$\max_{\prec}(P) = \{x \in P : \nexists y \text{ such that } x \prec y\}.$$

The subscript in \max_{\prec} will be suppressed when it is clear from the context.

Given (P, \prec) , I shall work with a probability space $(\Omega_P, \mathcal{F}_P, \mathbb{P}_P)$, with \mathbb{E}_P defined in the obvious way. The subscripts will be suppressed when they are clear from the context, as they will be for the rest of this section. The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is defined as follows. Set $\Omega = S_n \times [0, 1]$, where S_n is the permutation group on $[n]$, and $\mathcal{F} = \mathcal{P}(S_n) \times \mathcal{B}$, where \mathcal{B} is the Borel σ -algebra. Let $\mathbb{P} = \mu \times \lambda$, where μ is the uniform probability measure, that is,

$$\mu(\{\rho\}) = \frac{1}{n!}$$

for all $\rho \in S_n$, and λ is the Lebesgue measure. In other words, $(\rho, \delta) \in \Omega$ is picked uniformly at random. Given $(\rho, \delta) \in \Omega$, the ρ -co-ordinate will determine the order in which elements of P appear and the δ -co-ordinate will allow the introduction of randomness independent of this order into our algorithms. This will not be needed in Chapter 2; in Chapter 3, the δ -co-ordinate will determine an initial number of elements to reject without considering. The reason why continuous space and Lebesgue measure are used, despite the fact that all of the randomized strategies considered pick one of a finite number of options, is that this allows them all to lie in the same probability space.

Write $P^{[n]}$ for the set of permutations of P , and let $\pi : \Omega \rightarrow P^{[n]}$ be the random variable defined by

$$\pi(\rho, \delta)(i) = x_{\rho(i)}.$$

Let \mathcal{P}_t denote the set of all posets with vertex set $[t] = \{1, \dots, t\}$. Let $(P_t)_{t \in [n]}$ be a family of random variables with P_t representing the poset seen at time t . Formally, $P_t : \Omega \rightarrow \mathcal{P}_t$ and each $P_t(\rho, \delta) = ([t], \prec_t)$ is defined by

$$\forall i, j \in [t], i \prec_t j \iff \pi(i) \prec \pi(j).$$

The poset P_t is the natural description of what is seen at time t as the elements of P appear one by one.

Let $(\mathcal{F}_t)_{t \in [n]}$ be the sequence of σ -algebras with each \mathcal{F}_t generated by the random variables P_1, \dots, P_t , that is,

$$\mathcal{F}_t = \sigma(P_1, \dots, P_t) = \sigma(P_t),$$

the second equality holding since P_t is a labelled poset and thus its value determines the values of P_1, \dots, P_{t-1} . The σ -algebra \mathcal{F}_t can be thought of as the information known at time t about where we are in the universe Ω . Since P_t takes only finitely-many values, \mathcal{F}_t has a simple structure; it is the pre-images in Ω of the possible values of P_t and the unions of these pre-images. These pre-images are called the atoms of \mathcal{F}_t .

Let \mathcal{F}'_t be the projection of \mathcal{F}_t onto $\mathcal{P}(S_n)$. Since definitions have so far depended only on the ρ -co-ordinate of $(\rho, \delta) \in \Omega$, it is clear that, for each t ,

$$\mathcal{F}_t = \{A \times [0, 1] : A \in \mathcal{F}'_t\}.$$

In other words, (ρ_1, δ_1) and (ρ_2, δ_2) are in the same atom of \mathcal{F}_t if and only if ρ_1 and ρ_2 are in the same atom of \mathcal{F}'_t , which happens if and only if the labelled posets induced by the first t elements $\pi(1), \dots, \pi(t)$ are identical.

A *stopping time* is a random variable τ taking values in $[n]$ and satisfying the property

$$\{\tau = t\} \in \mathcal{F}_t,$$

that is, the decision to stop at time t is based only on the values of P_1, \dots, P_t .

I shall give a brief reminder of the formal definitions of conditional expectation and probability, which in the finite world are intuitive concepts. For more details, see page 304 of Galambos [30] or page 313 of Chung [23], for example. Let X be a random variable.

Then the *conditional expectation* of X given \mathcal{F}_t , denoted by $\mathbb{E}(X|\mathcal{F}_t)$ is any \mathcal{F}_t -measurable random variable satisfying

$$\int_A \mathbb{E}(X|\mathcal{F}_t) = \int_A X \quad \text{for all } A \in \mathcal{F}_t.$$

Any two random variables satisfying these conditions are equal with probability 1. As \mathcal{F}_t is finite, this random variable is constant on each atom of \mathcal{F}_t and takes the average value of X on that atom, that is,

$$\mathbb{E}(X|\mathcal{F}_t)(\omega) = \frac{\int_A X}{\mathbb{P}(A)} = \mathbb{E}(X|A),$$

where A is the atom of \mathcal{F}_t containing ω . In the same way that the probability of an event E is the expectation of the indicator function $\mathbb{1}(E)$ of this event, the *conditional probability* of E given \mathcal{F}_t is defined by

$$\mathbb{P}(E|\mathcal{F}_t) = \mathbb{E}(\mathbb{1}(E)|\mathcal{F}_t),$$

that is, since \mathcal{F}_t is finite,

$$\mathbb{P}(E|\mathcal{F}_t)(\omega) = \mathbb{E}(\mathbb{1}(E)|\mathcal{F}_t)(\omega) = \frac{\int_A \mathbb{1}(E)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(A)} = \mathbb{P}(E|A),$$

where A is the atom of \mathcal{F}_t containing ω .

Define the family of random variables $(Z_t)_{t \in [n]}$ by

$$Z_t = \mathbb{P}(\pi(t) \in \max(P)|\mathcal{F}_t),$$

that is, the random variable Z_t is the probability that the t^{th} element observed is maximal given P_1, \dots, P_t . The general aim will be to choose stopping times τ to maximize $\mathbb{P}(\pi(\tau) \in \max(P))$. This quantity is equal to $\mathbb{E}(Z_\tau)$ (see page 45 of Chow, Robbins and Siegmund [22], for example):

$$\begin{aligned} \mathbb{E}(Z_\tau) &= \int_{\Omega} Z_\tau = \sum_{t=1}^n \int_{\{\tau=t\}} Z_t = \sum_{t=1}^n \int_{\{\tau=t\}} \mathbb{P}(\pi(t) \in \max(P)|\mathcal{F}_t) \\ &= \sum_{t=1}^n \int_{\{\tau=t\}} \mathbb{1}(\pi(t) \in \max(P)) = \int_{\Omega} \mathbb{1}(\pi(\tau) \in \max(P)) = \mathbb{P}(\pi(\tau) \in \max(P)), \end{aligned}$$

the fourth equality holding by the definition of conditional expectation since $\{\tau = t\} \in \mathcal{F}_t$. These equivalent formulations will be useful later; this conditional probability can be treated as a pay-off offered at each step of the process.

Recall that \mathcal{F}'_t is the projection of \mathcal{F}_t onto $\mathcal{P}(S_n)$. A *randomized stopping time* is a random variable τ taking values in $[n]$ and satisfying the property

$$\{\tau = t\} \in \mathcal{F}'_t \times \mathcal{B},$$

that is, the decision to stop at time t is based on the values of P_1, \dots, P_t and on some \mathcal{B} -measurable random variable. The randomized stopping times considered in Chapter 3 will be convex combinations of a finite number of true stopping times, so if such a randomized stopping time gives a certain probability of success, then there is a true stopping time with at least that probability of success. In fact, this is true in general, as proved by Ghoussoub [38].

1.5. Useful theorems

In this section, I shall define backward induction formally and use it to solve the classical secretary problem. I shall also state a theorem that gives an optimal stopping time for monotone processes, to be defined later.

The next theorem, which is proved as Theorem 3.2 on page 50 of Chow, Robbins and Siegmund [22], gives in principle an optimal stopping time for any finite process, although in practice it might be hard to define such a stopping time more explicitly. It formalizes the concept of backward induction. Informally, it can be described as follows.

Define a new random variable for each t , the *value* of the process at time t . This is the expected pay-off ultimately accepted given what has happened so far. These values are calculated inductively, starting at the end. The value of the process at the final step is just the final pay-off offered. The value of the process at each earlier step is the maximum of the currently offered pay-off and the expected value of the process at the next step. An optimal strategy is to stop when the currently offered pay-off is at least the expected value at the next step.

In the theorem below, the pay-offs offered are the W_t and the values at each step are the γ_t . The σ -algebras \mathcal{A}_t represent what is known at time t . As a reminder, being

\mathcal{A}_t -measurable means that $\sigma(W_t) \subset \mathcal{A}_t$, that is, the value of W_t is determined by what is known at time t or, in the finite world, W_t is constant on each atom of \mathcal{A}_t . In fact, the nested condition means that $\mathcal{A}_t \supset \sigma(W_1, \dots, W_t)$. The conclusion of the theorem is that the strategy that stops at the first t when $W_t = \gamma_t$ (or, equivalently, when W_t is at least as large as the expected value of γ_{t+1} given \mathcal{A}_t) is indeed a stopping time and achieves the optimal value.

THEOREM 1.1. *Let $\mathcal{A}_1 \subset \dots \subset \mathcal{A}_n$ be a nested sequence of σ -algebras and let W_1, \dots, W_n be a sequence of random variables with each W_t being \mathcal{A}_t -measurable. Let $\mathcal{C}_{(\mathcal{A}_t)}$ be the class of stopping times relative to $(\mathcal{A}_t)_{t \in [n]}$ and let v^* be given by*

$$v^* = \sup_{\tau \in \mathcal{C}_{(\mathcal{A}_t)}} \mathbb{E}(W_\tau).$$

Define successively $\gamma_n, \gamma_{n-1}, \dots, \gamma_1$ by setting

$$\begin{aligned} \gamma_n &= W_n, \\ \gamma_t &= \max \left\{ W_t, \mathbb{E}(\gamma_{t+1} | \mathcal{A}_t) \right\}, \quad t = n-1, \dots, 1. \end{aligned}$$

Let

$$\tau^* = \min\{t : W_t = \gamma_t\}.$$

Then $\tau^* \in \mathcal{C}_{(\mathcal{A}_t)}$ and

$$\mathbb{E}(W_{\tau^*}) = \mathbb{E}(\gamma_1) = v^* \geq \mathbb{E}(W_\tau) \text{ for all } \tau \in \mathcal{C}_{(\mathcal{A}_t)}.$$

□

This theorem can now be used to justify the assertion in Section 1.2 that an optimal strategy is of the form “reject the first k candidates, and accept the next who is the best seen so far.” Recall from Section 1.4 that

$$Z_t = \mathbb{P}(\pi(t) \in \max(P) | \mathcal{F}_t),$$

that is, the probability that the t^{th} candidate is the best one given what is known at this point. Backward induction will be applied with the Z_t corresponding to the W_t , with \mathcal{F}_t to \mathcal{G}_t and with δ_t to γ_t .

Since the random variables Z_t are independent, the values of Z_1, \dots, Z_t give no information about the values of Z_{t+1}, \dots, Z_n , and therefore $\mathbb{E}(\delta_{t+1}|\mathcal{F}_t)$ is constant on all atoms of \mathcal{F}_t and equal to $\mathbb{E}(\delta_{t+1})$. Therefore, define the function $v : [n] \rightarrow \mathbb{R}$ by

$$v(t) = \mathbb{E}(\delta_t)$$

and note that backward induction tells us that the stopping time that stops at the first t such that $Z_t \geq \mathbb{E}(\delta_{t+1}|\mathcal{F}_t) = v(t+1)$ is optimal. By definition,

$$v(t) = \mathbb{E}(\max\{Z_t, v(t+1)\}) \geq v(t+1),$$

whereas $\frac{t}{n}$, the potential non-zero value of Z_t , is a non-decreasing function of t . Therefore, there exists k such that

$$\begin{aligned} \frac{t}{n} &< v(t+1) && \text{if } t \leq k, \\ \frac{t}{n} &\geq v(t+1) && \text{if } t > k, \end{aligned}$$

and an optimal strategy is “reject the first k candidates, and accept the next who is the best seen so far,” as claimed.

In fact, this argument is not even necessary, as the classical secretary problem is a sufficiently straightforward process that backward induction can be used to give an explicit optimal strategy. This is because in this case the δ_t can be calculated explicitly, as in the next lemma, and these define an optimal strategy.

LEMMA 1.2. *For all $t \leq n$, if*

$$\sum_{i=t}^n \frac{1}{i} \leq 1,$$

then

$$\mathbb{E}(\delta_t|\mathcal{F}_{t-1}) = \frac{t-1}{n} \sum_{i=t-1}^{n-1} \frac{1}{i},$$

that is, it is the constant random variable taking that value; otherwise,

$$\mathbb{E}(\delta_t|\mathcal{F}_{t-1}) = \mathbb{E}(\delta_{t+1}|\mathcal{F}_t).$$

PROOF. By definition, $\delta_n = Z_n$ and so

$$\mathbb{E}(\delta_n | \mathcal{F}_{n-1}) = \frac{1}{n} = \frac{n-1}{n} \cdot \frac{1}{n-1}.$$

For $t \leq n-1$, if

$$\frac{t}{n} \geq \frac{t}{n} \sum_{i=t}^n \frac{1}{i} = \mathbb{E}(\delta_{t+1} | \mathcal{F}_t),$$

then

$$\begin{aligned} \mathbb{E}(\delta_t | \mathcal{F}_{t-1}) &= \mathbb{P}(Z_t \geq \mathbb{E}(\delta_{t+1} | \mathcal{F}_t)) \cdot \mathbb{E}(Z_t | (Z_t \geq \mathbb{E}(\delta_{t+1} | \mathcal{F}_t))) \\ &\quad + \mathbb{P}(Z_t < \mathbb{E}(\delta_{t+1} | \mathcal{F}_t)) \cdot \mathbb{E}(\mathbb{E}(\delta_{t+1} | \mathcal{F}_t) | (Z_t < \mathbb{E}(\delta_{t+1} | \mathcal{F}_t))) \quad (1.1) \\ &= \frac{1}{t} \cdot \frac{t}{n} + \frac{t-1}{t} \cdot \frac{t}{n} \sum_{i=t}^{n-1} \frac{1}{i} \\ &= \frac{t-1}{n} \sum_{i=t-1}^{n-1} \frac{1}{i}. \end{aligned}$$

If

$$\frac{t}{n} < \frac{t}{n} \sum_{i=t}^n \frac{1}{i} = \mathbb{E}(\delta_{t+1} | \mathcal{F}_t),$$

then, from the formula in (1.1), it is the case that

$$\begin{aligned} \mathbb{E}(\delta_t | \mathcal{F}_{t-1}) &= 0 \cdot \frac{t}{n} + 1 \cdot \mathbb{E}(\delta_{t+1} | \mathcal{F}_t) \\ &= \mathbb{E}(\delta_{t+1} | \mathcal{F}_t). \end{aligned}$$

□

This lemma and the backward induction theorem clearly give the same optimal stopping time as before: accept the t^{th} candidate if he or she is the best so far and $\sum_{i=t}^n \frac{1}{i} > 1$.

The other main tool used in Part 1 of this dissertation is the most basic result for infinite processes. It is used in Chapter 2 for the reason that the twins case will be reduced to a process that, although finite, does not have a fixed number of steps, and so backward induction is inappropriate. The following theorem is slightly adapted from that on page 55 of Chow, Robbins and Siegmund [22], since the random variables of interest are all positive.

This theorem applies only to the *monotone* case, and monotonicity is a strong property: it is the property that, after the first time in the process where the currently offered pay-off is at least the expected pay-off at the next step, the same is true at all future times. The conclusion of the theorem is that the first time when this is true is an optimal stopping time.

THEOREM 1.3. *Let $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \dots$ be a nested sequence of σ -algebras and let W_1, W_2, \dots be a sequence of random variables with each W_t being \mathcal{A}_t -measurable. Let $\mathcal{C}_{(\mathcal{A}_t)}$ be the class of stopping times relative to $(\mathcal{A}_t)_{t \in [n]}$. For $t \in \mathbb{N}$, let*

$$A_t = \left\{ \mathbb{E}(W_{t+1} | \mathcal{A}_t) \leq W_t \right\}.$$

and suppose that

$$A_1 \subset A_2 \subset \dots \quad \text{and} \quad \bigcup_{t=1}^{\infty} A_t = \Omega. \quad (1.2)$$

Let

$$\tau^* = \min \left\{ t : W_t \geq \mathbb{E}(W_{t+1} | \mathcal{A}_t) \right\}.$$

Suppose that $\mathbb{P}(\tau^* < \infty) = 1$ and $\mathbb{E}(W_{\tau^*})$ exists and that

$$\liminf_t \int_{\{\tau^* > t\}} W_t = 0.$$

Then

$$\mathbb{E}(W_{\tau^*}) \geq \mathbb{E}(W_{\tau}) \quad \text{for all } \tau \in \mathcal{C}_{(\mathcal{A}_t)}.$$

□

If equation (1.2) holds then the process $(W_t, \mathcal{A}_t)_{t \in [n]}$ is said to be *monotone*.

CHAPTER 2

How to choose the best twins

2.1. Introduction

Sections 2.1 to 2.5 of this chapter are based on joint work with Michał Morayne and Grzegorz Kubicki [34], but with some significant reorganization of its presentation, particularly in Section 2.3. Sections 2.6 and 2.7 are my own work.

The poset initially considered in this chapter is the set of m pairs of identical twins. This can also be viewed as a version of the classical secretary problem where each element is seen exactly twice. The Hasse diagram of this poset is Figure 2.1.

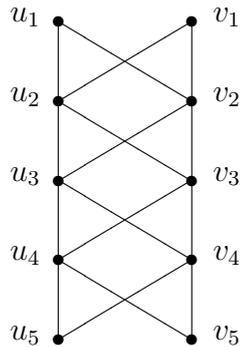


FIGURE 2.1. The poset $(U \cup V, \prec)$ for $m = 5$.

In Section 2.6, the poset considered is the natural extension to m sets of identical c -tuplets. The main results of this chapter are summarized in the following theorems.

THEOREM 2.1. *For $m \in \mathbb{N}$, let*

$$k_m = \min \left\{ k : \frac{2m}{k} + \sum_{j=k}^{m-1} \frac{1}{j} \leq 5 \right\}.$$

An optimal strategy for the secretary problem on m pairs of identical twins is to wait until candidates have been seen from at least k_m of the pairs and then to pick the next candidate who is the best so far and whose twin has already been seen. Asymptotically,

$$\lim_{m \rightarrow \infty} \frac{k_m}{m} = \frac{1}{x_0} \approx 0.4709,$$

where x_0 is the unique solution to $2x + \log x = 5$, and the probability of success tends to

$$\frac{1}{x_0} + \frac{4(x_0 - 1)^2}{3x_0} \left(\left(\frac{x_0}{x_0 - 1} \right)^{\frac{1}{2}} - 1 \right) \approx 0.7680.$$

THEOREM 2.2. For $c, m \in \mathbb{N}$ with $c \geq 2$, let

$$k_m^{(c)} = \min \left\{ k : \sum_{j=1}^{m-k} \left[\frac{\binom{m-j-1}{k-1}}{\binom{m-1}{k-1}} \prod_{i=2}^j \left(1 - \frac{1}{\binom{ci}{c}} \right) \right] \leq 1 \right\}.$$

An optimal strategy for the secretary problem on m sets of identical c -tuples is to wait until candidates have been seen from at least $k_m^{(c)}$ of the c -tuples and then to pick the next candidate who is the best so far and all of whose c -tuples have already been seen. For all m ,

$$\left(\frac{1}{2} - \frac{c+1}{2(2^{c+1}(c-1) - c - 1)} \right) m < k_m^{(c)} \leq \left\lceil \frac{m}{2} \right\rceil,$$

and the probability of success is at least

$$1 - \frac{c+1}{2^c(c-1)}.$$

At first, it might seem surprising that the secretary problem is easier in the twins case than on the total order, since there are fewer comparisons that can be made and so apparently less information. However, this is reconciled by the fact that if we attempt to compare two candidates then there are three possible outcomes rather than two, which gives us more information. Similarly, viewing the twins case as having two chances with each candidate makes it intuitive that it should be easier, which is correct.

Most of the notation used is as described in Section 1.4, and in Section 2.2 I shall describe the notation used specifically for the twins case. In Section 2.3, I shall use the monotone case theorem (Theorem 1.3) to find an optimal strategy for this problem and in Section 2.4 I shall give a formula for its probability of success. In Section 2.5, I shall describe the asymptotic behaviour of the threshold in the definition of the optimal strategy given and of the probability of success. I shall adapt the earlier method in Section 2.6 to find an optimal strategy and its probability of success for the natural generalization to c -tuples, and give some bounds that illustrate their behaviour. Finally, in Section 2.7, I

shall state what work there is still to be done in this case and suggest some other posets that one might wish to consider.

2.2. Notation and basic definitions

In this section, I shall define the poset consisting of m pairs of identical twins explicitly, as I shall refer to it in this way in the rest of the chapter.

Consider a set consisting of two chains $U = \{u_1, \dots, u_m\}$ and $V = \{v_1, \dots, v_m\}$ with u_1 and v_1 being maximal elements. For every i with $1 \leq i \leq m$, the elements u_i and v_i are incomparable. An element with a smaller subscript is higher than an element with a larger subscript, that is,

$$u_i \prec v_j \text{ if } i > j \quad \text{and} \quad v_j \prec u_i \text{ if } i < j.$$

The elements u_i and v_i are referred to as the twins at level i . The Hasse diagram of this partial order $(U \cup V, \prec)$ for $m = 5$ is given in Figure 2.1.

Suppose that the elements of $U \cup V$ are observed one by one in the order given by a random permutation π , with all $(2m)!$ permutations of $U \cup V$ equally likely. For every time t with $1 \leq t \leq 2m$, we observe the partial order induced by the set $\{\pi(1), \pi(2), \dots, \pi(t)\}$. As an example, suppose that we have a permutation

$$\pi = (u_3, v_2, u_1, v_3, u_5, v_1, u_4, v_5, v_4, u_2).$$

The induced orders observed for π up to $t = 6$ are given in Figure 2.2.

The goal is to choose the presently observed element to maximize the probability that this element is either u_1 or v_1 , that is, one of the two maximal elements of the poset $(U \cup V, \prec)$. In other words, we are looking for the stopping time τ such that $\mathbb{P}(\pi(\tau) \in \{u_1, v_1\})$ is maximal. The function τ , the optimal stopping time, depends itself on π . The decision at time t is based exclusively on the induced partial orders observed until time t and the order of appearance of their elements. For example, if the stopping time tells us to stop on the third element of the permutation given in Figure 2.2, then using this function we have to stop at time $t = 3$ for all permutations of $U \cup V$ whose first three elements have strictly decreasing subscripts.

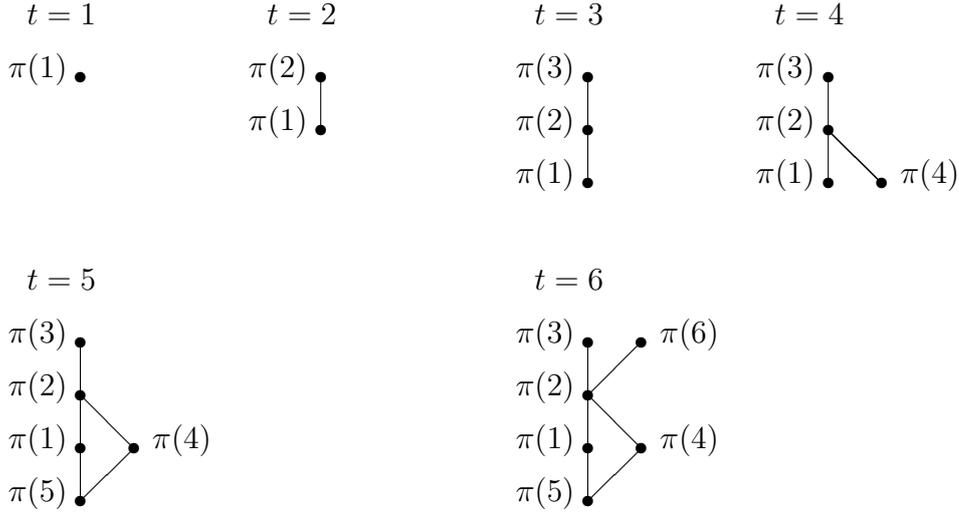


FIGURE 2.2. Induced orders when $\pi = (u_3, v_2, u_1, v_3, u_5, v_1, u_4, v_5, v_4, u_2)$ for $1 \leq t \leq 6$.

2.3. The optimal stopping time

In this section, I shall use the monotone case theorem to find an optimal stopping time for the process.

Define recursively the following stopping times:

$$\tau_i = \min \left\{ t > \tau_{i-1} : \pi(t) \in \max\{\pi(1), \pi(2), \dots, \pi(t)\} \text{ and} \right. \\ \left. \{\pi(1), \pi(2), \dots, \pi(t-1)\} \text{ contains the twin of } \pi(t) \right\},$$

with the assumptions that $\tau_0 = 0$ and if the set under the minimum is empty, then its minimum is $2m$. Recall that the twins come from m levels, with the elements u_i and v_i referred to as the twins at level i . Let

$$k_m = \min \left\{ k : \frac{2m}{k} + \sum_{j=k}^{m-1} \frac{1}{j} \leq 5 \right\}$$

and let

$$\tau = \min\{\tau_i : \text{the number of levels occupied by } \pi(1), \pi(2), \dots, \pi(\tau_i) \text{ is at least } k_m\},$$

with the assumption that if the set under minimum is empty, then $\tau = 2m$.

THEOREM 2.3. *The stopping time τ is optimal.*

Clearly, we should not consider stopping unless the currently observed element is maximal. Furthermore, if we are yet to see its twin then we should continue: if it is from level 1, then we shall have seen elements from at least as many levels by the time we see its twin and so shall be at least as confident that it is from level 1; if it is not from level 1, then we have a chance of finding this out by observing an element from a higher level. Thus, in order to maximize $\mathbb{P}(\pi(\tau) \in \{u_1, v_1\})$, it is necessary to consider only such stopping times τ whose values coincide with some τ_i 's.

Define the random variables

$$L_t = \text{level of the element } \pi(t).$$

Recall from Section 1.4 that

$$Z_t = \mathbb{P}(\pi(t) \in \{u_1, v_1\} | \mathcal{F}_t) = \mathbb{P}(L_t = 1 | \mathcal{F}_t).$$

The main aim is to show that the process $(Z_{\tau_i}, \mathcal{F}_{\tau_i})_{i \in \mathbb{N}}$ is monotone (see equation (1.2)), so that Theorem 1.3 can be applied. First note that if $\tau_i = 2m$ then $\tau_{i+1} = 2m$ and $\mathbb{E}(Z_{\tau_{i+1}} | \mathcal{F}_{\tau_i}) = Z_{\tau_i}$, so it is only necessary to consider those i for which $\tau_i < 2m$. For the rest of this section it will therefore be assumed that τ_i is a maximal element whose twin has already been seen, and that there are still some unseen elements.

A result is needed about the probability of success of the simple stopping time τ_1 . When using τ_1 , we stop at the first opportunity when we see a second twin who is maximal at that moment. Denote this probability by $P(m)$. Interestingly, the next lemma says that even the most naive strategy is successful with probability greater than $\frac{2}{3}$, which is much better than the corresponding $\frac{1}{n}$ for the total order on n elements.

LEMMA 2.4. *For the stopping time τ_1 the probability of success is equal to*

$$P(m) = \frac{2m + 1}{3m}.$$

PROOF. Observe that

$$\mathbb{P}(L_{\tau_1} = 1 | L_1 = 1) = 1,$$

since $\pi(\tau_1)$ can only be maximal at time τ_1 if it is from a level at most L_1 . I shall prove by induction on l that

$$\mathbb{P}(L_{\tau_1} = 1 | L_1 = l) = \frac{2}{3}$$

for all $l \geq 2$, from which the result follows, since then

$$\begin{aligned} P(m) &= \mathbb{P}(L_{\tau_1} = 1) = \sum_{l=1}^m \mathbb{P}(L_{\tau_1} = 1 | L_1 = l) \cdot \mathbb{P}(L_1 = l) \\ &= \frac{1}{m} \left(1 + \frac{2}{3}(m-1) \right) = \frac{2m+1}{3m}. \end{aligned}$$

If $l = 2$, then the claim is true, since the second twin from level 2 and the two twins from level 1 are identically distributed, and we win if and only if the first of these three to appear is from level 1. Now suppose that $l > 2$ and consider the level of the next element to appear from a level at most L_1 , at time τ' , say. If $1 < L_{\tau'} < l$, then the situation is the same as if $\pi(\tau')$ were the first element chosen, and by the induction hypothesis

$$\mathbb{P}(L_{\tau_1} = 1 | \{L_1 = l\} \cap \{1 < L_{\tau'} < l\}) = \frac{2}{3}.$$

Otherwise, we win if and only if one of the two twins from level 1 appears before the second twin from level l . Since these three are identically distributed, it is the case that

$$\mathbb{P}(L_{\tau_1} = 1 | \{L_1 = l\} \cap \{1 < L_{\tau'} < l\}^c) = \frac{2}{3}.$$

Thus, in either case, the probability is $\frac{2}{3}$, and the claim is proved. \square

I shall use the following combinatorial identity, which arises from consideration of the classical secretary problem, to simplify the formulae in the definition of the optimal stopping time and, later, its probability of success.

LEMMA 2.5. *For all $1 \leq k \leq m-1$,*

$$\sum_{j=1}^{m-k} \frac{1}{j} \cdot \frac{\binom{m-j-1}{k-1}}{\binom{m}{k}} = \frac{k}{m} \sum_{j=k}^{m-1} \frac{1}{j}.$$

PROOF. The proof proceeds by showing that both sides are equal to the probability of success in the classical secretary problem using the strategy: “reject the first k elements,

and accept the next maximal element observed.” Let W be the event that the best candidate is chosen.

LHS. For $0 \leq j \leq m - k$, let B_j be the event that the highest of the first k elements observed is at level $j + 1$. This means that there are $m - j - 1$ levels available for the remaining $k - 1$ elements. Let C_j be the event that the maximum element is the first element seen out of the highest j elements. Then B_j and C_j are independent, and

$$\begin{aligned} \mathbb{P}(W) &= \sum_{j=1}^{m-k} \mathbb{P}(B_j) \mathbb{P}(W|B_j) \\ &= \sum_{j=1}^{m-k} \mathbb{P}(B_j) \mathbb{P}(C_j) \\ &= \sum_{j=1}^{m-k} \frac{\binom{m-j-1}{k-1}}{\binom{m}{k}} \cdot \frac{1}{j}. \end{aligned}$$

RHS. For $k \leq j \leq m - 1$, let D_j be the event that the $(j + 1)^{\text{th}}$ element observed is the maximal element of the chain, and let E_j be the event that the highest of the first j elements observed is in the first k observed. Then D_j and E_j are independent, and

$$\begin{aligned} \mathbb{P}(W) &= \sum_{j=k}^{m-1} \mathbb{P}(D_j) \mathbb{P}(W|D_j) \\ &= \sum_{j=k}^{m-1} \mathbb{P}(D_j) \mathbb{P}(E_j) \\ &= \sum_{j=k}^{m-1} \frac{1}{m} \cdot \frac{k}{j}. \end{aligned}$$

Thus the lemma is proved. □

Let $N_k(i)$ denote the event that at time τ_i the number of levels occupied by the elements $\pi(1), \pi(2), \dots, \pi(\tau_i)$ is equal to k . Note that $N_k(i) \in \mathcal{F}_{\tau_i}$. The next two lemmas find the values that, when compared, determine whether or not a process is monotone.

LEMMA 2.6. *For all $\omega \in N_k(i)$,*

$$Z_{\tau_i}(\omega) = \frac{k}{m}.$$

PROOF. Elements from different levels are equally likely to appear in any position in the random order. Therefore, when k levels have been observed, the probability that one of them is the highest level is $\frac{k}{m}$, independent of the most recently observed element being from the highest level seen so far. \square

LEMMA 2.7. For all $\omega \in N_k(i)$,

$$\mathbb{E}(Z_{\tau_{i+1}} | \mathcal{F}_{\tau_i})(\omega) = \frac{2}{3} + \frac{k}{3m} \left(\sum_{j=k}^{m-1} \frac{1}{j} - 2 \right).$$

PROOF. Let

$$E = \int_A \mathbb{E}(Z_{\tau_{i+1}} | \mathcal{F}_{\tau_i}),$$

where $A \in \mathcal{F}_{\tau_i}$ is the atom containing ω . By the definition of conditional expectation, it is sufficient to show that

$$E = \left(\frac{2}{3} + \frac{k}{3m} \left(\sum_{j=k}^{m-1} \frac{1}{j} - 2 \right) \right) \mathbb{P}(A).$$

It is the case that

$$\begin{aligned} E &= \int_A Z_{\tau_{i+1}} = \int_A \mathbb{1}(L_{\tau_{i+1}} = 1) = \mathbb{P}(\{L_{\tau_{i+1}} = 1\} \cap A) \\ &= \sum_{j=1}^{m-k} \mathbb{P}(\{L_{\tau_{i+1}} = 1\} \cap \{L_{\tau_i} = j+1\} \cap A) \\ &= \sum_{j=1}^{m-k} \mathbb{P}(L_{\tau_{i+1}} = 1 | \{L_{\tau_i} = j+1\} \cap A) \cdot \mathbb{P}(L_{\tau_i} = j+1 | A) \cdot \mathbb{P}(A). \end{aligned}$$

As the probability of getting an element from the highest level at time τ_{i+1} does not depend on the pattern of the levels up to the time τ_i , and as A provides only a certain pattern how to put elements on any chosen k levels, it follows that, by Lemma 2.4,

$$\begin{aligned} E &= \sum_{j=1}^{m-k} \frac{2j+1}{3j} \cdot \frac{\binom{m-j-1}{k-1}}{\binom{m}{k}} \cdot \mathbb{P}(A) \\ &= \left(\sum_{j=1}^{m-k} \frac{2}{3} \cdot \frac{\binom{m-j-1}{k-1}}{\binom{m}{k}} + \sum_{j=1}^{m-k} \frac{1}{3j} \frac{\binom{m-j-1}{k-1}}{\binom{m}{k}} \right) \mathbb{P}(A) \end{aligned}$$

and, by Lemma 2.5,

$$\begin{aligned} E &= \left(\frac{2}{3} \cdot \frac{\binom{m-1}{k}}{\binom{m}{k}} + \frac{1}{3} \sum_{j=k}^{m-1} \frac{1}{j} \frac{\binom{m-1}{k-1}}{\binom{m}{k}} \right) \mathbb{P}(A) \\ &= \left(\frac{2}{3} \cdot \frac{m-k}{m} + \frac{1}{3} \cdot \frac{k}{m} \sum_{j=k}^{m-1} \frac{1}{j} \right) \mathbb{P}(A), \end{aligned}$$

which proves the lemma. \square

Let $A_i = \left\{ \mathbb{E}(Z_{\tau_{i+1}} | \mathcal{F}_{\tau_i}) \leq Z_{\tau_i} \right\}$ be the event that the pay-off offered at time τ_i is at least the expected value of the pay-off at time τ_{i+1} .

LEMMA 2.8. *For all $\omega \in N_k(i)$, it is the case that $\omega \in A_i$ if and only if*

$$\frac{2m}{k} + \sum_{j=k}^{m-1} \frac{1}{j} \leq 5.$$

PROOF. Let $\omega \in N_k(i)$. Then, by Lemmas 2.6 and 2.7, it is the case that $\omega \in A_i$ if and only if

$$\frac{2}{3} + \frac{k}{3m} \left(\sum_{j=k}^{m-1} \frac{1}{j} - 2 \right) = \mathbb{E}(Z_{\tau_{i+1}} | \mathcal{F}_{\tau_i})(\omega) \leq Z_{\tau_i}(\omega) = \frac{k}{m},$$

which gives the result. \square

The proof that τ is an optimal stopping time can now be completed.

PROOF OF THEOREM 2.3. It is sufficient to show that the process $(Z_{\tau_i}, \mathcal{F}_{\tau_i})_{i \in \mathbb{N}}$ is in the monotone case, that is, that $A_i \subset A_{i+1}$ for each i ; the monotone case theorem (Theorem 1.3) then says that τ is optimal.

Let $\omega \in A_i$ and let k be the number of levels observed at time τ_i , so that $\omega \in N_k(i)$. Then, by Lemma 2.8, the inequality

$$\frac{2m}{k} + \sum_{j=k}^{m-1} \frac{1}{j} \leq 5$$

holds. At time τ_{i+1} we must have seen more than k levels, and the left-hand side of the inequality decreases as k increases, and so $\omega \in A_{i+1}$. \square

2.4. The probability of success

In this section, I shall work out a formula for the probability of success when using an optimal stopping time.

THEOREM 2.9. *When using the optimal stopping time τ , the probability of success equals*

$$\mathbb{P}(L_{\pi(\tau)} = 1) = \frac{1}{3m} \left[2m + k_m - \left(k_m - \sum_{s=0}^{k_m-1} \prod_{r=1}^s \frac{2(m - k_m + r)}{2(m - k_m + r) + 1} \right) \left(3 - \sum_{j=k_m}^{m-1} \frac{1}{j} \right) \right]. \quad (2.1)$$

It is not surprising that evaluating the threshold produces a formula involving the harmonic series that cannot be expressed in closed form; a similar phenomenon occurs in the solution of the classical secretary problem.

For a permutation $\pi \in S(U \cup V)$, let

$$\tau_\pi = \min \{t : \pi(1), \pi(2), \dots, \pi(t) \text{ occupy exactly } k_m \text{ levels}\}.$$

Set

$$M_1 = \{\text{poset } \langle \pi(1), \pi(2), \dots, \pi(\tau_\pi) \rangle \text{ has exactly one maximal element}\}$$

and

$$M_2 = \{\text{poset } \langle \pi(1), \pi(2), \dots, \pi(\tau_\pi) \rangle \text{ has exactly two maximal elements}\}.$$

I shall begin by proving the following lemma.

LEMMA 2.10.

$$\mathbb{P}(M_1) = \frac{1}{k_m} \sum_{s=0}^{k_m-1} \prod_{r=1}^s \frac{2(m - k_m + r)}{2(m - k_m + r) + 1}. \quad (2.2)$$

PROOF. In the proof of this lemma, the levels will be referred to in the order in which they are observed $(1, 2, \dots, k_m)$, not their positions in the poset. Some of the elements will come from the same level as earlier elements, so one or two elements from each of these k_m levels will have been observed.

Let $\tau_\pi(j)$ be the smallest t such that $\pi(1), \pi(2), \dots, \pi(t)$ occupy exactly j levels (so, in particular, $\tau_\pi = \tau_\pi(k_m)$). For $i \leq j$, let $B_i(j)$ be the event that exactly one of

$\pi(1), \pi(2), \dots, \pi(\tau_\pi(j))$ occupies the i^{th} level, that is, at the first moment when j levels have been seen, exactly one element has been seen from the i^{th} level. Note that $\mathbb{P}(B_i(i)) = 1$ for all i , and also that, for $i \leq k_m$, the events are nested in the sequence $B_i(i) \supset B_i(i+1) \supset \dots \supset B_i(k_m)$. Thus

$$\begin{aligned} \mathbb{P}(B_i(k_m)) &= \mathbb{P}(B_i(i) \cap B_i(i+1) \cap \dots \cap B_i(k_m)) \\ &= \mathbb{P}(B_i(i)) \mathbb{P}(B_i(i+1)|B_i(i)) \dots \mathbb{P}(B_i(k_m)|B_i(k_m-1)). \end{aligned}$$

The probability $\mathbb{P}(B_i(j+1)|B_i(j))$ is easy to calculate. Indeed, suppose we are at time $\tau_\pi(j)$, having observed elements from j levels including exactly one element from the i^{th} level. There are $2(m-j)$ elements occupying levels not yet observed, and for $B_i(j+1)$ we require that one of these is observed before the second element from the i^{th} level. These $2(m-j)+1$ elements are identically distributed, so by symmetry

$$\mathbb{P}(B_i(j+1)|B_i(j)) = \frac{2(m-j)}{2(m-j)+1}.$$

Thus

$$\begin{aligned} \mathbb{P}(B_i(k_m)) &= \mathbb{P}(B_i(i)) \mathbb{P}(B_i(i+1)|B_i(i)) \dots \mathbb{P}(B_i(k_m)|B_i(k_m-1)) \\ &= \prod_{r=i}^{k_m-1} \frac{2(m-r)}{2(m-r)+1} \\ &= \prod_{r=1}^{k_m-i} \frac{2(m-k_m+r)}{2(m-k_m+r)+1}, \end{aligned}$$

following the convention that the empty product has value 1.

Let C_i be the event that the i^{th} level observed is the highest level observed by time τ_π , and let D_i be the event that both $B_i(k_m)$ and C_i hold, that is, that by time τ_π exactly one element has been observed from the i^{th} level and it is the highest observed so far. Note that the levels are equally likely to be observed in any order, so $\mathbb{P}(C_i) = \frac{1}{k_m}$, and

that $B_i(k_m)$ and C_i are independent, so

$$\begin{aligned}\mathbb{P}(D_i) &= \mathbb{P}(B_i(k_m) \cap C_i) \\ &= \mathbb{P}(B_i(k_m))\mathbb{P}(C_i) \\ &= \frac{1}{k_m} \prod_{r=1}^{k_m-i} \frac{2(m-k_m+r)}{2(m-k_m+r)+1}.\end{aligned}$$

Note finally that M_1 is the disjoint union of the D_i s, so

$$\begin{aligned}\mathbb{P}(M_1) &= \mathbb{P}(D_1) + \mathbb{P}(D_2) + \dots + \mathbb{P}(D_{k_m}) \\ &= \sum_{i=1}^{k_m} \frac{1}{k_m} \prod_{r=1}^{k_m-i} \frac{2(m-k_m+r)}{2(m-k_m+r)+1} \\ &= \frac{1}{k_m} \sum_{s=0}^{k_m-1} \prod_{r=1}^s \frac{2(m-k_m+r)}{2(m-k_m+r)+1},\end{aligned}$$

and the lemma is proved. \square

The proof of the formula for the probability of success can now be completed.

PROOF OF THEOREM 2.9. For each integer j with $2 \leq j \leq m - k_m + 2$, define M_1^j to be the subevent of M_1 where a second best element of $\langle \pi(1), \pi(2), \dots, \pi(\tau_\pi) \rangle$ is on level j . These subevents form a partition of M_1 and

$$\mathbb{P}(M_1^j) = \mathbb{P}(M_1^j | M_1) \mathbb{P}(M_1) = \frac{\binom{m-j}{k_m-2} (j-1)}{\binom{m}{k_m}} \mathbb{P}(M_1).$$

Similarly, for each integer j with $1 \leq j \leq m - k_m + 1$, define M_2^j to be the subevent of M_2 where the maximal elements of $\langle \pi(1), \pi(2), \dots, \pi(\tau_\pi) \rangle$ are on level j . These subevents form a partition of M_2 and

$$\mathbb{P}(M_2^j) = \mathbb{P}(M_2^j | M_2) \mathbb{P}(M_2) = \frac{\binom{m-j}{k_m-1}}{\binom{m}{k_m}} (1 - \mathbb{P}(M_1)).$$

Given M_2^j , the probability of success when using τ , which tells us to stop at the next τ_i , is equal to $\frac{2(j-1)+1}{3(j-1)}$, the probability of success when we use the simple stopping time τ_1 on the poset consisting of $j-1$ pairs of twins above the level j . Similarly, given M_1^j , the probability of success when using τ is equal to $\frac{2(j-1)+1}{3(j-1)}$, the probability of success when we use the simple stopping time τ_1 on the poset consisting of $j-1$ pairs of twins above

the level j of a second best element obtained so far, and with the maximal element being treated as the first random element observed in this reduced poset. The term with $j = 1$ in the second sum below disappears since there is no chance of stopping on a maximal element in that case. Therefore,

$$\begin{aligned} \mathbb{P}(L_\tau = 1) &= \sum_{j=2}^{m-k_m+2} \mathbb{P}(L_{\pi(\tau)} = 1 | M_1^j) \mathbb{P}(M_1^j) + \sum_{j=1}^{m-k_m+1} \mathbb{P}(L_{\pi(\tau)} = 1 | M_2^j) \mathbb{P}(M_2^j) \\ &= \sum_{j=2}^{m-k_m+2} \frac{2(j-1)+1}{3(j-1)} \cdot \frac{\binom{m-j}{k_m-2} (j-1)}{\binom{m}{k_m}} \mathbb{P}(M_1) \\ &\quad + \sum_{j=2}^{m-k_m+1} \frac{2(j-1)+1}{3(j-1)} \cdot \frac{\binom{m-j}{k_m-1}}{\binom{m}{k_m}} (1 - \mathbb{P}(M_1)), \end{aligned}$$

which, after changing the index of summation in both sums, becomes

$$\begin{aligned} \mathbb{P}(L_\tau = 1) &= \mathbb{P}(M_1) \sum_{j=1}^{m-k_m+1} \frac{2j+1}{3} \cdot \frac{\binom{m-j-1}{k_m-2}}{\binom{m}{k_m}} + (1 - \mathbb{P}(M_1)) \sum_{j=1}^{m-k_m} \frac{2j+1}{3j} \cdot \frac{\binom{m-j-1}{k_m-1}}{\binom{m}{k_m}} \\ &= \mathbb{P}(M_1) \frac{\sum_{j=1}^{m-k_m+1} \frac{2j}{3} \binom{m-j-1}{k_m-2} + \frac{1}{3} \sum_{j=1}^{m-k_m+1} \binom{m-j-1}{k_m-2}}{\binom{m}{k_m}} \\ &\quad + (1 - \mathbb{P}(M_1)) \frac{\sum_{j=1}^{m-k_m} \frac{2}{3} \binom{m-j-1}{k_m-1} + \sum_{j=1}^{m-k_m} \frac{1}{3j} \binom{m-j-1}{k_m-1}}{\binom{m}{k_m}}. \end{aligned}$$

Since, by standard combinatorial methods,

$$\begin{aligned} \sum_{j=1}^{m-k_m+1} j \binom{m-j-1}{k_m-2} &= \binom{m}{k_m}, \\ \sum_{j=1}^{m-k_m+1} \binom{m-j-1}{k_m-2} &= \binom{m-1}{k_m-1} \quad \text{and} \\ \sum_{j=1}^{m-k_m} \binom{m-j-1}{k_m-2} &= \binom{m-1}{k_m}, \end{aligned}$$

and, by Lemma 2.5,

$$\sum_{j=1}^{m-k_m} \frac{1}{j} \binom{m-j-1}{k_m-1} = \binom{m-1}{k_m-1} \sum_{j=k_m}^{m-1} \frac{1}{j},$$

it is the case that

$$\begin{aligned}
\mathbb{P}(L_\tau = 1) &= \mathbb{P}(M_1) \left(\frac{2}{3} + \frac{1}{3} \cdot \frac{\binom{m-1}{k_m-1}}{\binom{m-1}{k_m}} \right) + (1 - \mathbb{P}(M_1)) \frac{\frac{2}{3} \binom{m-1}{k_m} + \frac{1}{3} \binom{m-1}{k_m-1} \sum_{j=k_m}^{m-1} \frac{1}{j}}{\binom{m-1}{k_m}} \\
&= \mathbb{P}(M_1) \left(\frac{2}{3} + \frac{k_m}{3m} \right) + (1 - \mathbb{P}(M_1)) \left(\frac{2(m - k_m)}{3m} + \frac{k_m}{3m} \sum_{j=k_m}^{m-1} \frac{1}{j} \right) \\
&= \frac{2m + k_m}{3m} - (1 - \mathbb{P}(M_1)) \left(\frac{k_m}{m} - \frac{k_m}{3m} \sum_{j=k_m}^{m-1} \frac{1}{j} \right),
\end{aligned}$$

and substituting in the formula for $\mathbb{P}(M_1)$ from Lemma 2.10 gives the result. \square

2.5. Asymptotics

I shall begin this section with exact evaluations of, and a crude bound for, the threshold needed for the optimal stopping time τ and the probability of success for $m = 7$, and then determine the asymptotic behaviour of those quantities as $m \rightarrow \infty$.

EXAMPLE. For $m = 7$, the threshold $k_m = k_7 = 4$, since for $k = 4$ it is the case that

$$\frac{2 \cdot 7}{4} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} = \frac{247}{60} \leq 5$$

but for $k = 3$ the reverse inequality holds, namely,

$$\frac{2 \cdot 7}{3} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} = \frac{337}{60} > 5.$$

The probability of success is equal to

$$\begin{aligned}
\mathbb{P}(L_\tau = 1) &= \frac{1}{21} \left[18 - \left(4 - \left(1 + \frac{8}{9} + \frac{80}{99} + \frac{960}{1287} \right) \right) \left(3 - \frac{37}{60} \right) \right] \\
&= \frac{3001}{3780} \approx 0.7939.
\end{aligned}$$

OBSERVATION. By the definition of k_m , one can easily observe that for every m it must be the case that

$$k_m \leq \left\lceil \frac{m}{2} \right\rceil.$$

Indeed, if $k \geq \frac{m}{2}$ then

$$\frac{2m}{k} + \sum_{j=k}^{m-1} \frac{1}{j} \leq \frac{2m}{k} + \frac{m-k}{k} \leq 5.$$

The following theorems establish the asymptotic behavior of the threshold k_m and the limit of the probability of success as $m \rightarrow \infty$. Since m changes here, the optimal stopping time for the partial order consisting of m twins will be denoted by $\tau^{(m)}$. Denote the unique solution of the equation $2x + \log x = 5$ by x_0 , so that $x_0 \approx 2.12347$.

THEOREM 2.11. *The threshold k_m satisfies*

$$\frac{m}{x_0} < k_m < \frac{m}{x_0} + 2$$

for all $m \in \mathbb{N}$. In particular,

$$\lim_{m \rightarrow \infty} \frac{k_m}{m} = \frac{1}{x_0} \approx 0.4709.$$

PROOF. By the definition of k_m from Theorem 2.9, it is the case that k_m satisfies the inequality

$$5 \geq \frac{2m}{k_m} + \sum_{j=k_m}^{m-1} \frac{1}{j} > \frac{2m}{k_m} + \log \left(\frac{m}{k_m} \right),$$

the latter inequality by a simple integration argument. Since $2x + \log x$ is an increasing function, it must be the case that

$$\frac{m}{k_m} < x_0.$$

On the other hand, $k_m - 1$ does not satisfy the condition, so a similar argument yields

$$5 < \frac{2m}{k_m - 1} + \sum_{j=k_m-1}^{m-1} \frac{1}{j} < \frac{2m}{k_m - 1} + \log \left(\frac{m-1}{k_m - 2} \right) < \frac{2m}{k_m - 2} + \log \left(\frac{m}{k_m - 2} \right)$$

and hence

$$\frac{m}{k_m - 2} > x_0.$$

Thus

$$\frac{m}{x_0} < k_m < \frac{m}{x_0} + 2,$$

as required. □

I shall now turn to the asymptotic probability of success.

THEOREM 2.12. *The probability of success when using optimal stopping time $\tau^{(m)}$ satisfies*

$$\lim_{m \rightarrow \infty} \mathbb{P}(L_{\pi(\tau^{(m)})} = 1) = \frac{1}{x_0} + \frac{4(x_0 - 1)^2}{3x_0} \left(\left(\frac{x_0}{x_0 - 1} \right)^{\frac{1}{2}} - 1 \right) \approx 0.7680.$$

The proof uses the following lemma, which gives the asymptotic value of the expression in (2.2). As with $\tau^{(m)}$, the notation $\mathbb{P}(M_1^{(m)})$ is used to emphasize the dependence on m .

LEMMA 2.13. *The probability of $M_1^{(m)}$ occurring satisfies*

$$\lim_{m \rightarrow \infty} \mathbb{P}(M_1^{(m)}) = 2(x_0 - 1) \left(\left(1 + \frac{1}{x_0 - 1} \right)^{\frac{1}{2}} - 1 \right).$$

PROOF. Recall from Lemma 2.10 that $\mathbb{P}(M_1^{(m)})$ is given by

$$\mathbb{P}(M_1^{(m)}) = \frac{1}{k_m} \sum_{s=0}^{k_m-1} \prod_{r=1}^s \frac{2(m - k_m + r)}{2(m - k_m + r) + 1}.$$

Write $f(s)$ for the summand, that is,

$$f(s) = \prod_{r=1}^s \left(1 - \frac{1}{2(m - k_m + r) + 1} \right),$$

and consider $\log(f(s))$:

$$\begin{aligned} \log(f(s)) &= \sum_{r=1}^s \log \left(1 - \frac{1}{2(m - k_m + r) + 1} \right) \\ &= - \sum_{r=1}^s \left(\frac{1}{2(m - k_m + r) + 1} + \frac{1}{2} \left(\frac{1}{2(m - k_m + r) + 1} \right)^2 + \dots \right). \end{aligned}$$

For sufficiently large m , Theorem 2.11 shows that $\frac{k_m}{m} < \frac{1}{2}$, and hence that $\frac{1}{2(m - k_m + r) + 1} < \frac{1}{m}$ for all r . Thus, since $s < m$, the expression $\log(f(s))$ is bounded by

$$- \sum_{r=1}^s \frac{1}{2(m - k_m + r) + 1} - \left(\frac{1}{2m} + \frac{1}{3m^2} + \dots \right) < \log(f(s)) < - \sum_{r=1}^s \frac{1}{2(m - k_m + r) + 1}$$

and hence by

$$- \sum_{r=1}^s \frac{1}{2(m - k_m + r) + 1} - \left(\frac{1}{m} + \frac{1}{2m^2} + \dots \right) < \log(f(s)) < - \sum_{r=1}^s \frac{1}{2(m - k_m + r) + 1}. \quad (2.3)$$

The sum can be bounded further by integrals. Note that $\frac{1}{2x}$ is a decreasing function and obtain

$$\int_{\frac{1}{m}}^{\frac{s+1}{m}} \frac{1}{2\left(1 - \frac{k_m}{m} + \frac{1}{2m} + x\right)} dx \leq \sum_{r=1}^s \frac{1}{2(m - k_m + r) + 1} \leq \int_0^{\frac{s}{m}} \frac{1}{2\left(1 - \frac{k_m}{m} + \frac{1}{2m} + x\right)} dx$$

and hence

$$\frac{1}{2} \log \left(1 + \frac{s}{m - k_m + \frac{3}{2}} \right) \leq \sum_{r=1}^s \frac{1}{2(m - k_m + r) + 1} \leq \frac{1}{2} \log \left(1 + \frac{s}{m - k_m + \frac{1}{2}} \right). \quad (2.4)$$

Exponentiating (2.3) and substituting in the bounds in (2.4) gives

$$\left(1 - \frac{1}{m} \right) \left(1 + \frac{s}{m - k_m + \frac{1}{2}} \right)^{-\frac{1}{2}} < f(s) < \left(1 + \frac{s}{m - k_m + \frac{3}{2}} \right)^{-\frac{1}{2}}.$$

Noting that $f(s)$ is a decreasing function, the same technique is used to bound $\mathbb{P}(M_1^{(m)}) = \frac{1}{k_m} \sum_{s=0}^{k_m-1} f(s)$, which gives

$$\int_0^1 \left(1 - \frac{1}{m} \right) \left(1 + \frac{x}{\frac{m}{k_m} - 1 + \frac{1}{2k_m}} \right)^{-\frac{1}{2}} dx < \mathbb{P}(M_1^{(m)}) < \int_{-\frac{1}{k_m}}^{1 - \frac{1}{k_m}} \left(1 + \frac{x}{\frac{m}{k_m} - 1 + \frac{3}{2k_m}} \right)^{-\frac{1}{2}} dx$$

and hence

$$\begin{aligned} 2 \left(1 - \frac{1}{m} \right) \left(\frac{m}{k_m} - 1 + \frac{1}{2k_m} \right) \left(\left(\frac{m + \frac{1}{2}}{m - k_m + \frac{1}{2}} \right)^{\frac{1}{2}} - 1 \right) \\ < \mathbb{P}(M_1^{(m)}) < \\ 2 \left(\frac{m}{k_m} - 1 + \frac{3}{2k_m} \right) \left(\left(\frac{m + \frac{1}{2}}{m - k_m + \frac{3}{2}} \right)^{\frac{1}{2}} - \left(1 - \frac{1}{m - k_m + \frac{3}{2}} \right)^{\frac{1}{2}} \right). \end{aligned}$$

Recalling from Theorem 2.11 that $\lim_{m \rightarrow \infty} \frac{m}{k_m} = x_0$, take limits as $m \rightarrow \infty$ and notice that both sides of this inequality tend to the same limit. Hence,

$$\lim_{m \rightarrow \infty} \mathbb{P}(M_1^{(m)}) = 2(x_0 - 1) \left(\left(\frac{x_0}{x_0 - 1} \right)^{\frac{1}{2}} - 1 \right),$$

and the lemma is proved. \square

The proof of the asymptotic optimal probability of success can now be completed.

PROOF OF THEOREM 2.12. A similar integration argument to that in the proof of Lemma 2.13 gives

$$\lim_{m \rightarrow \infty} \sum_{j=k_m}^{m-1} \frac{1}{j} = \lim_{m \rightarrow \infty} \log \left(\frac{m}{k_m} \right) = \log x_0 = 5 - 2x_0.$$

Substituting these limiting values into the equation for success (2.1),

$$\mathbb{P}(L_\tau = 1) = \frac{1}{3m} \left[2m + k_m - \left(k_m - \sum_{s=0}^{k_m-1} \prod_{r=1}^s \frac{2(m - k_m + r)}{2(m - k_m + r) + 1} \right) \left(3 - \sum_{j=k_m}^{m-1} \frac{1}{j} \right) \right],$$

and writing $L = \lim_{m \rightarrow \infty} \mathbb{P}(L_{\tau(m)} = 1)$ gives

$$\begin{aligned} L &= \frac{1}{3} \left[2 + \frac{1}{x_0} - \left(\frac{1}{x_0} - \frac{1}{x_0} \cdot 2(x_0 - 1) \left[\left(\frac{x_0}{x_0 - 1} \right)^{\frac{1}{2}} - 1 \right] \right) (3 - (5 - 2x_0)) \right] \\ &= \frac{1}{x_0} + \frac{4(x_0 - 1)^2}{3x_0} \left(\left(\frac{x_0}{x_0 - 1} \right)^{\frac{1}{2}} - 1 \right), \end{aligned}$$

as required. □

2.6. How to choose the best c -tuplets

The posets most closely related to the one considered in this chapter are those that consist of m sets of c -tuplets, that is, with vertex set $U_1 \cup \dots \cup U_c$ where $U_i = \{u_i^1, \dots, u_i^m\}$, for fixed j the u_i^j s are incomparable and for all i_1, j_1, i_2, j_2 it is the case that

$$u_{i_1}^{j_1} \prec u_{i_2}^{j_2} \quad \text{if and only if} \quad j_1 > j_2.$$

In this section, I shall give an explicit optimal strategy and a formula for the probability of its success. Finding their asymptotic behaviour is an open problem, although I shall give some bounds that represent progress in that direction.

As with the twins case, the monotone case theorem can be used to find an explicit optimal strategy. Where it is unambiguous, I shall use notation from earlier in this chapter in this slightly different context without comment. Similarly to before, let

$$\begin{aligned} \tau_i^{(c)} &= \min \left\{ t > \tau_{i-1}^{(c)} : \pi(t) \in \max\{\pi(1), \pi(2), \dots, \pi(t)\} \text{ and} \right. \\ &\quad \left. \{\pi(1), \pi(2), \dots, \pi(t-1)\} \text{ contains all the } c\text{-tuplets of } \pi(t) \right\}, \end{aligned}$$

with the assumptions that $\tau_0^{(c)} = 0$ and if the set under the minimum is empty, then its minimum is cm .

To begin with, Lemma 2.4 can be generalized in the following way, where $\tau_1^{(c)}$ is the stopping time that stops at the first element that is maximal so far and all of whose c -tuplets have been seen.

LEMMA 2.14. *For the stopping time $\tau_1^{(c)}$ the probability of success is equal to*

$$P_c(m) = \prod_{i=2}^m \left(1 - \frac{1}{\binom{ci}{c}} \right).$$

PROOF. For $2 \leq i \leq m$, let N_i be the event that an element from level i is not chosen. Then

$$\begin{aligned} P_c(m) &= \mathbb{P}(N_2 \cap \dots \cap N_m) \\ &= \mathbb{P}(N_m) \mathbb{P}(N_{m-1} | N_m) \dots \mathbb{P}(N_2 | N_3 \cap \dots \cap N_m). \end{aligned}$$

Conditioned on $N_{i+1} \cap \dots \cap N_m$, an element from level i will be selected if and only if the c elements from this level all come before the $c(i-1)$ elements from the $i-1$ levels above. Hence,

$$\mathbb{P}(N_i | N_{i+1} \cap \dots \cap N_m) = 1 - \frac{1}{\binom{ci}{c}},$$

and the result follows. \square

This gives rise to an optimal stopping time. Similarly to before, let

$$k_m^{(c)} = \min \left\{ k : \sum_{j=1}^{m-k} \left[\frac{\binom{m-j-1}{k-1}}{\binom{m-1}{k-1}} \prod_{i=2}^j \left(1 - \frac{1}{\binom{ci}{c}} \right) \right] \leq 1 \right\}$$

and let

$$\tau^{(c)} = \min \left\{ \tau_i^{(c)} : \text{the number of levels occupied by } \pi(1), \pi(2), \dots, \pi(\tau_i) \text{ is at least } k_m^{(c)} \right\},$$

with the assumption that if the set under the minimum is empty then its minimum is cm .

THEOREM 2.15. *The stopping time $\tau^{(c)}$ is optimal for the secretary problem on m sets of identical c -tuplets.*

PROOF. The proof of this using the monotone case theorem is essentially the same as in Section 2.3. \square

Calculating the probability of success is slightly hard work algebraically, but follows broadly similar lines to the twins case. As then, let

$$L_t = \text{level of the element } \pi(t).$$

I shall use the multinomial coefficient notation

$$\binom{m}{r_1, \dots, r_c, m - r_1 - \dots - r_c}$$

to denote the number of ways of choosing c distinguishable sets of sizes r_1, \dots, r_c from a ground set of size m .

THEOREM 2.16. *The probability that the stopping time $\tau^{(c)}$ is successful is*

$$\begin{aligned} \mathbb{P}(L_{\tau^{(c)}} = 1) = & \sum_{i=1}^{c-1} \left[\left(\sum_{\substack{r_1 + \dots + r_c = k_m^{(c)} \\ r_1, r_i \geq 1}} \frac{\binom{m}{r_1, \dots, r_c, m - k_m^{(c)}} \binom{c}{1}^{r_1} \dots \binom{c}{c}^{r_c}}{\binom{cm}{r_1 + 2r_2 + \dots + cr_c}} \cdot \frac{r_1}{r_1 + 2r_2 + \dots + cr_c} \cdot \frac{r_i}{k_m^{(c)}} \right) \right. \\ & \left. \left(\frac{k_m^{(c)}}{m} + \sum_{j=2}^{m - k_m^{(c)} + 1} \frac{\binom{m-j}{k_m^{(c)} - 1}}{\binom{m}{k_m^{(c)}}} \left(1 - \frac{1}{\binom{c-j-i}{c-i}} \right) \prod_{k=2}^{j-1} \left(1 - \frac{1}{\binom{ck}{c}} \right) \right) \right] \\ & + \left(\sum_{\substack{r_1 + \dots + r_c = k_m^{(c)} \\ r_1, r_c \geq 1}} \frac{\binom{m}{r_1, \dots, r_c, m - k_m^{(c)}} \binom{c}{1}^{r_1} \dots \binom{c}{c}^{r_c}}{\binom{cm}{r_1 + 2r_2 + \dots + cr_c}} \cdot \frac{r_1}{r_1 + 2r_2 + \dots + cr_c} \cdot \frac{r_c}{k_m^{(c)}} \right) \\ & \left(\sum_{j=2}^{m - k_m^{(c)} + 1} \frac{\binom{m-j}{k_m^{(c)} - 1}}{\binom{m}{k_m^{(c)}}} \prod_{k=2}^{j-1} \left(1 - \frac{1}{\binom{ck}{c}} \right) \right). \end{aligned}$$

As with the twins case, I shall need a lemma that gives the probability that, at the first moment when $k_m^{(c)}$ levels have been seen, exactly i elements have been observed from the highest level seen. Let M_i be this event.

LEMMA 2.17. *Let i be an integer with $1 \leq i \leq c$. Then*

$$\mathbb{P}(M_i) = \sum_{\substack{r_1 + \dots + r_c = k_m^{(c)} \\ r_1, r_i \geq 1}} \frac{\binom{m}{r_1, \dots, r_c, m - k_m^{(c)}} \binom{c}{1}^{r_1} \dots \binom{c}{c}^{r_c}}{\binom{cm}{r_1 + 2r_2 + \dots + cr_c}} \cdot \frac{r_1}{r_1 + 2r_2 + \dots + cr_c} \cdot \frac{r_i}{k_m^{(c)}}.$$

PROOF. For $r_1 + \dots + r_c = k_m^{(c)}$, let A_{r_1, \dots, r_c} be the event that when $r_1 + 2r_2 + \dots + cr_c$ elements have been seen, the number of levels from which exactly j elements have been seen is r_j , for all j . Out of the $\binom{cm}{r_1 + 2r_2 + \dots + cr_c}$ possible choices for the first $r_1 + 2r_2 + \dots + cr_c$ elements, the number of choices such that A_{r_1, \dots, r_c} holds is as follows. First, there are $\binom{m}{r_1, \dots, r_c, m - k_m^{(c)}}$ choices for the levels with each number of elements seen. Then, for each j and for each of the r_j levels with j elements, there are $\binom{c}{j}$ choices for which of its elements have been seen. Therefore,

$$\mathbb{P}(A_{r_1, \dots, r_c}) = \frac{\binom{m}{r_1, \dots, r_c, m - k_m^{(c)}} \binom{c}{1}^{r_1} \dots \binom{c}{c}^{r_c}}{\binom{cm}{r_1 + 2r_2 + \dots + cr_c}}.$$

Now let B_{r_1, \dots, r_c} be the event that the number of levels from which exactly j elements have been seen is r_j , for all j , at the first moment when $k_m^{(c)}$ levels have been observed. Then B_{r_1, \dots, r_c} is only a non-empty event if $r_1 \geq 1$ and in that case B_{r_1, \dots, r_c} occurs if and only if A_{r_1, \dots, r_c} occurs and the last element observed is one of the singletons. By symmetry, the singletons are equally likely to be in any position in the first $r_1 + 2r_2 + \dots + cr_c$ elements, and therefore

$$\mathbb{P}(B_{r_1, \dots, r_c} | A_{r_1, \dots, r_c}) = \frac{r_1}{r_1 + 2r_2 + \dots + cr_c}.$$

Finally, let C_{r_1, \dots, r_c} be the event that B_{r_1, \dots, r_c} holds and that exactly i elements have been seen from the highest level. The highest level is equally likely to be any of the $r_1 + \dots + r_c = k_m^{(c)}$ levels chosen, and so

$$\mathbb{P}(C_{r_1, \dots, r_c} | B_{r_1, \dots, r_c}) = \frac{r_i}{k_m^{(c)}}.$$

For M_i to occur, there must be at least one level from which only one element has been seen (the last one observed), and at least one level from which exactly i elements have been seen (the highest one). Subject to these restrictions, if M_i occurs then C_{r_1, \dots, r_c} occurs for some choice of r_1, \dots, r_c with $r_1 + \dots + r_c = k_m^{(c)}$, and the C_{r_1, \dots, r_c} are disjoint

and all lie within M_i . Therefore,

$$\begin{aligned} \mathbb{P}(M_i) &= \sum_{\substack{r_1+\dots+r_c=k_m^{(c)} \\ r_1, r_i \geq 1}} \mathbb{P}(C_{r_1, \dots, r_c}) \\ &= \sum_{\substack{r_1+\dots+r_c=k_m^{(c)} \\ r_1, r_i \geq 1}} \frac{\binom{m}{r_1, \dots, r_c, m-k_m^{(c)}} \binom{c}{1}^{r_1} \cdots \binom{c}{c}^{r_c}}{\binom{cm}{r_1+2r_2+\dots+cr_c}} \cdot \frac{r_1}{r_1+2r_2+\dots+cr_c} \cdot \frac{r_i}{k_m^{(c)}}, \end{aligned}$$

as required. \square

Let M_i^j be the event that at the first moment when $k_m^{(c)}$ levels have been seen, there have been i elements from the highest level seen and that level is level j . Lemma 2.14 can easily be modified to give the following lemma, which gives the probability of success conditioned on M_i^j for all possible i and j .

LEMMA 2.18. (1) For i with $1 \leq i \leq c-1$,

$$\mathbb{P}(L_{\tau^{(c)}} = 1 | M_i^1) = 1.$$

(This is the case where $1 \leq i \leq c-1$ and $j = 1$.)

(2) For i, j with $1 \leq i \leq c-1$ and $2 \leq j \leq m - k_m^{(c)} + 1$,

$$\mathbb{P}(L_{\tau^{(c)}} = 1 | M_i^j) = \left(1 - \frac{1}{\binom{c-j-i}{c-i}}\right) \prod_{k=2}^{j-1} \left(1 - \frac{1}{\binom{ck}{c}}\right).$$

(This is the case where $1 \leq i \leq c-1$ and $2 \leq j \leq m - k_m^{(c)} + 1$.)

(3)

$$\mathbb{P}(L_{\tau^{(c)}} = 1 | M_c^1) = 0.$$

(This is the case where $i = c$ and $j = 1$.)

(4) For j with $2 \leq j \leq m - k_m^{(c)} + 1$,

$$\mathbb{P}(L_{\tau^{(c)}} = 1 | M_c^j) = \prod_{k=2}^{j-1} \left(1 - \frac{1}{\binom{ck}{c}}\right).$$

(This is the case where $i = c$ and $2 \leq j \leq m - k_m^{(c)} + 1$.)

\square

The proof of the formula for the probability of success can now be completed.

PROOF OF THEOREM 2.16. By the same argument as in the proof of Theorem 2.9, it is the case that

$$\begin{aligned} \mathbb{P}(M_i^j) &= \frac{\binom{m-j}{k_m^{(c)}-1}}{\binom{m}{k_m^{(c)}}} \mathbb{P}(M_i) \\ &= \frac{\binom{m-j}{k_m^{(c)}-1}}{\binom{m}{k_m^{(c)}}} \sum_{\substack{r_1+\dots+r_c=k_m^{(c)} \\ r_1, r_i \geq 1}} \frac{\binom{m}{r_1, \dots, r_c, m-k_m^{(c)}} \binom{c}{1}^{r_1} \dots \binom{c}{c}^{r_c}}{\binom{cm}{r_1+2r_2+\dots+cr_c}} \cdot \frac{r_1}{r_1+2r_2+\dots+cr_c} \cdot \frac{r_i}{k_m^{(c)}}. \end{aligned}$$

The events $\{M_i^j : 1 \leq i \leq c, 1 \leq j \leq m - k_m^{(c)} + 1\}$ partition the space and thus

$$\mathbb{P}(L_{\tau^{(c)}} = 1) = \sum_{i=1}^c \sum_{j=1}^{m-k_m^{(c)}+1} \mathbb{P}(L_{\tau^{(c)}} = 1 | M_i^j) \mathbb{P}(M_i^j),$$

and the proof is completed using Lemma 2.18. \square

Recall that

$$k_m^{(c)} = \min \left\{ k : \sum_{j=1}^{m-k} \left[\frac{\binom{m-j-1}{k-1}}{\binom{m-1}{k-1}} \prod_{i=2}^j \left(1 - \frac{1}{\binom{ci}{c}} \right) \right] \leq 1 \right\}.$$

The following lemma will be helpful for finding bounds for $k_m^{(c)}$.

LEMMA 2.19. *For all integers $c \geq 2$, and for all positive integers k and m with $k \leq m$, it is the case that*

$$\left(1 - \frac{c+1}{2^c(c-1)} \right) \left(\frac{m}{k} - 1 \right) < \sum_{j=1}^{m-k} \left[\frac{\binom{m-j-1}{k-1}}{\binom{m-1}{k-1}} \prod_{i=2}^j \left(1 - \frac{1}{\binom{ci}{c}} \right) \right] \leq \frac{m}{k} - 1.$$

PROOF. For the upper bound, first observe that the value of the product is at most one. By standard combinatorial arguments,

$$\sum_{j=1}^{m-k} \binom{m-j-1}{k-1} = \binom{m-1}{k}$$

and so

$$\sum_{j=1}^{m-k} \frac{\binom{m-j-1}{k-1}}{\binom{m-1}{k-1}} = \frac{m-k}{k} = \frac{m}{k} - 1,$$

which gives the upper bound.

For the lower bound, first observe that

$$\binom{ci}{c} = \frac{ci}{c} \cdot \frac{ci-1}{c-1} \cdot \dots \cdot \frac{ci-c+1}{1} \geq i^c.$$

Therefore,

$$\begin{aligned} \prod_{i=2}^j \left(1 - \frac{1}{\binom{ci}{c}}\right) &> 1 - \sum_{i=2}^{\infty} i^{-c} \\ &> 1 - \frac{1}{2^c} - \int_{i=2}^{\infty} x^{-c} dx \\ &= 1 - \frac{c+1}{2^c(c-1)}. \end{aligned} \tag{2.5}$$

Since this bound is independent of j , it can be treated as a constant factor and the sum evaluated as before, and the lower bound follows. \square

The reason that in (2.5) I bounded the sum only from $i = 3$, rather than the whole sum, was just to make the upper and lower bounds exponentially close. If I had not done so, then the first factor in the lower bound in Lemma 2.19 would have been $1 - \frac{1}{c-1}$, and the first factor in the lower bound in the next theorem would have been $\frac{1}{2} - \frac{1}{2(2c-3)}$.

THEOREM 2.20. *For all integers $c \geq 2$ and $m \geq 1$, it is the case that*

$$\left(\frac{1}{2} - \frac{c+1}{2(2^{c+1}(c-1) - c - 1)}\right) m < k_m^{(c)} \leq \left\lceil \frac{m}{2} \right\rceil.$$

In particular, for all $m \geq 1$,

$$\lim_{c \rightarrow \infty} k_m^{(c)} = \left\lceil \frac{m}{2} \right\rceil.$$

PROOF. If

$$k \leq \left(\frac{1}{2} - \frac{c+1}{2(2^{c+1}(c-1) - c - 1)}\right) m = \frac{1 - \frac{c+1}{2^c(c-1)}}{2 - \frac{c+1}{2^c(c-1)}} m,$$

then, by Lemma 2.19,

$$\sum_{j=1}^{m-k} \left[\frac{\binom{m-j-1}{k-1}}{\binom{m-1}{k-1}} \prod_{i=2}^j \left(1 - \frac{1}{\binom{ci}{c}}\right) \right] > \left(1 - \frac{c+1}{2^c(c-1)}\right) \left(\frac{m}{k} - 1\right) \geq 1,$$

and the lower bound follows.

If

$$k \geq \frac{m}{2},$$

then, by Lemma 2.19,

$$\sum_{j=1}^{m-k} \left[\frac{\binom{m-j-1}{k-1}}{\binom{m-1}{k-1}} \prod_{i=2}^j \left(1 - \frac{1}{\binom{ci}{c}} \right) \right] \leq \frac{m}{k} - 1 \leq 1,$$

and the upper bound follows. \square

A lower bound for the probability of success comes from observing that the proof of Lemma 2.19 bounds the probability of success of the simple stopping time τ_1 given in Lemma 2.14 by

$$P_c(m) \geq 1 - \frac{c+1}{2^c(c-1)},$$

and an optimal stopping time must do at least as well. This argument is equivalent to bounding the products in the formula for the probability of success in Theorem 2.16.

c	$\frac{1}{2} - \frac{c+1}{2(2^{c+1}(c-1)-c-1)}$	$k_{100}^{(c)}$	$k_{1000}^{(c)}$	$\lim_{m \rightarrow \infty} \frac{k_m^{(c)}}{m}$	$1 - \frac{c+1}{2^c(c-1)}$	$w_{100}^{(c)}$	$\lim_{m \rightarrow \infty} w_m^{(c)}$
1	-	38	369	0.3679	-	0.3708	0.3679
2	0.2000	48	471	0.4709	0.2500	0.7697	0.7680
3	0.4286	50	493	?	0.7500	0.9354	?
4	0.4725	50	499	?	0.8958	?	?
5	0.4880	50	500	?	0.9531	?	?
6	0.4945	50	500	?	0.9781	?	?
7	0.4974	50	500	?	0.9896	?	?
8	0.4987	50	500	?	0.9950	?	?
9	0.4994	50	500	?	0.9976	?	?
10	0.4997	50	500	?	0.9988	?	?

FIGURE 2.3. A table showing bounds for and exact and asymptotic values of $k_m^{(c)}$ and $w_m^{(c)} = \mathbb{P}(L_{\tau^{(c)}} = 1)$ for small values of c .

For convenience and to emphasise the dependence on m , for given c and m let

$$w_m^{(c)} = \mathbb{P}(L_{\tau^{(c)}} = 1)$$

be the probability of success of $\tau^{(c)}$ on m sets of c -tuplets. Figure 2.3 summarizes the bounds from this section, and gives exact values of $k_{100}^{(c)}$, $k_{1000}^{(c)}$ and $w_{100}^{(c)}$, calculated using

Maple™ 13. Unfortunately, the formula in Theorem 2.16 is too complicated to be calculated efficiently for anything but small c ; similarly, these values of m are near the limit of what I could calculate.

2.7. Open problems

I shall first summarize the main goals for the c -tuplets case.

- (1) Simplify the formula for the probability of success in Theorem 2.16.
- (2) For fixed c , prove that $\lim_{m \rightarrow \infty} \frac{k_m^{(c)}}{m}$ exists.
- (3) If $\lim_{m \rightarrow \infty} \frac{k_m^{(c)}}{m}$ exists, find its value as the root of an equation or as a function of $\lim_{m \rightarrow \infty} \frac{k_m^{(c-1)}}{m}$.
- (4) For fixed c , prove that $\lim_{m \rightarrow \infty} w_m^{(c)}$ exists.
- (5) If $\lim_{m \rightarrow \infty} w_m^{(c)}$ exists, find its value as the root of an equation or as a function of the value for $\lim_{m \rightarrow \infty} w_m^{(c-1)}$.

I have already shown that for any fixed m ,

$$\lim_{c \rightarrow \infty} k_m^{(c)} = \left\lceil \frac{m}{2} \right\rceil$$

and

$$\lim_{c \rightarrow \infty} w_m^{(c)} = 1,$$

which gives the values of

$$\lim_{c \rightarrow \infty} \lim_{m \rightarrow \infty} \frac{k_m^{(c)}}{m}$$

and

$$\lim_{c \rightarrow \infty} \lim_{m \rightarrow \infty} w_m^{(c)}$$

if they exist.

In this chapter, I have assumed that the aim is to pick the best candidate. It is almost certainly a harder problem, as with the original secretary problem, to try to minimize the expected rank instead. Alternatively, the full information case would be interesting to study; here, m random samples would be taken from a known distribution and each replicated $c - 1$ times, and all cm of these values revealed to the selector one by one.

Inspired further by the results in Chapter 1, one could allow the interviewer to go back to an earlier candidate, who is still available with a certain probability, or assume that the number of sets of c -tuplets is unknown but comes from a known distribution.

A modification of the c -tuplets poset that might be worth thinking about is a ‘pyramid’ with i candidates at level i . It seems likely that this would lie between the classical secretary problem and the twins version in terms of difficulty. It is also natural to try to solve the problem of optimal choice, or any of these variations, on other common posets, for example

- (1) $Q_k = \{0, 1\}^k$, where $(x_1, \dots, x_k) \prec (y_1, \dots, y_k)$ if $(x_1, \dots, x_k) \neq (y_1, \dots, y_k)$ and $x_i \leq y_i$ for $1 \leq i \leq k$,
- (2) $[m]^2$, where $(x_1, x_2) \prec (y_1, y_2)$ if $(x_1, x_2) \neq (y_1, y_2)$ and $x_i \leq y_i$ for $i = 1, 2$, or even
- (3) $[m]^k$, where $(x_1, \dots, x_k) \prec (y_1, \dots, y_k)$ if $(x_1, \dots, x_k) \neq (y_1, \dots, y_k)$ and $x_i \leq y_i$ for $1 \leq i \leq k$, which includes both of the previous examples.

CHAPTER 3

The secretary problem on an unknown poset

3.1. Introduction

This chapter is based on joint work with Robert Morris [35], and is substantially the same as the paper.

In this chapter, I shall show that for the problem of optimal choice on any poset there is an algorithm that is successful with probability at least $\frac{1}{e}$, so, in this sense, the total order is the hardest possible partial order. In fact, this algorithm depends only on the size of the poset and its number of maximal elements, so it is universal for any family where these are given. It is therefore natural to ask which is the hardest partial order with a given number of maximal elements. The most obvious choice is the poset consisting of k disjoint chains. I shall give an asymptotically sharp lower bound on the probability of success in the problem of optimal choice on k disjoint chains, and show that it is at least as hard as on any poset with k maximal elements and of width k , that is, whose largest antichain has size k .

More precisely, the main aim is to prove the following two theorems.

THEOREM 3.1. *Let (P, \prec) be a poset with k maximal elements and of width k . Then there is an algorithm for the secretary problem on (P, \prec) depending only on $|P|$ and k that is successful with probability at least p_k , where*

$$p_k = \begin{cases} \frac{1}{e} & \text{if } k = 1, \\ k^{-1} \sqrt{\frac{1}{k}} & \text{if } k > 1, \end{cases} \quad (3.1)$$

and these are the best possible such bounds.

THEOREM 3.2. *Let (P, \prec) be a poset with k maximal elements. Then there is an algorithm for the secretary problem on (P, \prec) depending only on $|P|$ and k that is successful with probability at least $\frac{1}{e}$, and this is the best possible such bound.*

Robert Morris and I conjectured that both of these theorems can be improved, by removing the width condition in Theorem 3.1 and the dependence on k in Theorem 3.2; see Conjectures 3.20 and 3.21 in Section 3.4.

The poset consisting of k disjoint chains was considered by Kuchta and Morayne [56], but with a restriction on the order in which the elements are observed: those from the first chain all appear in a random order, then those from the second chain, and so on. This poset is also related to multicriteria extensions of the secretary problem. In the multicriteria version due to Gnedin [41], each element is ranked independently in $k > 1$ different criteria, and the selector wishes to select an element that is maximal in at least one of them. This is equivalent to the problem on k equally-sized disjoint chains with the elements appearing one at a time from each chain in the same cyclic order. Interestingly, the asymptotic value of the probability of success in Theorem 3.1, ${}^{k-1}\sqrt{\frac{1}{k}}$, is the same as in the multicriteria version.

This chapter is organized as follows. In Section 3.2, I shall describe a (randomized) algorithm for choosing an element of a given poset, and prove lower bounds for its probability of success on various families of posets. In Section 3.3, I shall show that these bounds are best possible, by proving that, for the poset that consists of k disjoint chains of length x (which lies in each of these families), there is no strategy that is successful with probability greater than $p_k + o(1)$ (as $x \rightarrow \infty$). Finally, I shall suggest some conjectures and open problems in Section 3.4.

3.2. Lower bounds

Throughout this section, p is a real number satisfying $0 < p < 1$. Recall that $\pi(t)$ is the t^{th} element of the poset P that we see, and that P_t is a poset with vertex set $[t]$ that is isomorphic to the poset seen at time t . I shall prove lower bounds for the probability of success of the following randomized algorithm on different families of posets.

ALGORITHM. Given a poset with n elements, of which k are maximal, let $X(p) \sim \text{Bin}(n, p)$. Reject the first $X(p)$ elements and accept the first subsequent element where the following condition holds: the poset induced by the elements seen so far (including the

currently observed element) has at most k maximal elements and the currently observed element is one of them.

This algorithm gives rise to the following randomized stopping time, $\tau_k(p)$.

Recall that $\Omega = S_n \times [0, 1]$ and that \mathbb{P} is the uniform probability measure on Ω . Let $X(p) : \Omega \rightarrow \{0, \dots, n\}$ be the random variable defined by

$$X(p)(\rho, \delta) = \min \left\{ x \geq 0 : \sum_{i=0}^x \binom{n}{i} p^i (1-p)^{n-i} \geq \delta \right\},$$

so that

$$\mathbb{P}(X(p) = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

and $X(p) = X(p)(\rho, \delta)$ is independent of ρ . Then $\tau_k(p)$ is defined by

$$\tau_k(p) = \begin{cases} \min \{t > X(p) : |\max(P_t)| \leq k \text{ and } t \in \max(P_t)\} & \text{if this exists,} \\ n & \text{otherwise.} \end{cases}$$

Given the definition of $\tau_k(p)$, it makes sense to consider another random variable, the set of $X(p)$ elements that we reject without considering. Denote this random variable by $S(p)$, where

$$S(p) = \{\pi(t) : t \leq X(p)\}.$$

The following simple property of $S(p)$ is important. It means that one can think of the algorithm as rejecting each element of P with probability p , independently of the others and of the random ordering π , which makes it easier to work out a formula for the probability that it is successful.

LEMMA 3.3. *The events $\{x \in S(p)\}_{x \in P}$ are independent and $\mathbb{P}(x \in S(p)) = p$ for all $x \in P$.*

PROOF. The random variables π and $X(p)$ can be generated with the required distributions in the following way. Put each element of P in $S(p)$ with probability p independently of all other elements. Let π consist of a uniformly random ordering of the elements of $S(p)$ followed by a uniformly random ordering of $P \setminus S(p)$. By symmetry, π is

a uniformly random ordering of P , and $X(p) = |S(p)|$ is a binomial random variable independent of π . The events $\{x \in S(p)\}_{x \in P}$ depend only on π and $X(p)$, and by construction the properties in the statement of the lemma hold. \square

The following identity will be useful for simplifying the formula for the probability of success.

LEMMA 3.4. *For all integers $k \geq 1$, the following holds:*

$$\sum_{s=0}^{\infty} \binom{k+s-1}{k-1} (1-p)^s = \frac{1}{p^k}.$$

PROOF. Suppose that we have a coin that comes up heads with probability p and tails with probability $1-p$, and that we toss it infinitely many times. Then, with probability 1, we shall see at least k heads, and the k^{th} head comes up in position $k+s$ for some $s \geq 0$. In this case, we know that $k-1$ of the first $k+s-1$ tosses are heads and the remaining s are tails, and so, summing over the probabilities that the k^{th} head comes in each position, we have

$$\sum_{s=0}^{\infty} \binom{k+s-1}{k-1} p^k (1-p)^s = 1.$$

\square

In order to prove Theorem 3.1, I shall first calculate a lower bound for the probability that $\tau_k(p)$ is successful on the poset consisting of k disjoint chains. Recall that p is a real number satisfying $0 < p < 1$ and that $\pi(\tau_k(p))$ is the element that the algorithm $\tau_k(p)$ selects.

THEOREM 3.5. *Let (P, \prec) be a poset consisting of k disjoint chains. Then*

$$\mathbb{P}\left(\pi(\tau_k(p)) \in \max(P)\right) > \begin{cases} p \log\left(\frac{1}{p}\right) & \text{if } k = 1, \\ \frac{k}{k-1} p (1-p^{k-1}) & \text{if } k > 1. \end{cases}$$

PROOF. First note that $\pi(\tau_k(p)) \in \max(P)$ in the exceptional case where $S(p) = P$ and $\pi(\tau_k(p)) = \pi(n) \in \max(P)$, an event with probability $\frac{k}{n} p^n$. This tends to 0 as $n \rightarrow \infty$, and the bounds in the theorem are obtained by considering only the cases where $X(p) < n$

and hence $\pi(\tau_k(p)) \notin S(p)$. However, later on, in the proof of Lemma 3.11, the fact that these bounds are for a slightly smaller event will be important.

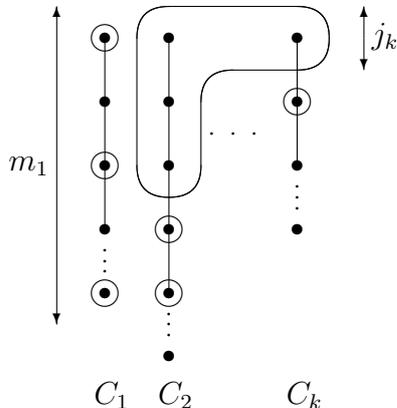


FIGURE 3.1. An example of k disjoint chains with the elements of $S(p)$ circled. This illustrates an instance of the event $A_{0,3,\dots,1}$. The region enclosed by the solid curve marks the $j_1 + \dots + j_k$ elements that might be selected.

Let the k chains be denoted by C_1, \dots, C_k and have lengths m_1, \dots, m_k . Let A_{j_1, \dots, j_k} be the event that for each i there are j_i elements from C_i not in $S(p)$ above the highest element from C_i in $S(p)$ (see Figure 3.1), that is,

$$A_{j_1, \dots, j_k} = \bigcap_{i=1}^k \left\{ \left| \left\{ x \in C_i \setminus S(p) : \nexists y \in C_i \cap S(p) \text{ such that } x \prec y \right\} \right| = j_i \right\}.$$

For $j_i < m_i$, this means that the top j_i elements are not in $S(p)$ but the $(j_i + 1)^{\text{th}}$ is. For $j_i = m_i$, this means that there are no elements from the i^{th} chain in $S(p)$. Note that if A_{j_1, \dots, j_k} occurs then $\pi(\tau_k(p))$ will be the first element observed from the $j_1 + \dots + j_k$ elements not in $S(p)$ that are at the tops of their chains.

The events $\{A_{j_1, \dots, j_k} : 0 \leq j_1 \leq m_1, \dots, 0 \leq j_k \leq m_k\}$ partition the whole space. Thus, writing $Q_k(p)$ for $\mathbb{P}(\pi(\tau_k(p)) \in \max(P))$,

$$\begin{aligned} Q_k(p) &= \sum_{0 \leq j_1 \leq m_1, \dots, 0 \leq j_k \leq m_k} \mathbb{P}(\pi(\tau_k(p)) \in \max(P) \mid A_{j_1, \dots, j_k}) \mathbb{P}(A_{j_1, \dots, j_k}) \\ &> \sum_{\substack{0 \leq j_1 \leq m_1, \dots, 0 \leq j_k \leq m_k \\ (j_1, \dots, j_k) \neq (0, \dots, 0)}} \frac{|\{i : j_i > 0\}|}{j_1 + \dots + j_k} (1-p)^{j_1 + \dots + j_k} p^{|\{i : j_i < m_i\}|}. \end{aligned} \quad (3.2)$$

Since $1 + (1 - p) + (1 - p)^2 + \dots = \frac{1}{p}$, this can be written as

$$\begin{aligned}
Q_k(p) &> \sum_{\substack{0 \leq j_1 \leq m_1, \dots, 0 \leq j_k \leq m_k \\ (j_1, \dots, j_k) \neq (0, \dots, 0)}} \frac{|\{i : j_i > 0\}|}{j_1 + \dots + j_k} (1 - p)^{j_1 + \dots + j_k} p^k (1 + (1 - p) + (1 - p)^2 + \dots)^{|\{i : j_i = m_i\}|} \\
&= \sum_{\substack{j_1, \dots, j_k \geq 0 \\ (j_1, \dots, j_k) \neq (0, \dots, 0)}} \frac{|\{i : j_i > 0\}|}{\min\{j_1, m_1\} + \dots + \min\{j_k, m_k\}} (1 - p)^{j_1 + \dots + j_k} p^k \\
&> \sum_{\substack{j_1, \dots, j_k \geq 0 \\ (j_1, \dots, j_k) \neq (0, \dots, 0)}} \frac{|\{i : j_i > 0\}|}{j_1 + \dots + j_k} (1 - p)^{j_1 + \dots + j_k} p^k.
\end{aligned}$$

Rewriting the last line as a sum over $r = |\{i : j_i > 0\}|$ and $s = j_1 + \dots + j_k$ gives

$$Q_k(p) > \sum_{r=1}^k \sum_{s=r}^{\infty} \left| \left\{ (j_1, \dots, j_k) : |\{i : j_i > 0\}| = r \text{ and } j_1 + \dots + j_k = s \right\} \right| \cdot \frac{r}{s} (1 - p)^s p^k.$$

The rest of the proof is a straightforward calculation. To calculate

$$\left| \left\{ (j_1, \dots, j_k) : |\{i : j_i > 0\}| = r \text{ and } j_1 + \dots + j_k = s \right\} \right|,$$

we note that there are $\binom{k}{r}$ ways of choosing the indices i with $j_i > 0$ and there are then $\binom{s-1}{r-1}$ ways for r non-zero numbers to add up to s . Thus

$$Q_k(p) > \sum_{r=1}^k \sum_{s=r}^{\infty} \binom{k}{r} \binom{s-1}{r-1} \frac{r}{s} (1 - p)^s p^k = kp^k \sum_{r=1}^k \sum_{s=r}^{\infty} \binom{k-1}{r-1} \binom{s-1}{r-1} \frac{1}{s} (1 - p)^s.$$

Reversing the order of summation,

$$Q_k(p) > kp^k \sum_{s=1}^{\infty} \frac{1}{s} (1 - p)^s \sum_{r=1}^{\min\{k, s\}} \binom{s-1}{r-1} \binom{k-1}{k-r}.$$

The second sum is easily evaluated (a result known as Vandermonde's identity) to give

$$Q_k(p) > kp^k \sum_{s=1}^{\infty} \frac{1}{s} (1 - p)^s \binom{k+s-2}{k-1}. \quad (3.3)$$

It remains to evaluate the sum in the above inequality. Write

$$V_k(p) = \sum_{s=1}^{\infty} \frac{1}{s} (1 - p)^s \binom{k+s-2}{k-1}.$$

THEOREM 3.7. *Let (P, \prec) be a poset with n elements. Suppose that (P, \prec) has k maximal elements and that none of its antichains has size greater than k . Then*

$$\mathbb{P}\left(\pi(\tau_k(p)) \in \max(P)\right) > \begin{cases} p \log\left(\frac{1}{p}\right) & \text{if } k = 1, \\ \frac{k}{k-1}p(1 - p^{k-1}) & \text{if } k > 1. \end{cases}$$

PROOF. Dilworth's theorem (Theorem 3.6) implies that P takes the form of k chains with some comparisons in between them. Clearly, the k elements of $\max(P)$ lie at the top of the k chains. The proof therefore proceeds in an almost identical manner to that of Theorem 3.5. The only difference is that the denominator in each term of (3.2) is now at most, rather than equal to, $j_1 + \dots + j_k$ (see Figure 3.2), so the expression in this line is still a lower bound. The calculations that make up the remainder of the proof of Theorem 3.5 therefore follow in the same way. \square

The values that maximize the function in Theorem 3.7 are

$$p_k = \begin{cases} \frac{1}{e} & \text{if } k = 1, \\ \sqrt[k-1]{\frac{1}{k}} & \text{if } k > 1. \end{cases} \quad (3.4)$$

This has the following consequence, which gives the lower bounds in Theorem 3.1.

COROLLARY 3.8. *Let (P, \prec) be a poset with n elements. Suppose that (P, \prec) has k maximal elements and that none of its antichains has size greater than k . Then*

$$\mathbb{P}\left(\pi(\tau_k(p_k)) \in \max(P)\right) > p_k.$$

\square

It is curious to note that the expected proportion of elements that we reject without considering is the same as the probability of success; it is not clear to me why this should be the case.

The next objective is to prove the following theorem, which, with the right choice of p , will give a lower bound of $\frac{1}{e}$ for all posets, as in Theorem 3.2.

THEOREM 3.9. *Let (P, \prec) be a poset with n elements. Suppose that (P, \prec) has k maximal elements. Then*

$$\mathbb{P}\left(\pi(\tau_k(p)) \in \max(P)\right) > kp^k \log\left(\frac{1}{p}\right).$$

The proof will use two simple lemmas. The first states that the linear order is the hardest of all posets with a unique maximal element.

LEMMA 3.10. *Let (P, \prec) be a poset with n elements. Suppose that (P, \prec) has exactly one maximal element. Then the probability that $\tau_1(p)$ is successful on (P, \prec) is at least the probability that it is successful on a linear extension of P , and hence*

$$\mathbb{P}\left(\pi(\tau_k(p)) \in \max(P)\right) > p \log\left(\frac{1}{p}\right).$$

PROOF. Begin by taking an arbitrary linear extension of \prec , that is, a partial order \prec' such that any two elements are comparable and such that $x \prec y \Rightarrow x \prec' y$. (It is clear that such a partial order exists.) Denote the unique element in $\max_{\prec}(P) = \max_{\prec'}(P)$ by x_{\max} .

Given this new poset, (P, \prec') , define random variables π' , $X'(p)$, $S'(p)$ and $\tau'_1(p)$ in the same way as π , $X(p)$, $S(p)$ and $\tau_1(p)$ were defined given (P, \prec) . Couple the random variables $(\pi, X(p), S(p), \tau_1(p))$ and $(\pi', X'(p), S'(p), \tau'_1(p))$ in the obvious way; set $\pi' = \pi$ and $X'(p) = X(p)$, and hence $S'(p) = S(p)$. This means that the elements appear in the same order in both instances, and the same set $S(p)$ is rejected in both cases. The induced posets observed in the process on (P, \prec') are linear extensions of those observed in the process on (P, \prec) . The proof proceeds by showing that if $\pi(\tau_1(p)) \neq x_{\max}$ then $\pi'(\tau'_1(p)) \neq x_{\max}$, that is, if $\tau_1(p)$ fails in the process on (P, \prec) then $\tau'_1(p)$ fails on (P, \prec') . From this, the result follows, since the probability of success is therefore at least as large on (P, \prec) as on (P, \prec') , and Theorem 3.5 applied to (P, \prec') gives the lower bound.

If we reach x_{\max} then it will be accepted, since it must be the unique maximal element in the poset induced by the elements observed so far. Thus $\pi(\tau_1(p)) \neq x_{\max}$ if either

- (i) $x_{\max} \in S(p)$ or
- (ii) after rejecting $S(p)$, we accept an element that appears earlier than x_{\max} .

In case (i), $\tau'_1(p)$ must fail on (P, \prec') for the same reason, since $S'(p) = S(p)$. In case (ii), such an element must be the unique maximal element of the poset induced by what we have seen so far, and this is still the case in any linear extension. Therefore, with $\tau'_1(p)$, if this element is observed then it must be accepted, and so we still accept an element that appears earlier than x_{\max} . It follows that in either case $\pi'(\tau'_1(p)) \neq x_{\max}$. \square

The next lemma gives a lower bound for the probability of success restricting our attention to the case when all but one of the maximal elements of our poset are in $S(p)$. This turns out to be enough to prove Theorem 3.9.

LEMMA 3.11. *Let (P, \prec) be a poset with n elements. Suppose that (P, \prec) has k maximal elements. Then*

$$\mathbb{P}\left(\pi(\tau_k(p)) \in \max(P) \mid |\max(P) \cap S(p)| = k - 1\right) > \frac{p}{1-p} \log\left(\frac{1}{p}\right).$$

PROOF. First observe that we may assume that $k = 1$, for the following reason. The condition that $|\max(P) \cap S(p)| = k - 1$ means that the $k - 1$ maximal elements in $\max(P) \cap S(p)$ will be maximal for the remainder of the process, so when using $\tau_k(p)$ these and all elements dominated by at least one of these may be ignored, and a unique maximal element from the remaining elements will be selected. Those elements form a poset (P', \prec') with a unique maximal element x_{\max} , which is not in $S(p)$. Since all elements are in $S(p)$ with probability p independently of the others, the situation is the same as if we were working with (P', \prec') and conditioning on $x_{\max} \notin S(p)$. Looking for one of at most k maximal elements in P using $\tau_k(p)$ is the same as looking for a unique maximal element in P' using $\tau_1(p)$.

Assume from now on that $k = 1$; Lemma 3.10 will be used to prove the result in this case. Lemma 3.10 used the bound from Theorem 3.5, and recall from the proof of that theorem that the lower bound for $\mathbb{P}(\pi(\tau_1(p)) = x_{\max})$ is in fact a lower bound for $\mathbb{P}(\{S(p) \neq P\} \cap \{\pi(\tau_1(p)) = x_{\max}\})$.

Set

$$M = \{x_{\max} \in S(p)\}$$

and

$$W = \{S(p) \neq P\} \cap \{\pi(\tau_1(p)) = x_{\max}\}.$$

Note that if $S(p) \neq P$ and $x_{\max} \in S(p)$ then $\pi(\tau_1(p)) \neq x_{\max}$ and hence $W \subset M^c$.

Therefore,

$$\mathbb{P}(W) = \mathbb{P}(W \cap M^c) = \mathbb{P}(M^c)\mathbb{P}(W|M^c) = (1-p)\mathbb{P}(W|M^c).$$

Since, by the bound from Lemma 3.10,

$$\mathbb{P}(W) > p \log \left(\frac{1}{p} \right),$$

and the quantity that we are interested in is $\mathbb{P}(W|M^c)$, the result follows. \square

The theorem can now be proved.

PROOF OF THEOREM 3.9. It is clear that

$$\mathbb{P}\left(|\max(P) \cap S(p)| = k-1\right) = kp^{k-1}(1-p).$$

Thus, for general k ,

$$\begin{aligned} \mathbb{P}\left(\pi(\tau_k(p)) \in \max(P)\right) &> \mathbb{P}\left(\pi(\tau_k(p)) \in \max(P) \mid |\max(P) \cap S(p)| = k-1\right) \\ &\quad \cdot \mathbb{P}\left(|\max(P) \cap S(p)| = k-1\right) \\ &> \frac{p}{1-p} \log \left(\frac{1}{p} \right) \cdot kp^{k-1}(1-p) \\ &= kp^k \log \left(\frac{1}{p} \right), \end{aligned}$$

as required. \square

This theorem has the following consequence. The probability $e^{-\frac{1}{k}}$ is chosen to maximize the function in Theorem 3.9 and gives the lower bound in Theorem 3.2. As described in Chapter 1, it is well-known that $\frac{1}{e}$ is the best possible lower bound for the probability of success in the classical secretary problem, and so this completes the proof of Theorem 3.2.

COROLLARY 3.12. *Let (P, \prec) be a poset with n elements of which k are maximal.*

Then

$$\mathbb{P}\left(\pi\left(\tau_k\left(e^{-\frac{1}{k}}\right)\right) \in \max(P)\right) > \frac{1}{e}.$$

□

3.3. Upper bound

In this section, I shall show that the bound in Corollary 3.8 is best possible. The proof of Theorem 3.5 shows that the probability of success of the stopping time $\tau_k(p_k)$ on k disjoint chains decreases towards the given lower bound as the chains increase in length. This might suggest that the probability of success of an optimal strategy is reduced as the chains increase in length and thus, to prove that the bounds are best possible, one should consider chains of length tending to infinity. The main theorem in this section, Theorem 3.13, does just that; for sufficiently long chains, the probability of success of an optimal stopping time can be made arbitrarily close to that in Corollary 3.8, and so $\tau_k(p_k)$ is asymptotically optimal. Since the poset consisting of k disjoint chains satisfies the conditions of Corollary 3.8, the bounds given are the best possible such bounds.

Define $D_k(x)$ to be the poset consisting of k disjoint chains, each of size x . It might be useful at this point to recall some definitions from Section 1.4. The probability space associated with the poset $D_k(x)$ is denoted by $(\Omega_{D_k(x)}, \mathcal{F}_{D_k(x)}, \mathbb{P}_{D_k(x)})$, but the subscripts are suppressed when they are clear from the context. The poset induced by the first t elements observed is isomorphic to the random variable P_t , which is a poset on vertex set $[t]$, and \mathcal{F}_t is the σ -algebra generated by P_1, \dots, P_t , which represents what is known at time t . A stopping time is a random variable τ taking values in $[n]$ and satisfying the property

$$\{\tau = t\} \in \mathcal{F}_t.$$

Let $\mathcal{C}_{(\mathcal{F}_t)}$ denote the class of all such stopping times, and extend this notation to any sequence of σ -algebras in the analogous way.

The aim is to find an upper bound for $\mathbb{E}(Z_\tau)$ that holds for all stopping times τ , where

$$Z_t = \mathbb{P}\left(\pi(t) \in \max(D_k(x)) \middle| \mathcal{F}_t\right).$$

Theorem 3.13 states that, as $x \rightarrow \infty$, the limit of the probability of success of an optimal stopping time on $D_k(x)$ is no greater than p_k . Since Corollary 3.8 showed the existence of a stopping time with probability of success at least p_k , Theorem 3.13 shows that this is the best possible such bound and so gives Theorem 3.1.

THEOREM 3.13. *Let p_k be as defined in (3.4). Then*

$$\lim_{x \rightarrow \infty} \sup_{\tau \in \mathcal{C}(\mathcal{F}_t)} \mathbb{E}_{D_k(x)}(Z_\tau) \leq p_k.$$

Note that the supremum is over stopping times in $\mathcal{C}(\mathcal{F}_t)$, which means that we are allowed to use the extra information from the structure of the posets, not just the payoffs that we are offered.

The following observation is important and so is recorded as a lemma.

LEMMA 3.14. *When $(P, \prec) = D_k(x)$, it is the case that*

$$Z_t = \begin{cases} \frac{y}{x} & \text{if } t \in \max(P_t), \pi(t) \in C \text{ and } |C \cap \{\pi(1), \dots, \pi(t)\}| = y, \\ 0 & \text{if } t \notin \max(P_t), \end{cases}$$

where C is one of the k chains in $D_k(x)$.

PROOF. The maximal element of a chain C is equally likely to be at any position in the order in which its x elements are observed. Therefore, when y elements have been observed from this chain, the probability that one of them is the maximal element is $\frac{y}{x}$, independent of the most recently observed element being maximal. \square

The proof of Theorem 3.13 will proceed roughly as follows. At time t we expect to have seen approximately $\frac{t}{k}$ elements from each chain. Therefore, since all orders are equally likely, the t^{th} element observed is maximal in what has been seen so far with probability approximately $\frac{1}{t/k} = \frac{k}{t}$. By Lemma 3.14, if this happens then it is a maximal element of $D_k(x)$ with probability approximately $\frac{t}{kx}$. Therefore, Z_t is approximately distributed as

$$Z_t = \begin{cases} \frac{t}{kx} & \text{with probability } \frac{k}{t}, \\ 0 & \text{with probability } 1 - \frac{k}{t}. \end{cases} \quad (3.5)$$

If Z_t were distributed exactly like this with the Z_t all independent of each other, then the proof would not be difficult to complete. Since the potential non-zero value of Z_t increases with t , it is straightforward to show (as in the classical secretary problem; more details will be given later in this section) that an optimal strategy is, for some I , to ignore the first I elements and accept the next non-zero Z_t . I will denote the associated stopping time by τ_I and make some rough calculations. This is only an outline of the more precise argument that will be given later; in particular, \approx is only intended to have an intuitive meaning and does not stand for any well-defined relation. The approximate distribution of Z_t gives

$$\begin{aligned} \mathbb{E}(Z_{\tau_I}) &= \sum_{t=I+1}^{kx} \mathbb{P}(\tau_I = t) \mathbb{E}(Z_t | \tau_I = t) \\ &\approx \sum_{t=I+1}^{kx} \mathbb{P}\left(\left\{Z_i = 0 \text{ for all } i \in \{I+1, \dots, t-1\}\right\} \cap \{Z_t > 0\}\right) \cdot \frac{t}{kx} \\ &\approx \sum_{t=I+1}^{kx} \left(\prod_{i=I+1}^{t-1} \left(1 - \frac{k}{i}\right)\right) \cdot \frac{k}{t} \cdot \frac{t}{kx} \\ &= \frac{1}{x} \sum_{t=I+1}^{kx} \prod_{i=I+1}^{t-1} \left(1 - \frac{k}{i}\right). \end{aligned}$$

I shall apply this formula in the case where k is much smaller than I , so $1 - \frac{k}{i}$ can be approximated by $e^{-\frac{k}{i}}$, and sums are twice approximated by integrals to give

$$\begin{aligned} \mathbb{E}(Z_{\tau_I}) &\approx \frac{1}{x} \sum_{t=I+1}^{kx} e^{-\sum_{i=I+1}^{t-1} \frac{k}{i}} \approx \frac{1}{x} \sum_{t=I+1}^{kx} e^{-k \log\left(\frac{t}{I}\right)} \\ &= \frac{I^k}{x} \sum_{t=I+1}^{kx} t^{-k} \approx \begin{cases} \frac{I}{x} \cdot \log\left(\frac{x}{I}\right) & \text{if } k = 1, \\ \frac{k}{k-1} \cdot \frac{I}{kx} \cdot \left(1 - \left(\frac{I}{kx}\right)^{k-1}\right) & \text{if } k > 1. \end{cases} \end{aligned}$$

This is the formula in Theorem 3.7 with $p = \frac{I}{kx}$ and is thus maximized, as in Corollary 3.8, when $\frac{I}{kx} = p_k$, in which case $\mathbb{E}(Z_{\tau_I}) \approx p_k$. Therefore the bounds in Corollary 3.8 are best possible, and if these calculations had been exact then the proof of Theorem 3.1 would be complete.

Unfortunately, Z_t is not distributed exactly as in (3.5). In order to conclude that an optimal stopping time is of the simple form above, it would be nice to use the principle of

backward induction described in Section 1.5. This formalizes the intuitive principle that, in a finite game, an optimal strategy is simply to analyse at each step whether or not we expect our situation to improve by continuing, and to do so if and only if this is the case.

The reason why the sequence of random variables $(Z_t)_{t \in [n]}$ is difficult to analyse is that the values they can take vary depending on how the process unfolds. However, it is very likely that at any time we shall have seen approximately the same number of elements from each chain. The proof will therefore proceed by defining a sequence of random variables $(Y_t)_{t \in [n]}$, which act as asymptotically almost sure upper bounds for Z_t and are easier to analyse. To obtain these bounds, each chain will be split into m segments, each of length l , and the process split into m sets of kl observations. These lengths l are margins of error beyond which the number of elements observed from a chain is not expected to deviate. Initially, I shall fix m and let $l \rightarrow \infty$ to find an upper bound for $\mathbb{E}(Y_\tau)$ and hence $\mathbb{E}(Z_\tau)$ in terms of m . Letting $m \rightarrow \infty$ will then give a best possible result.

This means that the precise statement proved for Theorem 3.13 is in fact

$$\lim_{m \rightarrow \infty} \limsup_{l \rightarrow \infty} \sup_{\tau \in \mathcal{C}(\mathcal{F}_t)} \mathbb{E}_{D_k(lm)}(Z_\tau) \leq p_k.$$

However, this is purely a matter of convenience; it is clear that the proof can be extended to posets $D_k(x)$ where x is not a multiple of m by dividing each chain into m almost equal rather than exactly equal segments.

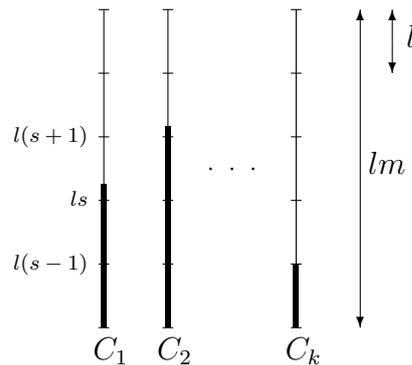


FIGURE 3.3. This figure shows the number of elements observed from each chain. In this example, after a total of kl_s elements have been observed we see that $U_{C_1,s}$ and $U_{C_k,s}$ hold but $U_{C_2,s}$ does not.

It is necessary to show that the process behaves in this approximately uniform manner with high probability as $l \rightarrow \infty$. I shall first define what it means to be approximately

uniform in one particular chain C at time $kl s$, an event called $U_{C,s}$ (see Figure 3.3), and then what it means to be approximately uniform everywhere at all times, an event called U .

Given one of the chains, C , and for all $s \in \{0, \dots, m\}$, let $U_{C,s}$ be the event that when we have observed $kl s$ elements in total we have observed between $l(s-1)$ and $l(s+1)$ elements from C , that is,

$$U_{C,s} = \left\{ l(s-1) \leq \left| C \cap \{\pi(1), \dots, \pi(kl s)\} \right| \leq l(s+1) \right\}.$$

For all $t \in \{0, \dots, klm\}$, let $s(t)$ be the unique integer s such that $kl(s-1) < t \leq kl s$, that is,

$$s(t) = \left\lceil \frac{t}{kl} \right\rceil. \quad (3.6)$$

Let U be the event that for all t when we have observed t elements in total we have observed between $l(s(t)-2)$ and $l(s(t)+1)$ elements from each chain, that is,

$$U = \bigcap_{i,t} \left\{ l(s(t)-2) \leq \left| C_i \cap \{\pi(1), \dots, \pi(t)\} \right| \leq l(s(t)+1) \right\}.$$

I shall use Lemma 3.16, which states that the process is approximately uniform with high probability and follows easily from Lemma 3.15.

LEMMA 3.15. *Let $m \geq 1$ be an integer, let C be one of the chains in $D_k(lm)$ and let $s \in \{0, \dots, m\}$. Then*

$$\lim_{l \rightarrow \infty} \mathbb{P}_{D_k(lm)}(U_{C,s}) = 1.$$

PROOF. I shall show that the probability that we have observed more than $l(s+1)$ or fewer than $l(s-1)$ elements tends to zero as $l \rightarrow \infty$. (If $s = 0$ or $s = m$ then only one of these tails needs to be considered.)

Assume C and s are given. Let N be the number of elements observed from chain C when $kl s$ elements have been observed in total, and write $f(x) = \mathbb{P}(N = x)$. Then

$$f(x) = \frac{\binom{lm}{x} \binom{(k-1)lm}{kl s - x}}{\binom{klm}{kl s}}$$

and hence

$$\frac{f(x+1)}{f(x)} = \frac{(lm-x)(kl s - x)}{(x+1)((k-1)lm - kl s + x + 1)}.$$

Note that this is a decreasing function in x . Thus, if $x \geq ls$, then

$$\frac{f(x+1)}{f(x)} < \frac{l(m-s)(k-1)ls}{ls(k-1)l(m-s)} = 1 \quad (3.7)$$

and, if $x+1 \leq ls$, then

$$\frac{f(x+1)}{f(x)} > \frac{l(m-s)(k-1)ls}{ls(k-1)l(m-s)} = 1. \quad (3.8)$$

Observe also that, if $x \geq l(s+1)$, then

$$\frac{f(x+1)}{f(x)} < \frac{(m-s-1)((k-1)s-1)}{(s+1)((k-1)(m-s)+1)} = c_{s+1} < 1 \quad (3.9)$$

and, if $x+1 \leq l(s-1)$, then

$$\frac{f(x+1)}{f(x)} > \frac{(m-s+1)((k-1)s+1)}{(s-1)((k-1)(m-s)-1)} = \frac{1}{c_{s-1}} > 1, \quad (3.10)$$

where c_{s+1} and c_{s-1} are constants depending on k , m and s but not l .

Since (3.7) and (3.8) imply that $f(l(s+1)) < \dots < f(ls)$ and $f(l(s-1)) < \dots < f(ls)$, it follows that $f(l(s+1)) < \frac{1}{l}$ and $f(l(s-1)) < \frac{1}{l}$. Finally, (3.9) and (3.10) give

$$\mathbb{P}(N > l(s+1)) < \sum_{i=0}^{\infty} \frac{c_{s+1}^i}{l} = \frac{1}{l(1-c_{s+1})} \rightarrow 0 \text{ as } l \rightarrow \infty$$

and

$$\mathbb{P}(N < l(s-1)) < \sum_{i=0}^{\infty} \frac{c_{s-1}^i}{l} = \frac{1}{l(1-c_{s-1})} \rightarrow 0 \text{ as } l \rightarrow \infty,$$

which proves the claim. \square

LEMMA 3.16. *Let $m \geq 1$ be an integer. Then*

$$\lim_{l \rightarrow \infty} \mathbb{P}_{D_k(lm)}(U) = 1.$$

PROOF. This lemma follows simply from the previous lemma: choose l sufficiently large that each of the $k(m+1)$ events $U_{C_{i,s}}$ occurs with probability at least $1 - \delta$. Then, trivially, all $k(m+1)$ events hold with probability at least $1 - k(m+1)\delta$. It is easy to see that

$$\bigcap_{C_{i,s}} U_{C_{i,s}} \subset U,$$

since the events $U_{C_i, s(t)-1}$ and $U_{C_i, s(t)}$ imply that

$$l(s(t) - 2) \leq \left| C_i \cap \{\pi(1), \dots, \pi(t)\} \right| \leq l(s(t) + 1),$$

and so this holds for every t and i , as required. \square

The next lemma states that in order to prove Theorem 3.13, it suffices to show that its statement is true conditioned on U occurring. This formalizes the intuition that, since the process is asymptotically almost surely uniform (that is, since $\lim_{l \rightarrow \infty} \mathbb{P}_{D_k(lm)}(U) \rightarrow 1$), we may assume that it is uniform.

Recall that $\mathcal{C}_{(\mathcal{F}_t)}$ is the class of all stopping times relative to the σ -algebras $\mathcal{F}_t = \sigma(P_1, \dots, P_t)$, that is, the decision to stop at time t depends only on P_1, \dots, P_t .

LEMMA 3.17. *For all m ,*

$$\limsup_{l \rightarrow \infty} \sup_{\tau \in \mathcal{C}_{(\mathcal{F}_t)}} \mathbb{E}_{D_k(lm)}(Z_\tau) \leq \limsup_{l \rightarrow \infty} \sup_{\tau \in \mathcal{C}_{(\mathcal{F}_t)}} \mathbb{E}_{D_k(lm)}(Z_\tau | U).$$

PROOF. By Lemma 3.16, for all $\varepsilon > 0$ the value of l may be chosen sufficiently large that $\mathbb{P}_{D_k(lm)}(U^c) \leq \varepsilon$. By definition, it is also the case that $Z_t \leq 1$ for all t . Therefore, for all τ ,

$$\begin{aligned} \mathbb{E}_{D_k(lm)}(Z_\tau) &= \mathbb{E}_{D_k(lm)}(Z_\tau | U) \mathbb{P}_{D_k(lm)}(U) + \mathbb{E}_{D_k(lm)}(Z_\tau | U^c) \mathbb{P}_{D_k(lm)}(U^c) \\ &\leq \mathbb{E}_{D_k(lm)}(Z_\tau | U) + \varepsilon. \end{aligned}$$

Taking suprema gives

$$\sup_{\tau \in \mathcal{C}_{(\mathcal{F}_t)}} \mathbb{E}_{D_k(lm)}(Z_\tau) \leq \sup_{\tau \in \mathcal{C}_{(\mathcal{F}_t)}} \mathbb{E}_{D_k(lm)}(Z_\tau | U) + \varepsilon.$$

Since ε is arbitrary, the result follows. \square

Now that I have shown that the process is asymptotically almost surely uniform, and in light of the previous lemma, I shall assume that U occurs. The next aim is to find a process that offers at least the pay-offs that Z_t offers but is easier to analyse. This means that I shall be able to find an upper bound for the probability of success of an optimal stopping time on this process and hence on Z_t .

I shall now define the random variables Y_t that act as upper bounds for the $(Z_t|U)$ and are easier to analyse. These random variables are not strict upper bounds, in the sense that the random variables are not coupled in any way. However, conditioned on U occurring, Z_t is less than the potential non-zero value of Y_t and the probability that Z_t is non-zero is less than the probability that Y_t is non-zero, and I shall be able to show that an optimal strategy for the game on Z_t has a lower expected pay-off than an optimal strategy for the game on Y_t .

I shall often need to refer to the potential non-zero value of Y_t and the probability that it takes this value; set

$$y_t = \frac{s(t) + 1}{m} \quad \text{and} \quad p_t = \begin{cases} \frac{1}{l(s(t)-2)} & \text{if } s(t) \geq 3, \\ 1 & \text{if } s(t) \leq 2, \end{cases}$$

where $s(t)$ is as defined in (3.6). Now define a sequence of independent random variables $(Y_t)_{t \in [n]}$ by

$$Y_t = \begin{cases} y_t & \text{with probability } p_t, \\ 0 & \text{with probability } 1 - p_t. \end{cases}$$

These will not be explicitly defined on Ω as there is no need, although it is of course straightforward to do so.

The next lemma states the intuitive principle that we expect to do at least as well in the game with the random variables Y_t as in the game with the random variables Z_t conditioned on U occurring. Analogously to \mathcal{F}_t for Z_t , let $(\mathcal{G}_t)_{t \in [n]}$ be defined by $\mathcal{G}_t = \sigma(Y_1, \dots, Y_t)$.

LEMMA 3.18. *For all $l, m \in \mathbb{N}$ with $m \geq 3$,*

$$\sup_{\tau \in \mathcal{C}(\mathcal{F}_t)} \mathbb{E}_{D_k(lm)}(Z_\tau | U) < \sup_{\tau \in \mathcal{C}(\mathcal{G}_t)} \mathbb{E}_{D_k(lm)}(Y_\tau).$$

This will be proved shortly. Putting Lemmas 3.17 and 3.18 together tells us that

$$\lim_{l \rightarrow \infty} \sup_{\tau \in \mathcal{C}(\mathcal{F}_t)} \mathbb{E}_{D_k(lm)}(Z_\tau) \leq \lim_{l \rightarrow \infty} \sup_{\tau \in \mathcal{C}(\mathcal{G}_t)} \mathbb{E}_{D_k(lm)}(Y_\tau).$$

The proof of the lemma relies on Theorem 1.1, which was the theorem that formalized backward induction.

PROOF OF LEMMA 3.18. For convenience, I shall continue to use n to denote the number of elements in $D_k(lm)$, that is, $n = klm$. I shall apply Theorem 1.1 to the sequences $(Y_t)_{t \in [n]}$ and $(Z_t)_{t \in [n]}$, conditioned on U occurring, to show that the optimal expected pay-off for $(Y_t)_{t \in [n]}$ is at least as large for $(Z_t)_{t \in [n]}$.

First consider what happens with the sequence $(Y_t)_{t \in [n]}$. Recall that $(\mathcal{G}_t)_{t \in [n]}$ are defined by $\mathcal{G}_t = \sigma(Y_1, \dots, Y_t)$ and let $(\alpha_t)_{t \in [n]}$ be defined for $(Y_t)_{t \in [n]}$ as $(\gamma_t)_{t \in [n]}$ were for $(W_t)_{t \in [n]}$ in the backward induction theorem, that is,

$$\begin{aligned}\alpha_n &= Y_n, \\ \alpha_t &= \max \{Y_t, \mathbb{E}(\alpha_{t+1} | \mathcal{G}_t)\}, \quad t = n-1, \dots, 1.\end{aligned}$$

Since the random variables Y_t are independent, the values of Y_1, \dots, Y_t give no information about the values of Y_{t+1}, \dots, Y_n , and therefore $\mathbb{E}(\alpha_{t+1} | \mathcal{G}_t)$ is constant on all atoms of \mathcal{G}_t and equal to $\mathbb{E}(\alpha_{t+1})$. Therefore, define the function $v : [n] \rightarrow \mathbb{R}$ by

$$v(t) = \mathbb{E}(\alpha_t)$$

and note that backward induction tells us that the stopping time that stops at the first t such that $Y_t \geq \mathbb{E}(\alpha_{t+1} | \mathcal{G}_t) = v(t+1)$ is optimal. By definition,

$$v(t) = \mathbb{E}(\max\{Y_t, v(t+1)\}) \geq v(t+1),$$

whereas y_t , the potential non-zero value of Y_t , is a non-decreasing function of t . Therefore, there exists I such that

$$\begin{aligned}y_t &< v(t+1) \quad \text{if } t \leq I \quad \text{and} \\ y_t &\geq v(t+1) \quad \text{if } t > I,\end{aligned}$$

and an optimal strategy for the game on Y_t is ‘reject the first I elements, and accept the next with a non-zero pay-off.’

Recall that the distribution of Y_t is given by

$$Y_t = \begin{cases} y_t & \text{with probability } p_t, \\ 0 & \text{with probability } 1 - p_t. \end{cases}$$

From this, it follows that

$$v(n) = \mathbb{E}(\alpha_n) = \mathbb{E}(Y_n) = p_n y_n = \frac{m+1}{m} \cdot \frac{1}{l(m-2)} > \frac{1}{lm} = \frac{k}{n} \quad (3.11)$$

and that, for $1 \leq t \leq n-1$,

$$v(t) = \mathbb{E}(\alpha_t) = \mathbb{E}(\max\{Y_t, \mathbb{E}(\alpha_{t+1})\}) = \begin{cases} p_t y_t + (1-p_t)v(t+1) & \text{if } t > I, \\ v(t+1) & \text{if } t \leq I. \end{cases} \quad (3.12)$$

Now consider the sequence of random variables $(Z_t)_{t \in [n]}$. Since the statement of the lemma is conditioned on U occurring, define a new sequence of σ -algebras $(\mathcal{H}_t)_{t \in [n]}$ by

$$\mathcal{H}_t = \sigma(\mathcal{F}_t \cup \{U\})$$

and consider only $\omega \in U$. Analogously to γ_t and α_t , let

$$\beta_n = Z_n$$

and, for $1 \leq t \leq n-1$, let

$$\beta_t = \max\{Z_t, \mathbb{E}(\beta_{t+1} | \mathcal{H}_t)\}.$$

Recalling that $\mathcal{C}_{(\mathcal{F}_t)}$ denotes the class of stopping times relative to \mathcal{F}_t , observe that $\mathcal{F}_t \subset \mathcal{H}_t$, and hence $\mathcal{C}_{(\mathcal{F}_t)} \subset \mathcal{C}_{(\mathcal{H}_t)}$ and

$$\sup_{\tau \in \mathcal{C}_{(\mathcal{F}_t)}} \mathbb{E}_{D_k(lm)}(Z_\tau | U) \leq \sup_{\tau \in \mathcal{C}_{(\mathcal{H}_t)}} \mathbb{E}_{D_k(lm)}(Z_\tau | U).$$

Note that intuitively this is obvious: it just says that having extra information (that the event U occurs) can only help in choosing our stopping time τ .

Recall that $\mathbb{E}(\beta_t|\mathcal{H}_{t-1})(\omega)$ denotes the expected value of the game at time t , if we have so far seen the first $t - 1$ elements of P and are told whether or not U holds (that is, whether or not $\omega \in U$). I shall prove the following claim by induction on $n - t$.

CLAIM. For all $\omega \in U$ and for all $t \in [n]$,

$$\mathbb{E}(\beta_t|\mathcal{H}_{t-1})(\omega) < v(t),$$

where $\mathcal{H}_0 = \{\emptyset, U, U^c, \Omega\}$ and so $\mathbb{E}(\beta_1|\mathcal{H}_0)(\omega) = \mathbb{E}(\beta_1|U)$.

PROOF OF CLAIM. First, observe that if $\omega \in U$, then

$$Z_t(\omega) \leq y_t \quad \text{and} \quad \mathbb{P}(Z_t > 0 | \mathcal{H}_{t-1})(\omega) \leq p_t,$$

by Lemma 3.14 and the definition of U . Moreover, by Lemma 3.14 and (3.11),

$$\mathbb{E}(\beta_n|\mathcal{H}_{n-1}) = \frac{1}{n/k} \cdot 1 = \frac{k}{n} < v(n),$$

which proves the case $n - t = 0$.

Now let $1 \leq t \leq n - 1$ and assume that the result holds for $t + 1$. Since \mathcal{H}_{t-1} is finite, the random variable $\mathbb{E}(\beta_t|\mathcal{H}_{t-1})$ is constant on each of its atoms. Given $\omega \in U$, let $A \in \mathcal{H}_{t-1}$ be the atom containing ω , so that

$$\begin{aligned} \mathbb{E}(\beta_t|\mathcal{H}_{t-1})(\omega) &= \mathbb{E}(\beta_t|A) \\ &= \mathbb{P}(Z_t \geq \mathbb{E}(\beta_{t+1}|\mathcal{H}_t) | A) \cdot \mathbb{E}(Z_t | (Z_t \geq \mathbb{E}(\beta_{t+1}|\mathcal{H}_t)) \cap A) \\ &\quad + \mathbb{P}(Z_t < \mathbb{E}(\beta_{t+1}|\mathcal{H}_t) | A) \cdot \mathbb{E}(\mathbb{E}(\beta_{t+1}|\mathcal{H}_t) | (Z_t < \mathbb{E}(\beta_{t+1}|\mathcal{H}_t)) \cap A). \end{aligned}$$

Now, $(Z_t < \mathbb{E}(\beta_{t+1}|\mathcal{H}_t)) \in \mathcal{H}_t$ and $\mathcal{H}_{t-1} \subset \mathcal{H}_t$, and so $(Z_t < \mathbb{E}(\beta_{t+1}|\mathcal{H}_t)) \cap A \in \mathcal{H}_t$. Note first that, since $\omega \in U$ and $U \in \mathcal{H}_{t-1}$, it must be the case that $A \subset U$, and second that $(Z_t < \mathbb{E}(\beta_{t+1}|\mathcal{H}_t)) \cap A$ is non-empty since $\mathbb{P}(Z_t < \mathbb{E}(\beta_{t+1}|\mathcal{H}_t) | A) \geq 1 - p_t > 0$.

For all $\omega' \in (Z_t < \mathbb{E}(\beta_{t+1}|\mathcal{H}_t)) \cap A$, by the induction hypothesis,

$$\mathbb{E}(\beta_{t+1}|\mathcal{H}_t)(\omega') < v(t + 1)$$

and therefore

$$\mathbb{E} \left(\beta_{t+1} \middle| (Z_t < \mathbb{E}(\beta_{t+1} | \mathcal{H}_t)) \cap A \right) < v(t+1).$$

Moreover,

$$\mathbb{E} \left(Z_t \middle| (Z_t \geq \mathbb{E}(\beta_{t+1} | \mathcal{H}_t)) \cap A \right) \leq y_t,$$

since $Z_t \leq y_t$ whenever U holds and $A \subset U$. Since $y_t \leq v(t+1)$ if and only if $t \leq I$, and $\mathbb{P}(Z_t \geq \mathbb{E}(\beta_{t+1} | \mathcal{H}_t) | A) \leq \mathbb{P}(Z_t > 0 | A) \leq p_t < 1$, it follows that

$$\mathbb{E}(\beta_t | \mathcal{H}_{t-1})(\omega) < \begin{cases} p_t y_t + (1 - p_t)v(t+1) & \text{if } t > I, \\ v(t+1) & \text{if } t \leq I, \end{cases}$$

and hence that $\mathbb{E}(\beta_t | \mathcal{H}_{t-1})(\omega) < v(t)$, by (3.12). This completes the induction step, and hence proves the claim. \square

The lemma follows from the claim, since, by the backward induction theorem (Theorem 1.1), it follows that

$$\sup_{\tau \in \mathcal{C}(\mathcal{H}_t)} \mathbb{E}_{D_k(lm)}(Z_\tau | U) = \mathbb{E}(\beta_1 | U) < v(1) = \mathbb{E}(\alpha_1) = \sup_{\tau \in \mathcal{C}(\mathcal{G}_t)} \mathbb{E}_{D_k(lm)}(Y_\tau),$$

as required. \square

In the final lemma before the proof of Theorem 3.13, backward induction is used to show that an optimal stopping time for the process with the Y_t takes the simple form of rejecting the first klu^* elements, for some integer u^* , and accepting the next non-zero pay-off.

LEMMA 3.19. For $u^* \in \{0, \dots, m-1\}$, let

$$\tau_{u^*} = \begin{cases} \min \{t > klu^* : Y_t > 0\} & \text{if this exists,} \\ n & \text{otherwise.} \end{cases}$$

Then

$$\sup_{\tau \in \mathcal{C}(\mathcal{G}_t)} \mathbb{E}_{D_k(lm)}(Y_\tau) = \sup_{u^* \in \{0, \dots, m-1\}} \mathbb{E}_{D_k(lm)}(Y_{\tau_{u^*}})$$

PROOF. It has already been shown in the proof of Lemma 3.18 that an optimal strategy takes the form ‘ignore the first I , and accept the next non-zero pay-off.’ In

fact, I must be a multiple of kl : suppose, for contradiction, that $I = klu^* + r$, where $r \in \{1, \dots, kl - 1\}$. Then it must be the case that

$$v(I + 1) \leq \frac{kl(s(I + 1) + 1)}{m} = \frac{kl(u^* + 2)}{m},$$

since we would be willing to stop at time $I + 1$, but then

$$v(I) \leq \max \left\{ \frac{kl(u^* + 2)}{m}, v(I + 1) \right\} \leq \frac{kl(u^* + 2)}{m} = \frac{kl(s(I) + 1)}{m},$$

which is a contradiction since we would not be willing to stop at time I . \square

The proof of the main theorem in this section can now be completed.

PROOF OF THEOREM 3.13. All that remains is to calculate and maximize $\mathbb{E}(Y_{\tau_{u^*}})$ over u^* , where τ_{u^*} is the smallest $t > klu^*$ such that $Y_t > 0$. These calculations are very similar to those on page 68, but also include error terms which tend to zero as $l, m \rightarrow \infty$.

Assume first that $u^* \rightarrow \infty$ as $m \rightarrow \infty$, as the calculation will show that this is a valid assumption. Indeed, if $t = o(n)$ then $y_t = o(1)$, whereas a probability of success will be obtained that is separated from zero. We should therefore never accept a pay-off for $t = o(n)$. In particular, for sufficiently large m it is the case that $u^* \geq 3$, and so

$$\mathbb{E}(Y_{\tau_{u^*}}) = \sum_{u=u^*}^{m-1} \sum_{r=1}^{kl} \mathbb{P}(Y_{klu^*+1} = 0, \dots, Y_{klu^*+r-1} = 0, Y_{klu^*+r} > 0) \cdot y_{klu^*+r}.$$

Recall that the Y_t are independent, and that $s(klu + h) = u + 1$ when $1 \leq h \leq kl$. Hence, using the convention that the empty product takes the value 1, it follows that

$$\mathbb{E}(Y_{\tau_{u^*}}) = \sum_{u=u^*}^{m-1} \sum_{r=1}^{kl} \left(\prod_{q=u^*+1}^u \left(1 - \frac{1}{l(q-2)} \right)^{kl} \right) \left(1 - \frac{1}{l(u-1)} \right)^{r-1} \cdot \frac{1}{l(u-1)} \cdot \frac{u+2}{m}.$$

The formula for the sum of the geometric progression $\left(1 - \frac{1}{l(u-1)} \right)^{r-1}$ gives

$$\mathbb{E}(Y_{\tau_{u^*}}) = \sum_{u=u^*}^{m-1} \left(\prod_{q=u^*+1}^u \left(1 - \frac{1}{l(q-2)} \right)^{kl} \right) \left(1 - \left(1 - \frac{1}{l(u-1)} \right)^{kl} \right) \cdot \frac{u+2}{m},$$

and multiplying out the term $\left(1 - \left(1 - \frac{1}{l(u-1)}\right)^{kl}\right)$ gives

$$\begin{aligned} \mathbb{E}(Y_{\tau_{u^*}}) &= \sum_{u=u^*}^{m-1} \left(\prod_{q=u^*+1}^u \left(1 - \frac{1}{l(q-2)}\right)^{kl} \right) \cdot \frac{u+2}{m} \\ &\quad - \sum_{u=u^*+1}^m \left(\prod_{q=u^*+1}^u \left(1 - \frac{1}{l(q-2)}\right)^{kl} \right) \cdot \frac{u+1}{m}. \end{aligned}$$

Parts of these two sums cancel each other out, leaving

$$\begin{aligned} \mathbb{E}(Y_{\tau_{u^*}}) &= \frac{u^*+2}{m} - \left(\prod_{q=u^*+1}^m \left(1 - \frac{1}{l(q-2)}\right)^{kl} \right) \cdot \frac{m+1}{m} \\ &\quad + \sum_{u=u^*+1}^{m-1} \left(\prod_{q=u^*+1}^u \left(1 - \frac{1}{l(q-2)}\right)^{kl} \right) \cdot \frac{1}{m}. \end{aligned}$$

Now, let $\varepsilon > 0$ be arbitrary, choose $m = m(\varepsilon)$ and $l = l(m, \varepsilon)$ sufficiently large, and recall that therefore $u^* = u^*(\varepsilon)$ may be chosen to be sufficiently large also. Since $(1 - \frac{1}{n})^n < \frac{1}{e} < (1 - \frac{1}{n})^{n-1}$ for all $n \geq 2$, it is the case that

$$\prod_{q=u^*+1}^m \left(1 - \frac{1}{l(q-2)}\right)^{kl} \geq \exp\left(-\frac{kl-1}{l} \sum_{q=u^*+1}^m \frac{1}{q-2}\right) \geq \left(\frac{u^*}{m}\right)^k - \frac{\varepsilon}{2},$$

and similarly

$$\prod_{q=u^*+1}^u \left(1 - \frac{1}{l(q-2)}\right)^{kl} \leq \exp\left(-k \sum_{q=u^*+1}^u \frac{1}{q-2}\right) \leq \left(\frac{u^*}{u}\right)^k.$$

Setting $p = \frac{u^*}{m}$, it follows that

$$\mathbb{E}(Y_{\tau_{u^*}}) \leq \begin{cases} p - p + p \log\left(\frac{1}{p}\right) + \varepsilon & \text{if } k = 1, \\ p - p^k + \frac{p}{k-1} (1 - p^{k-1}) + \varepsilon & \text{if } k > 1. \end{cases}$$

As before, these expressions are maximized when $p = p_k$, and thus

$$\lim_{m \rightarrow \infty} \lim_{l \rightarrow \infty} \mathbb{E}(Y_{\tau_{u^*}}) \leq p_k + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this completes the proof. \square

Putting Corollary 3.8 and Theorem 3.13 together gives Theorem 3.1.

3.4. Open problems

It seems likely that these theorems can be improved. Firstly, Robert Morris and I [35] conjectured that Theorem 3.1 can be extended to all posets with k maximal elements. It is not inconceivable that the algorithm in Corollary 3.8 works on all posets with k maximal elements rather than just those whose width is also k ; if not, it would be good to find some other algorithm dependent only on k that does so.

CONJECTURE 3.20. *Let (P, \prec) be a poset with k maximal elements. Then there is an algorithm for the secretary problem on (P, \prec) that is successful with probability at least p_k , where p_k is as defined in (3.1).*

Our algorithm, which gives the bound in Theorem 3.2, depends only on the size of the poset and the number of maximal elements. Our second conjecture was that the latter piece of information is not needed.

CONJECTURE 3.21. *Let (P, \prec) be a poset. Then there is an algorithm for the secretary problem on (P, \prec) that depends only on $|P|$ and is successful with probability at least $\frac{1}{e}$.*

In light of the threshold probabilities $e^{-\frac{1}{k}}$ in Corollary 3.12, we considered the stopping time

$$\tau = \begin{cases} \min \left\{ t : t > e^{-\frac{1}{m}} n, \text{ where } m = |\max(P_t)| \text{ and } t \in \max(P_t) \right\} & \text{if this exists,} \\ n & \text{otherwise,} \end{cases}$$

and tried to show that this algorithm might be used to prove Conjecture 3.21, but unfortunately we did not make any progress with its analysis. Kozik [54] considered this stopping time independently at the same time, and was able to show that it was successful with probability strictly greater than $\frac{1}{4}$ for sufficiently large posets. Kozik's proof is based on the analysis of many cases by computer, and he does not specify how much greater. However, whereas the bound of $\frac{1}{4}$ is best possible for Preater's algorithm, Kozik's analysis is not at all tight and, like Robert Morris and me, he is hopeful that this algorithm might be successful with the best possible probability of $\frac{1}{e}$.

After I had finished writing this dissertation, Michał Morayne alerted me to the fact that, very recently, Freij and Wästlund [29] had proved Conjecture 3.21. Their strategy

and proof of its probability of success are beautifully simple. Firstly, they assume that the elements are observed at n times uniformly distributed in $[0, 1]$. Secondly, in addition to this, they assign a weight uniformly distributed in $[0, 1]$ to each element as it appears, independently of the others, and define the *greedy maximum* of a poset (P, \prec) as follows. Let $z_0 \in P$ be the element of minimal weight. If z_0 is not a maximal element, then let z_1 be the element of minimal weight in $\{x : z_0 \prec x\}$. Repeat this process, producing a chain $z_0 \prec z_1 \prec z_2 \prec \dots$, until a maximal element is reached. This element is the greedy maximum.

Their strategy is to wait until time $\frac{1}{e}$ and then to select the next element that is the greedy maximum out of what has been seen so far; their analysis of it is surprisingly brief. Writing A_k for the event that the k^{th} element observed is the greedy maximum out of what has been seen so far, they first show that the joint distribution of $(A_k)_{k \in [n]}$ is independent of the ground poset, and in particular the same as for the single chain. They then show that the probability that a maximal element is the greedy maximum at time t , conditioned on it appearing at time t , is equal to the conditional probability that it is the greedy maximum of the ground poset given that its weight is at most t ; this is at least the probability that it is the greedy maximum of the ground poset. Finally, they carry out a simple calculation for the conditional probability that a maximal element is accepted given that it appears at time t , integrate over t and sum over the maximal elements, and find that with probability at least $\frac{1}{e}$ one of them is accepted.

In this chapter, the aim was to select a maximal element; the same problems can be posed for the minimum rank version instead. In a general poset, the rank of an element x is most naturally defined as the maximum number of elements in a chain whose minimal element is x . In light of the results in this chapter, the following questions are appealing:

- (1) Is it the case that the single chain is the hardest poset for the minimal expected rank problem, that is, is it the case that for every poset there is a strategy that selects an element with expected rank at most about 3.8695?
- (2) Is there a strategy that depends only on the size of the poset and its number of maximal elements, or even just its size, that selects an element of finite expected rank for every poset where these are given?

As with the problems that I suggested at the end of Chapter 2, one could also allow the interviewer to go back to a previous candidate, who is still available with a certain probability, and see how this affects the problem.

This chapter has been about unknown posets with a known number of elements. In Chapter 1, I described a version of the problem where the poset is a chain whose length comes from a known distribution. Distributions that did not seem too unpleasant nevertheless forced the expected rank to infinity. What distributions on the number of elements in an unknown poset admit a strategy that chooses a maximal element with positive probability, or a strategy that selects an element with finite expected rank?

Part 2

Generalizations of acyclic colourings

CHAPTER 4

Acyclic colourings of graphs

4.1. Colouring problems

Throughout the history of graph theory, colouring problems have been widely studied. I shall summarize some of the highlights in this chapter, and show how this part of this dissertation fits into that journey. For a more detailed history and for proofs of these results, see Chapter V of Bollobás [11], whose notation I shall use.

A *proper vertex-colouring* of a graph $G = (V, E)$ is an assignment of colours to its vertices such that the end-points of any edge are differently coloured. The smallest number of colours needed to do this is called the *chromatic number* $\chi(G)$ of G . It is obvious that any graph with maximum degree $d = \Delta(G)$ can be properly coloured with at most $d + 1$ colours: if the vertices are coloured with a palette of $d + 1$ colours one by one in any order, then for each vertex v there is always at least one colour available that has not been used to colour its neighbours, and so v can be given this colour.

The following theorem, due to Brooks [16], says that one can usually do slightly better.

THEOREM 4.1. *Let G be a connected graph. Then $\chi(G) \leq \Delta(G)$ unless G is a complete graph or a cycle of odd length, in which case $\chi(G) = \Delta(G) + 1$. \square*

The chromatic number of a graph can also be bounded if it can be drawn on a particular surface without its edges crossing. The *Euler characteristic* of a surface S , unfortunately usually denoted by $\chi(S)$, is the invariant

$$\chi(S) = V - E + F,$$

where V , E and F are the numbers of vertices, edges and faces in a *triangulation* of S , that is, a graph drawn on S all of whose faces are triangles. In the following theorem, due to Heawood [47], a bound is given for surfaces other than the plane or, equivalently, the sphere, which has Euler characteristic 2.

THEOREM 4.2. *The chromatic number of a graph drawn on a closed surface of Euler characteristic $\chi \leq 1$ is at most*

$$h(\chi) = \left\lfloor \frac{7 + \sqrt{49 - 24\chi}}{2} \right\rfloor.$$

□

The function $h(\chi)$ is known as the Heawood bound, which was proved to be best possible for every surface except the Klein bottle by Ringel and Youngs [72]. The Klein bottle has Euler characteristic 0, but every graph that can be drawn on it has chromatic number at most 6, rather than 7. The famous four-colour theorem, proved by Appel and Haken with computational assistance from Koch [7, 8, 9], states that every graph drawn on the plane can be coloured using at most four colours.

It is just as natural to colour a graph's edges as its vertices; in a *proper edge-colouring* the edges meeting at any one vertex are all coloured differently. This gives rise to the *chromatic index* $\chi'(G)$ of G , the smallest number of colours needed to do this. In this case, it is obvious that $\Delta(G)$ is a lower bound for $\chi'(G)$, since the edges incident to a vertex of maximum degree must all be differently coloured. What is surprising is that the best possible upper bound is only one larger, which was proved by Vizing [78] in the following theorem.

THEOREM 4.3. *Let G be a graph. Then $\Delta(G) \leq \chi'(G) \leq \Delta(G) + 1$.* □

There are of course many more results concerning colourings, and I shall briefly mention some in only one more direction. An *L -colouring* is a proper vertex-colouring where the colour of each vertex v is chosen from a given list $L(v)$ of possible colours for that vertex. The *list-chromatic number* $\chi_l(G)$ is the minimum k such that G can be L -coloured for any L where $|L(v)| \geq k$ for all v . It is clearly the case that $\chi_l(G) \geq \chi(G)$ and one might expect always to have equality, since it seems most restrictive when the $L(v)$ are all the same. This turns out not to be true; Voigt [79] proved that there are some planar graphs with list-chromatic number greater than 4. However, Thomassen [76] proved that the list chromatic number of planar graphs is at most 5.

The *list-chromatic index* $\chi'_l(G)$ is the corresponding concept for edge colourings, and again it must be the case that $\chi'_l(G) \geq \chi'(G)$. It is conjectured (by whom first is not known) that $\chi'_l(G) = \chi'(G)$ for all graphs G . Galvin [31] proved the conjecture for bipartite graphs and Kahn [48] proved that $\chi'_l(G) = (1 + o(1))\chi'(G)$ as $\chi'(G) \rightarrow \infty$. The full conjecture is still open.

4.2. From Nash-Williams to acyclic colourings

Perhaps surprisingly, given their name, acyclic colourings did not evolve as generalizations of proper colourings; their history, which I shall outline in this section, is more closely related to the ‘acyclic’ part of the term. An acyclic graph is often called a *forest*, and I shall use the terms interchangeably.

Suppose that a graph G is the union of k forests $F_1 \cup \dots \cup F_k$. For all subgraphs $H \subset G$ and for all i , it must be the case that

$$|E(F_i \cap H)| \leq |V(H)| - 1,$$

as otherwise F_i would contain a cycle somewhere in H . Summing over the forests, it is therefore the case that

$$|E(H)| \leq k(|V(H)| - 1).$$

A fundamental theorem of Nash-Williams [65] (whose proof relies on his work in an earlier paper [64]) asserts that this necessary condition is also sufficient.

THEOREM 4.4. *Let G be a graph. Suppose that every subgraph $H \subset G$ satisfies*

$$|E(H)| \leq k(|V(H)| - 1), \tag{4.1}$$

for some integer k . Then G can be decomposed into at most k forests. □

The minimum value of k such that G can be decomposed into at most k forests, or equivalently such that G satisfies (4.1), is called the *arboricity* of G .

This concept was generalized to *point-arboricity* by Chartrand, Geller and Hedetniemi [19], and formally defined and further explored by Chartrand, Kronk and Wall [20]. The point-arboricity of a graph is the minimum number of colours needed to colour the

vertices (not necessarily properly) so that the graph induced by each colour class is a forest, that is, is acyclic.

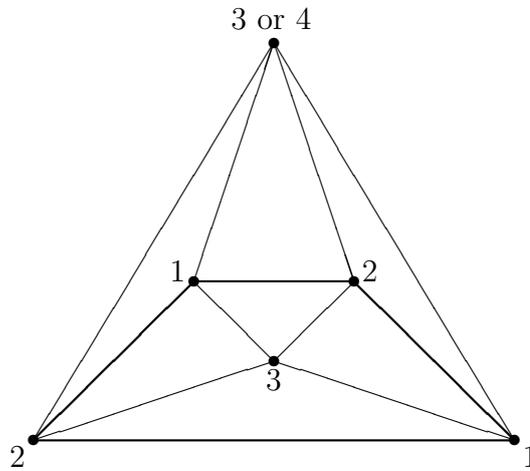


FIGURE 4.1. This graph can be 4-coloured in two different ways, shown on this figure up to relabelling of the colours. Neither of them is acyclic, since the vertices coloured 1 and 2 form a cycle.

Acyclic colourings were introduced as a natural next step by Grünbaum [44]. A proper vertex-colouring of a graph is *acyclic* if every graph induced by the union of *two* colour classes is acyclic, that is, there is no two-coloured cycle. The minimum number of colours needed to do this is the graph's *acyclic chromatic number*. Whereas the point-arboricity of a graph is clearly at most the chromatic number, since an independent set is certainly acyclic, the acyclic chromatic number must be at least as large. Grünbaum exhibited simple planar graphs with acyclic colouring number 5, for example the octahedron (see Figure 4.1), and conjectured that the acyclic chromatic number of all planar graphs is at most 5. Grünbaum himself proved that it was at most 9; this was improved to 8 by Mitchem [58], to 7 by Albertson and Berman [2] and to 6 by Kostochka [50]. Finally, the full conjecture was proved by Borodin [13].

4.3. Variants

In this section, I shall demonstrate how many traditional colouring problems, as described in Section 4.1, have been transferred to the acyclic arena. I shall also describe the natural extensions of these problems that I shall consider in this part of this dissertation. Kostochka [51] proved that it is an NP-complete problem to decide for given G and k

if the acyclic chromatic number of G is at most k ; thus, it makes sense to try to find sufficient conditions for it to be small.

Given their introduction, it is unsurprising that acyclic chromatic numbers were next bounded for graphs drawn on surfaces other than the plane. The *genus* $g(S)$ of a surface S is the maximum number of closed curves along which S can be cut without being disconnected or, informally, the number of handles that need to be added to a sphere to obtain S . The plane has genus 0 and, for any surface S , it is the case that $\chi(S) = 2 - 2g(S)$. A graph is said to be of *genus* g if the smallest genus of a surface on which it can be drawn without its edges crossing is g . Albertson and Berman [1] instigated this study and proved the following theorem [3].

THEOREM 4.5. *Any graph of genus $g > 0$ can be acyclically coloured with $4g + 4$ colours.*

When bounding the acyclic chromatic number, the shortest cycles are critical, since these are less likely to receive three colours in a proper colouring. The *girth* of a graph is the length of its shortest cycle. Borodin, Kostochka and Woodall [15] proved that planar graphs of girth at least 5 have acyclic chromatic number at most 4 and that those of girth at least 7 have acyclic chromatic number at most 3.

A bound for the acyclic chromatic number of a graph as a function of its maximum degree was found by Alon, McDiarmid and Reed [4]. In the following theorem, $A(d)$ is the maximum possible value of the acyclic chromatic number of a graph of maximum degree d .

THEOREM 4.6.

$$\Omega\left(\frac{d^{\frac{4}{3}}}{(\log d)^{\frac{1}{3}}}\right) \leq A(d) \leq O\left(d^{\frac{4}{3}}\right).$$

My results are generalizations of this one, so it is worth outlining its proof. The upper bound is proved by colouring a graph of maximum degree d randomly and then using the Erdős-Lovász local lemma. This essentially says that if we have a collection of bad events, each of which happens with small probability and is independent of most other bad events, then with positive probability none of them happens. I shall state this precisely in Section 4.5. For acyclic colourings, the most obvious bad events are,

- (1) for each edge, that the vertices that it connects are the same colour, and,
- (2) for each even cycle, that its vertices are coloured with alternating colours.

There is no need to consider odd cycles explicitly, since in a proper colouring they will receive at least three colours anyway.

These are not quite the events considered, for two reasons. Firstly, considering cycles of all lengths makes it harder to find a useful bound for the number of other bad events that each bad event might depend on. Secondly, if two vertices u and v are opposite vertices in many 4-cycles, then by avoiding bad events of the second type we effectively end up insisting that u and v are differently coloured too many times to get best possible bounds. The first problem is easily dealt with; 4-cycles and paths of length 4 are considered, since longer cycles contain a path of length 4. The second problem is overcome by calling u and v ‘special’ if they have many common neighbours and are therefore in many 4-cycles and, for each special pair u, v , introducing the bad event that u and v are the same colour. If these bad events do not occur then properly coloured 4-cycles containing a special pair receive at least three colours, and there remain only 4-cycles not containing a special pair. The final type of bad event is that such 4-cycles are coloured with alternating colours, and each vertex cannot be in too many of them without being in a special pair.

They proved that the upper bound is best possible up to a logarithmic factor by considering a random graph $G_{n,p}$ on n vertices with each edge present with probability p independently of the others. They showed that if $p \geq 3 \left(\frac{\log n}{n}\right)^{\frac{1}{4}}$ then, however the vertices are coloured with at most $\frac{n}{2}$ colours, the probability that every 4-cycle receives at least three colours is so small that the probability that *any* colouring has every 4-cycle receiving at least three colours is less than one. At the same time, for $p = 3 \left(\frac{\log n}{n}\right)^{\frac{1}{4}}$, with high probability the maximum degree of the graph d satisfies $d < 2pn = 6n^{\frac{3}{4}}(\log n)^{\frac{1}{4}}$, and therefore a graph exists with maximum degree d that cannot be acyclically coloured with fewer than $\frac{n}{2} = \Omega\left(\frac{d^{\frac{4}{3}}}{(\log d)^{\frac{1}{3}}}\right)$ colours.

In terms of precise evaluation of $A(d)$, it is obvious that $A(1) = 2$ and $A(2) = 3$. It is known that $A(3) = 4$ (mentioned without proof by Grünbaum [44]; see also Skulrattanakulchai [73]), $A(4) = 5$ (Burnstein [17]), $A(5) \leq 8$ and $A(6) \leq 12$ (Kothapalli, Varagani,

Venkaiah and Yadav [52, 53]). No lower bounds better than $d + 1$ are known for $d = 5, 6$, for which the example is the complete graph K_{d+1} .

In the same paper as earlier, Alon, McDiarmid and Reed [4] defined the edge acyclic chromatic number $A'(G)$ and its maximum $A'(d)$ for graphs of maximum degree d in the obvious way. They proved that $A'(d) \leq 64d$, without trying to optimize the constant. When quoting this result in a later paper, Molloy and Reed [59] showed that the same proof can easily be modified to give a bound of $16d$. Alon, Sudakov and Zaks [6] conjectured that in fact $A'(d) \leq d + 2$. They proved that $A'(G_{n,d}) \leq d + 2$ for almost all random d -regular graphs $G_{n,d}$ and showed that there exists $c > 0$ such that $A'(G) \leq d + 2$ for all graphs G with $\Delta(G) = d$ and girth at least $cd \log d$. They also commented that this conjecture would be best possible, by considering acyclic edge colourings of the complete graph K_{2n} . The largest that any one colour class can be is n , when it is a perfect matching, and no two colour classes can have more than $2n - 1$ edges without containing a cycle. This means that the largest colour class has size at most n and all the others at most $n - 1$, so there are at least $2n + 1$ colour classes. Nešetřil and Wormald [66] improved the bound for almost all random d -regular graphs to $d + 1$.

As with the vertex version, further improvements have been made for graphs of large girth. Muthu, Narayanan and Subramanian [63] improved the bound to $6d$ for graphs with girth at least 9 and to $4.52d$ for graphs of girth at least 220.

Gerke, Greenhill and Wormald [37] defined the r -acyclic edge chromatic number to be the minimum number of colours in a proper edge-colouring in which each cycle receives at least r colours. They showed that $(r - 2)d$ is asymptotically almost surely (as $d \rightarrow \infty$) a bound for the r -acyclic edge chromatic number of a random d -regular graph.

List colourings have natural acyclic analogues, which I shall not define formally as they are obvious. Borodin, Fon-Der Flaass, Kostochka, Raspaud and Sopena [14] proved that the list-acyclic chromatic index of a planar graph is at most 7.

When acyclic colourings are considered in isolation, rather than as a generalization of arboricity, there is no particular reason why the subgraphs chosen to be subject to extra conditions should be cycles. Again in the same paper as before, Alon, McDiarmid and Reed [4] discussed P_k -free colourings; these are proper vertex colourings such that

no path with k vertices is 2-coloured. They pointed out that, since there are at most $d + d(d - 1) = d^2$ vertices within distance 2 of any vertex, a greedy algorithm can be used to colour a graph in a P_3 -free fashion with at most $d^2 + 1$ colours. They also commented that a modification of their work in that paper proves that there is always a P_k -free colouring for $k \geq 4$ with at most $O\left(d^{\frac{k-1}{k-2}}\right)$ colours.

4.4. Generalized acyclic colourings

The problems that I shall consider in this part of this dissertation are extensions of Theorem 4.6, which gave a bound for $A(d)$, the maximum possible acyclic chromatic number of a graph of maximum degree d . There are many possibilities for generalizations, and these are just some of them. I shall introduce some more general versions of acyclic colourings, which will be formally defined in Section 4.5.

As described in Section 4.3, since it is the shortest cycles that are most likely to receive only two colours in a proper colouring, several authors have imposed a minimum girth condition on the underlying graph to be coloured. My approach is slightly different; I shall continue to allow all graphs of maximum degree d , but only require long cycles to receive at least three colours. Specifically, I shall introduce the *length- l -acyclic chromatic number*, where cycles of length at least l must be 3-coloured. Since an odd cycle is already 3-coloured in a proper colouring, it is only necessary to consider length- $2m$ -acyclic colourings for integers m .

Recall from Section 4.3 that the r -acyclic edge chromatic number was defined to be the minimum number of colours in a proper edge-colouring in which each cycle receives at least r colours. I shall do the same for the vertex-colouring version, but to avoid ambiguity with the length- l -acyclic chromatic number, I shall call it the *c -colour acyclic chromatic number*. This is a natural concept to consider; for point-arboricity the defining condition is that every cycle receives at least two colours, for an acyclic colouring it is that every cycle receives at least three colours. What happens if we demand that every cycle receives at least c colours? Point-arboricity corresponds to the 2-colour acyclic chromatic number and the usual acyclic chromatic number to the 3-colour acyclic chromatic number.

During my oral examination, I discovered that my bounds on the c -colour acyclic chromatic number (Section 5.3) had already been proved by Greenhill and Pikhurko [43], along with corresponding results for edge-colourings.

I also mentioned P_k -free colourings in Section 4.3, which are an example of a variation on the theme of acyclic colourings in which the subgraphs that must receive extra colours are not cycles. One way to think of the acyclic chromatic number as a generalization of the chromatic number is as follows. In a proper colouring, every 1-regular connected subgraph (edge) receives at least two colours. In an acyclic colouring, every 2-regular connected subgraph (cycle) receives at least three colours. The next step is to ask what happens if every r -regular connected subgraph must receive at least three colours.

However, this seems quite restrictive, as there are not likely to be many r -regular connected subgraphs. An alternative, related approach is to require subgraphs with minimum degree at least r to receive at least three colours. This is still a natural thing to do: a proper colouring is one in which every subgraph with minimum degree at least 1 receives at least two colours, and an acyclic colouring is one in which every subgraph with minimum degree at least 2 receives at least three colours. I shall define the *degree- r chromatic number* of a graph to be the minimum number of colours needed to colour the graph properly in such a way that this happens. The methods that I shall use work just as well for r -regular subgraphs; I have chosen to proceed in this way only because it seems more interesting.

This definition can be easily extended to hypergraphs; here, a *degree- r colouring* of a u -uniform hypergraph is a colouring such that every edge is multicoloured and every subhypergraph with minimum degree at least r receives at least $u + 1$ colours. This is the first case to consider since every subhypergraph with at least one edge already receives at least u colours in a multicolouring.

4.5. Definitions and a useful tool

In the next two chapters, I shall bound the following quantities for (hyper)graphs with maximum degree d . These bounds will be asymptotically best possible up to constant or logarithmic factors. In Chapter 5, I shall bound the two that concern cycles in graphs: the length- l -acyclic chromatic number and the c -colour chromatic number. In Chapter 6,

I shall bound the degree- r chromatic number of a u -uniform hypergraph. This is best separated into two cases: where $u = 2$, that is, for simple graphs, and where $u \geq 3$.

Note that the maxima in the following definitions all exist; trivial upper bounds are $d^2 + 1$, $d^{c-1} + 1$ and $((u-1)d)^2 + 1$ colours, sufficient to colour vertices differently from those at distance at most 2, $c-1$ and 2 away.

The length- l -acyclic chromatic number of a graph. A colouring $f : V(G) \rightarrow [x]$ is a length- l -acyclic colouring if it is a proper colouring and every cycle of length at least l receives at least three colours. For a graph G and an integer l , let

$$A^{(l)}(G) = \min\{x : G \text{ can be length-}l\text{-acyclically coloured using } x \text{ colours}\}.$$

For integers l, d , let

$$A^{(l)}(d) = \max\{A^{(l)}(G) : \Delta(G) = d\}.$$

Note that a length- l -acyclic colouring is also a length- $(l+1)$ -acyclic colouring, so $A^{(l+1)}(G) \leq A^{(l)}(G)$ and $A^{(l+1)}(d) \leq A^{(l)}(d)$. Also, since a cycle of odd length must receive three colours in a proper colouring, it must be the case that $A^{(2m-1)}(G) = A^{(2m)}(G)$ and $A^{(2m-1)}(d) = A^{(2m)}(d)$ for any integer m .

The c -colour acyclic chromatic number of a graph. A colouring $f : V(G) \rightarrow [x]$ is a c -colour acyclic colouring if it is a proper colouring, every cycle of length less than c is multicoloured and every cycle of length at least c receives at least c colours. For a graph G and an integer c , let

$$A_c(G) = \min\{x : G \text{ can be } c\text{-acyclically coloured using } x \text{ colours}\}.$$

For integers c, d , let

$$A_c(d) = \max\{A_c(G) : \Delta(G) = d\}.$$

Note that a $(c+1)$ -colour acyclic colouring is also a c -colour acyclic colouring, so $A_c(G) \leq A_{c+1}(G)$ and $A_c(d) \leq A_{c+1}(d)$.

The degree- r chromatic number of a u -uniform hypergraph. Let H be a u -uniform hypergraph. A colouring $f : V(H) \rightarrow [x]$ is a degree- r colouring if every edge is multicoloured and every subhypergraph K with $\delta(K) \geq r$ receives at least $u+1$ colours.

For a u -uniform hypergraph H and an integer r , let

$$D_r^{(u)}(H) = \min\{x : H \text{ can be degree-}r \text{ coloured using } x \text{ colours}\}.$$

For integers r, d , let

$$D_r^{(u)}(d) = \max\{D_r^{(u)}(H) : \Delta(H) = d \text{ and } H \text{ is a } u\text{-uniform hypergraph}\}.$$

I shall bound $D_r^{(u)}$ separately when $u = 2$ and when $u \geq 3$.

It is not easy to guess what the upper bounds should be, even knowing that they should include Theorem 4.6. In fact, in the case of c -colour acyclic colourings, this theorem is in fact an exceptional case and so unhelpful in providing intuition.

As with most results in this area, the proofs of these bounds will all use random colourings and the Erdős-Lovász local lemma [25], in its nonsymmetric form (see page 64 of Alon and Spencer [5]).

LEMMA 4.7. *Let A_1, \dots, A_n be events in an arbitrary probability space. Let the graph $G = (V, E)$ with $V = [n]$ be a dependency graph for the events A_i , that is, assume that for each i , A_i is independent of the family of events $\{A_j : ij \notin E\}$. If there are reals $0 < y_i < 1$ such that for all i*

$$\mathbb{P}(A_i) \leq y_i \prod_{ij \in E} (1 - y_j),$$

then

$$\mathbb{P}\left(\bigcap_i A_i^c\right) \geq \prod_{i=1}^n (1 - y_i) > 0,$$

so that with positive probability no event A_i occurs.

The events A_i are commonly called *bad* events.

To prove that the upper bounds that I shall give are asymptotically best possible up to constant or logarithmic factors, I shall give explicit or probabilistic constructions. In the probabilistic constructions, the number of colours needed will be proportional to the number of vertices, so I shall need to bound this in terms of the maximum degree of a random graph. Bollobás [12] (page 65, Corollary 3.4) proved the following precise result

about the maximum degree Δ of a random graph $G_{n,p}$ with n vertices and with each edge present with probability $p = p(n)$ independent of the others, and with $q = 1 - p$.

THEOREM 4.8. *Suppose $\frac{pqn}{(\log n)^3} \rightarrow \infty$ and y is a fixed real number. Then the maximal degree Δ of $G_{n,p}$ satisfies*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\Delta < pn + (2pqn \log n)^{\frac{1}{2}} \left(1 - \frac{\log \log n}{4 \log n} - \frac{\log(2\pi^{\frac{1}{2}})}{2 \log n} + \frac{y}{2 \log n} \right) \right] = e^{-e^{-y}}.$$

I shall use only the trivial consequence that, under these conditions, asymptotically almost surely the maximum degree satisfies $\Delta < 2pn$. In fact, I shall only ever use probabilities p of a particular form, and the lemma that I shall apply is as follows.

LEMMA 4.9. *Let $\alpha \geq 0$, $0 < \beta < 1$ and $C \geq \frac{1}{2}$ be real numbers, and let*

$$p = C \frac{(\log n)^\alpha}{n^\beta}.$$

Then asymptotically almost surely the maximal degree Δ of $G_{n,p}$ satisfies

$$n > \frac{\left(\frac{\Delta}{2C}\right)^{\frac{1}{1-\beta}}}{\left(\frac{1}{1-\beta} \log \Delta\right)^{\frac{\alpha}{1-\beta}}}. \quad (4.2)$$

PROOF. The bounds on β mean that Theorem 4.8 can be applied, and so asymptotically almost surely

$$\Delta < 2Cn^{1-\beta}(\log n)^\alpha,$$

from which it follows that

$$n > \frac{\left(\frac{\Delta}{2C}\right)^{\frac{1}{1-\beta}}}{(\log n)^{\frac{\alpha}{1-\beta}}}. \quad (4.3)$$

Either the bound in (4.2) holds, in which case there is nothing to prove, or it does not hold, in which case

$$\log n \leq \frac{1}{1-\beta} \log \Delta - \frac{\alpha}{1-\beta} \log \log \Delta - \frac{1}{1-\beta} \log \left(\frac{2C}{(1-\beta)^\alpha} \right) \leq \frac{1}{1-\beta} \log \Delta,$$

and (4.2) follows from (4.3). □

CHAPTER 5

Long acyclic colourings and acyclic colourings with many colours

5.1. Introduction

I have grouped the two results in this chapter together as they both involve generalizations of the acyclic chromatic number in a way that still relates to cycles in graphs, which is not the case with the results in the next chapter. Alon, McDiarmid and Reed [4] showed that a graph with maximum degree d can be acyclically coloured with $O(d^{\frac{4}{3}})$ colours. In this chapter, I shall answer the following questions:

- (1) How many colours do we need if only cycles of length at least l must receive at least three colours (Section 5.2)?
- (2) How many colours do we need if cycles must receive at least c colours (Section 5.3)?

In Section 5.4, I shall suggest some natural open problems following on from these.

5.2. The length- l -acyclic chromatic number

Recall that the length- l -acyclic chromatic number is the minimum number of colours needed to colour a graph properly so that each cycle of length at least l receives at least three colours. Recall also that, in a proper colouring, cycles of odd length will automatically receive at least three colours. In this section, I shall bound the maximum possible length- l -acyclic chromatic number of graphs of a given maximum degree d , denoted by $A^{(l)}(d)$.

THEOREM 5.1. *For all integers $d, m \geq 2$, it is the case that*

$$\frac{\left(\frac{d}{6}\right)^{\frac{2m}{2m-1}}}{m \left(\frac{2m}{2m-1} \log d\right)^{\frac{1}{2m-1}}} < A^{(2m)}(d) = A^{(2m-1)}(d) < 14(2m+1)d^{\frac{2m}{2m-1}}.$$

This has the following immediate corollary.

COROLLARY 5.2. *For a fixed integer $m \geq 2$, it is the case that*

$$\Omega \left(\frac{d^{\frac{2m}{2m-1}}}{(\log d)^{\frac{1}{2m-1}}} \right) \leq A^{(2m)}(d) = A^{(2m-1)}(d) \leq O \left(d^{\frac{2m}{2m-1}} \right)$$

as $d \rightarrow \infty$. □

I shall begin by proving the lower bound.

THEOREM 5.3. *For n sufficiently large there is a graph G on n vertices with maximum degree $\Delta(G) = d$ satisfying*

$$\frac{n}{m} \geq \frac{\left(\frac{d}{6}\right)^{\frac{2m}{2m-1}}}{m \left(\frac{2m}{2m-1} \log d\right)^{\frac{1}{2m-1}}} = \Omega \left(\frac{d^{\frac{2m}{2m-1}}}{(\log d)^{\frac{1}{2m-1}}} \right)$$

such that there is no length- $2m$ -acyclic colouring with at most $\frac{n}{m}$ colours.

PROOF. This proof is a generalization of one by Alon, McDiarmid and Reed [4].

Let

$$p = 3 \left(\frac{\log n}{n} \right)^{\frac{1}{2m}},$$

let n be large and divisible by m^2 , and let $G \in \mathcal{G}_{n,p}$.

Suppose that $f : V(G) \rightarrow \left[\frac{n}{m} \right]$ is a colouring that uses at most $\frac{n}{m}$ colours. From each class, discard as few vertices as possible to make its size divisible by m . Note that at most $m - 1$ vertices are discarded from each class, so at most $\frac{(m-1)n}{m}$ vertices are discarded in total and there remain at least $\frac{n}{m}$ vertices. Since the size of each colour class is now divisible by m , at least $\frac{n}{m^2}$ distinct monochromatic m -tuples can be taken from these vertices. Choose $\frac{n}{m^2}$ of these and call them $T_1, \dots, T_{\frac{n}{m^2}}$.

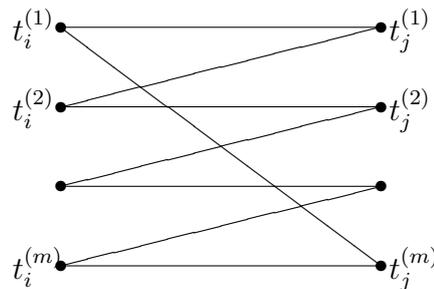


FIGURE 5.1. A $2m$ -cycle of the form $t_i^{(1)}t_j^{(1)}t_i^{(2)}t_j^{(2)} \dots t_i^{(k)}t_j^{(k)}$.

For each m -tuple T_i , fix an ordering of the vertices $t_i^{(1)}, \dots, t_i^{(m)}$. Let $C_{i,j}$ be the event that G contains the cycle $t_i^{(1)}t_j^{(1)}t_i^{(2)}t_j^{(2)} \dots t_i^{(m)}t_j^{(m)}$ (see Figure 5.1). If any event $C_{i,j}$ occurs then this colouring cannot be a length- $2m$ -acyclic colouring of G . Let A_f be the event that f is a length- $2m$ -acyclic colouring. Then

$$\begin{aligned} \mathbb{P}(A_f) &\leq \mathbb{P}\left(\bigcap_{i < j} C_{i,j}^c\right) \\ &= (1 - p^{2m})^{\binom{n/m^2}{2}}, \end{aligned}$$

since the events $C_{i,j}$ are independent. Writing E for the event that G can be length- $2m$ -acyclically coloured with at most $\frac{n}{m}$ colours,

$$\begin{aligned} \mathbb{P}(E) &\leq \sum_{f: V(G) \rightarrow \left[\frac{n}{m}\right]} \mathbb{P}(A_f) \\ &\leq n^n (1 - p^{2m})^{\binom{n/m^2}{2}} \\ &\leq \exp\left(n \log n - \binom{n/m^2}{2} p^{2m}\right) \\ &= o(1), \end{aligned}$$

which means that there exists a graph G that cannot be length- $2m$ -acyclically coloured with at most $\frac{n}{m}$ colours, as required. By Lemma 4.9, the graph G can be chosen so that, in addition, the bound in the theorem is satisfied. \square

I shall now prove the upper bound.

THEOREM 5.4. *Let G be a graph with $\Delta(G) = d$ and let $m \geq 2$ be an integer. Then it is possible to colour G length- $2m$ -acyclically with at most $14(2m+1)d^{\frac{2m}{2m-1}} = O(d^{\frac{2m}{2m-1}})$ colours.*

PROOF. In this and the next chapter, there will be several proofs that use the Erdős-Lovász local lemma. I shall apply the method in more detail in this proof than in later ones.

Call a pair of vertices $\{u, v\}$ *special* if u and v have at least $d^{\frac{2m-2}{2m-1}}$ common neighbours.

Let $f : V(G) \rightarrow [x]$ be a random function where each vertex receives a colour from $[x]$ uniformly at random.

The proof uses the Erdős-Lovász local lemma. The bad events to be considered are as follows:

- (1) For each edge uv , let $W_{uv} = \{f(u) = f(v)\}$.
- (2) For each special pair $\{u, v\}$, let $X_{u,v} = \{f(u) = f(v)\}$.
- (3) For each $2m$ -cycle $C = v_0v_1 \dots v_{2m-1}$ containing no special pairs of vertices distance 2 apart, where $v_0, v_1, \dots, v_{2m-1}$ are distinct vertices, let

$$Y_C = \{f(v_0) = f(v_2) = \dots = f(v_{2m-2}) \text{ and } f(v_1) = f(v_3) = \dots = f(v_{2m-1})\}.$$

- (4) For each path $P = v_0v_1 \dots v_{2m}$ of length $2m$, where v_0, v_1, \dots, v_{2m} are distinct vertices, let

$$Z_P = \{f(v_0) = f(v_2) = \dots = f(v_{2m}) \text{ and } f(v_1) = f(v_3) = \dots = f(v_{2m-1})\}.$$

If none of the events W_{uv} occurs then f is a proper colouring. The only way that a cycle of even length can be 2-coloured in a proper colouring is if its colours alternate, and so if in addition none of the events $X_{u,v}$ and Y_C occurs then every cycle of length $2m$ receives at least three colours whether or not it contains special pairs distance 2 apart. If it is also the case that none of the events Z_P occurs then every cycle of length at least $2m + 2$ receives at least three colours.

CLAIM 1. *Each vertex v is in at most*

- (1) d edges,
- (2) $d^{\frac{2m}{2m-1}}$ special pairs,
- (3) $d^{\frac{2m(2m-2)}{2m-1}}$ cycles of length $2m$ not containing any special pairs of vertices distance 2 apart, and
- (4) $(m + 1)d^{2m}$ paths of length $2m$.

PROOF OF CLAIM. (1) By supposition, the maximum degree $\Delta(G) = d$.

- (2) Since $\Delta(G) = d$, the vertex v has at most d neighbours, each of which has at most d neighbours, and so there are at most d^2 paths of length 2 starting at v . By definition, for each special pair that v is in there are at least $d^{\frac{2m-2}{2m-1}}$ paths of

length 2 starting at v and so v is in at most

$$\frac{d^2}{d^{\frac{2m-2}{2m-1}}} = d^{\frac{4m-2-(2m-2)}{2m-1}} = d^{\frac{2m}{2m-1}}$$

special pairs.

- (3) Without loss of generality, assume that v plays the role of v_0 in the cycle C . Since $\Delta(G) = d$, there are at most d choices for v_1 , for each of these there are most d choices for v_2 and so on, and so there at most d^{2m-2} choices for the path $v_0 \dots v_{2m-2}$. Since $\{v_0, v_{2m-2}\}$ is not a special pair, there are at most $d^{\frac{2m-2}{2m-1}}$ choices for their common neighbour v_{2m-1} and so in total v is in at most

$$d^{2m-2+\frac{2m-2}{2m-1}} = d^{\frac{2m(2m-2)}{2m-1}}$$

cycles of length $2m$ not containing any special pairs of vertices distance 2 apart.

- (4) Without loss of generality, assume that v plays the role of one of v_0, v_1, \dots, v_m in $P = v_0 v_1 \dots v_{2m}$, which gives $m + 1$ possibilities. For each of these there are at most d^{2m} paths P , which means that v is in at most $(m + 1)d^{2m}$ paths of length $2m$.

□

By multiplying the number of vertices in a pair, cycle or path by the number of edges, special pairs, cycles without special pairs and paths that each vertex can be in, the number of edges in the dependency graph from an event of each type to events of each type is at most that shown in the table below.

		to			
		$W_{u,v}$	$X_{u,v}$	Y_C	Z_P
from	$W_{u,v}$	$2d$	$d^{\frac{2m}{2m-1}}$	$2d^{\frac{2m(2m-2)}{2m-1}}$	$2(m+1)d^{2m}$
	$X_{u,v}$	$2d$	$2d^{\frac{2m}{2m-1}}$	$2d^{\frac{2m(2m-2)}{2m-1}}$	$2(m+1)d^{2m}$
	Y_C	$2md$	$2md^{\frac{2m}{2m-1}}$	$2md^{\frac{2m(2m-2)}{2m-1}}$	$2m(m+1)d^{2m}$
	Z_P	$(2m+1)d$	$(2m+1)d^{\frac{2m}{2m-1}}$	$(2m+1)d^{\frac{2m(2m-2)}{2m-1}}$	$(2m+1)(m+1)d^{2m}$

CLAIM 2. *The probabilities of the events of each type are given by*

- (1) $\mathbb{P}(W_{u,v}) = \frac{1}{x}$,
- (2) $\mathbb{P}(X_{u,v}) = \frac{1}{x}$,

$$(3) \mathbb{P}(Y_C) = \frac{1}{x^{2m-2}}, \text{ and}$$

$$(4) \mathbb{P}(Z_P) = \frac{1}{x^{2m-1}}.$$

PROOF OF CLAIM. All four statements follow from the fact that f colours vertices uniformly at random with colours taken from $[x]$. \square

I shall now apply the Erdős-Lovász local lemma (Lemma 4.7). The weightings used are double the probabilities, that is, let

$$\begin{aligned} y_1 &= 2\mathbb{P}(W_{u,v}) = \frac{2}{x}, \\ y_2 &= 2\mathbb{P}(X_{u,v}) = \frac{2}{x}, \\ y_3 &= 2\mathbb{P}(Y_C) = \frac{2}{x^{2m-2}} \quad \text{and} \\ y_4 &= 2\mathbb{P}(Z_P) = \frac{2}{x^{2m-1}}, \end{aligned}$$

and the four statements that need to be proved are

$$\begin{aligned} \mathbb{P}(W_{u,v}) &\leq 2\mathbb{P}(W_{u,v}) \left(1 - \frac{2}{x}\right)^{2d} \left(1 - \frac{2}{x}\right)^{2d \frac{2m}{2m-1}} \\ &\quad \left(1 - \frac{2}{x^{2m-2}}\right)^{2d \frac{2m(2m-2)}{2m-1}} \left(1 - \frac{2}{x^{2m-1}}\right)^{2(m+1)d^{2m}}, \end{aligned}$$

$$\begin{aligned} \mathbb{P}(X_{u,v}) &\leq 2\mathbb{P}(X_{u,v}) \left(1 - \frac{2}{x}\right)^{2d} \left(1 - \frac{2}{x}\right)^{2d \frac{2m}{2m-1}} \\ &\quad \left(1 - \frac{2}{x^{2m-2}}\right)^{2d \frac{2m(2m-2)}{2m-1}} \left(1 - \frac{2}{x^{2m-1}}\right)^{2(m+1)d^{2m}}, \end{aligned}$$

$$\begin{aligned} \mathbb{P}(Y_C) &\leq 2\mathbb{P}(Y_C) \left(1 - \frac{2}{x}\right)^{2md} \left(1 - \frac{2}{x}\right)^{2md \frac{2m}{2m-1}} \\ &\quad \left(1 - \frac{2}{x^{2m-2}}\right)^{2md \frac{2m(2m-2)}{2m-1}} \left(1 - \frac{2}{x^{2m-1}}\right)^{2m(m+1)d^{2m}}, \quad \text{and} \end{aligned}$$

$$\begin{aligned} \mathbb{P}(Z_P) &\leq 2\mathbb{P}(Z_P) \left(1 - \frac{2}{x}\right)^{(2m+1)d} \left(1 - \frac{2}{x}\right)^{(2m+1)d \frac{2m}{2m-1}} \\ &\quad \left(1 - \frac{2}{x^{2m-2}}\right)^{(2m+1)d \frac{2m(2m-2)}{2m-1}} \left(1 - \frac{2}{x^{2m-1}}\right)^{(2m+1)(m+1)d^{2m}}. \end{aligned}$$

Since $m \geq 2$, it suffices to show the fourth statement, that is,

$$\left(1 - \frac{2}{x}\right)^{(2m+1)d} \left(1 - \frac{2}{x}\right)^{(2m+1)d^{\frac{2m}{2m-1}}} \left(1 - \frac{2}{x^{2m-2}}\right)^{(2m+1)d^{\frac{2m(2m-2)}{2m-1}}} \left(1 - \frac{2}{x^{2m-1}}\right)^{(2m+1)(m+1)d^{2m}} \geq \frac{1}{2}.$$

It suffices to choose x such that each factor is at least $\frac{6}{7}$, since $(\frac{6}{7})^4 = \frac{1296}{2401} > \frac{1}{2}$, and it therefore suffices to choose x such that

$$x \geq \max \left\{ 14(2m+1)d, 14(2m+1)d^{\frac{2m}{2m-1}}, 2^{m-2} \sqrt{14(2m+1)d^{\frac{2m}{2m-1}}}, 2^{m-1} \sqrt{14(2m+1)(m+1)d^{\frac{2m}{2m-1}}} \right\},$$

for which it suffices to take $x = 14(2m+1)d^{\frac{2m}{2m-1}}$. \square

5.3. The c -colour acyclic chromatic number

The results in this section had already been proved by Greenhill and Pikhurko [43], which I did not know until my oral examination.

Recall that the c -colour acyclic chromatic number is the minimum number of colours needed to colour a graph properly so that each cycle receives at least c colours. In this section, I shall bound the maximum possible c -colour acyclic chromatic number of graphs of a given maximum degree d , denoted by $A_c(d)$.

THEOREM 5.5. *For all integers $d, k \geq 2$, it is the case that*

$$\max \left\{ \left(\frac{d}{k}\right)^k, \frac{\left(\frac{d}{4}\right)^k}{2k \log d} \right\} < A_{2k}(d) \leq A_{2k+1}(d) < 6(k+2)(2k)^{k+1} d^k.$$

This has the following immediate corollary.

COROLLARY 5.6. *For a fixed integer $k \geq 2$, it is the case that*

$$A_{2k}(d), A_{2k+1}(d) = \Theta(d^k)$$

as $d \rightarrow \infty$. \square

It is interesting that this bound does not include the bound on the 3-colour acyclic chromatic number of Alon, McDiarmid and Reed [4]. It will become clearer why this is the case once Theorem 5.11 has been proved; I shall say more at the time.

For $c = 4$, a probabilistic lower bound can be found in a similar way to Theorem 5.3.

THEOREM 5.7. *For n sufficiently large there is a graph G on n vertices with maximum degree $\Delta(G) = d$ satisfying*

$$\frac{n}{2} \geq \frac{d^2}{400 \log d}$$

such that there is no 4-colour acyclic colouring with at most $\frac{n}{2}$ colours.

PROOF. Let

$$p = 5\sqrt{\frac{\log n}{n}},$$

let n be large and divisible by 4, and let $G \in \mathcal{G}_{n,p}$.

Suppose that $f : V(G) \rightarrow [\frac{n}{2}]$ is a colouring that uses at most $\frac{n}{2}$ colours. By discarding at most one vertex from each colour class, at least $\frac{n}{4}$ monochromatic pairs can be found. Let P be a set of $\frac{n}{4}$ of these pairs and split the remaining singletons into two sets S_1 and S_2 each of size $\frac{n}{4}$.

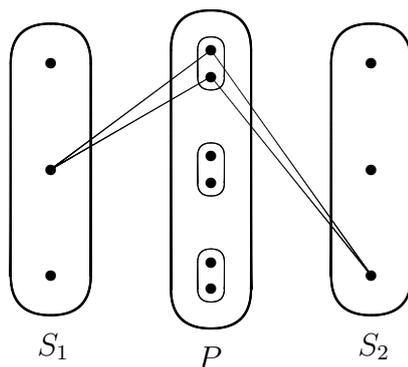


FIGURE 5.2. The type of 4-cycle that is avoided.

If any pair in P has a common neighbour in S_1 and a common neighbour in S_2 then f cannot be 4-colour acyclic (see Figure 5.2). For each pair $p_i \in P$ let $N_{p_i,j}$ be the event that the pair p_i has a common neighbour in S_j and let A_f be the event that f is a 4-colour

acyclic colouring. Then

$$\begin{aligned} \mathbb{P}(A_f) &\leq \mathbb{P}\left(\bigcap_{p_i \in P} (N_{p_i,1}^c \cup N_{p_i,2}^c)\right) \\ &\leq \left(2(1-p^2)^{\frac{n}{4}}\right)^{\frac{n}{4}}, \end{aligned}$$

since the events $N_{p_i,j}$ are independent. Writing E for the event that G can be 4-colour acyclically coloured with at most $\frac{n}{2}$ colours,

$$\begin{aligned} \mathbb{P}(E) &\leq \sum_{f:V(G) \rightarrow [\frac{n}{2}]} \mathbb{P}(A_f) \\ &\leq n^n \left(2(1-p^2)^{\frac{n}{4}}\right)^{\frac{n}{4}} \\ &\leq \exp\left(n \log n + \frac{n}{4} \log 2 - \frac{p^2 n^2}{16}\right) \\ &= o(1), \end{aligned}$$

which means that there exists a graph G that cannot be 4-colour acyclically coloured with at most $\frac{n}{2}$ colours, as required. By Lemma 4.9, the graph G can be chosen so that, in addition, the bound in the theorem is satisfied. \square

For $c > 4$, this does not seem easy to generalize. A sensible alternative approach to finding a lower bound is to find a graph of small maximum degree such that any two vertices are in a cycle of length at most c , so that all vertices must be coloured differently. One way of doing this is to join two graphs of small diameter together. The most obvious graph with small diameter for a given maximum degree is a tree.

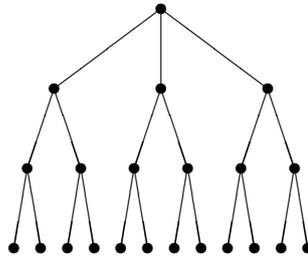


FIGURE 5.3. The tree $T_{d,h}$ for $d = h = 3$.

Let $T_{d,h}$ be the tree where all vertices except the leaves have degree d and the leaves are all distance h from a fixed root vertex (see Figure 5.3). Let T_1 and T_2 be copies of

$T_{d, \lfloor \frac{k-1}{2} \rfloor}$ and form G by taking T_1 and T_2 and adding edges between vertices $v_1 \in T_1$ and $v_2 \in T_2$ if they are copies of the same vertex $v \in T_{d, \lfloor \frac{k-1}{2} \rfloor}$. Then the number of vertices in G is

$$\begin{aligned} 2 \left(1 + d + d(d-1) + \dots + d(d-1)^{\lfloor \frac{k-1}{2} \rfloor - 1} \right) &= 2 + 2d \left(\frac{(d-1)^{\lfloor \frac{k-1}{2} \rfloor} - 1}{d-2} \right) \\ &> \left(\frac{d}{2} \right)^{\lfloor \frac{k-1}{2} \rfloor} = \Omega \left(d^{\lfloor \frac{k-1}{2} \rfloor} \right), \end{aligned}$$

and any two vertices of G are in a cycle of length at most $2k$. Therefore, G cannot be $2k$ -acyclically coloured (or $(2k+1)$ -acyclically coloured) with less than $\Omega \left(d^{\lfloor \frac{k-1}{2} \rfloor} \right)$ colours.

This is a long way from the lower bound of $\Omega \left(\frac{d^k}{\log d} \right)$. Using probabilistic techniques, Bollobás [10] (see also page 259 of Bollobás [12], Theorem 10.10) found graphs of small diameter with much smaller maximum degree:

THEOREM 5.8. *Let h be a positive constant, $k = k(n) \geq 2$ a natural number, and define $p = p(n, h, k)$, $0 < p < 1$, by*

$$p^k n^{k-1} = \log \left(\frac{n^2}{h} \right).$$

Suppose that $\frac{pn}{(\log n)^3} \rightarrow \infty$. Then for $G \in \mathcal{G}_{n,p}$ we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{diam}(G) = k) = e^{-\frac{h}{2}} \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbb{P}(\text{diam}(G) = k + 1) = 1 - e^{-\frac{h}{2}}.$$

□

Using this, I can obtain a lower bound that is best possible up to logarithmic factors.

THEOREM 5.9. *For n sufficiently large there is a graph G on n vertices with maximum degree $\Delta(G) = d$ satisfying*

$$n \geq \frac{d^k}{2^{2k+1} k \log d} = \Omega \left(\frac{d^k}{\log d} \right)$$

such that any two vertices are in a cycle of length at most $2k$. In particular, a $2k$ - or $(2k+1)$ -colour acyclic colouring of such a graph must use at least n colours.

PROOF. Set

$$p = \left(\frac{2 \log n}{n^{k-1}} \right)^{\frac{1}{k}},$$

which corresponds to $h = 1$ in Theorem 5.8, and $q = 2p$. Let $G_{n,q} \in \mathcal{G}_{n,q}$ and obtain G_1 and G_2 by putting each edge of $G_{n,q}$ in G_1 or G_2 uniformly at random independently of all other edges. Since $q = 2p$, this coupling means that $G_1 \in \mathcal{G}_{n,p}$ and $G_2 \in \mathcal{G}_{n,p}$, but G_1 and G_2 are disjoint. By Theorem 5.8,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\text{diam}(G_1) = k) = \lim_{n \rightarrow \infty} \mathbb{P}(\text{diam}(G_2) = k) = e^{-\frac{1}{2}} \approx 0.6065.$$

Take n sufficiently large that $\mathbb{P}(\text{diam}(G_1) = k) = \mathbb{P}(\text{diam}(G_2) = k) > \frac{3}{5}$ and so

$$\mathbb{P}\left(\left(\text{diam}(G_1) = k\right) \wedge \left(\text{diam}(G_2) = k\right)\right) > \frac{1}{5}.$$

Since this probability is greater than zero, this means that there is a graph G with any two vertices in a cycle of length at most $2k$. By Lemma 4.9, the graph G can be chosen so that, in addition, the bound in the theorem is satisfied. \square

In fact, an explicit construction does even better and gives the best possible lower bound up to a constant factor.

THEOREM 5.10. *For n sufficiently large there is a graph G on n vertices with maximum degree $\Delta(G) = d$ satisfying*

$$n > \left(\frac{d}{k} \right)^k = \Omega(d^k),$$

such that any two non-adjacent vertices are in a cycle of length at most $2k$ (and so a $2k$ - or $(2k + 1)$ -colour acyclic colouring must use at least n colours).

PROOF. Let $G = (V, E)$ be defined as follows. For some large positive integer b , let $V = [b]^k$ and let $uv \in E$ if u and v differ in exactly one co-ordinate. Then G is a $k(b-1)$ -regular graph. For all non-adjacent u, v , there is a cycle of length at most $2k$ containing u and v as follows. Start with u and then, working from left to right, for each digit u_i where $u_i \neq v_i$, change it to v_i , and then do the same to get back to u from v (see Figure 5.4). All vertices are therefore adjacent or in a cycle of length at most $2k$, and must be differently coloured in a $2k$ - or $(2k + 1)$ -colour acyclic colouring.

$(1, 2, 3, \mathbf{2}, 1)$
 $(\mathbf{3}, 2, 3, 2, 1)$
 $(3, 2, \mathbf{2}, 2, 1)$
 $(3, 2, 2, \mathbf{3}, 1)$
 $(\mathbf{1}, 2, 2, 3, 1)$
 $(1, 2, \mathbf{3}, 3, 1)$

FIGURE 5.4. A 6-cycle for $b = 3$, $k = 5$, $u = (1, 2, 3, 2, 1)$ and $v = (3, 2, 2, 3, 1)$, with the new digit in each k -tuple in bold type.

Setting $d = k(b - 1)$, this is a graph G with $n = |V(G)|$ and $\Delta(G) = d$ satisfying

$$n = \left(\frac{d}{k} + 1 \right)^k,$$

and the result follows. □

For fixed k , the probabilistic bound is worse as $d \rightarrow \infty$, but it is better for small d , that is, for d up to about

$$\exp \left(\frac{\binom{k}{4}^{k-1}}{8} \right).$$

To prove an upper bound of the same order of magnitude as the explicit construction, I shall again use the Erdős-Lovász local lemma.

THEOREM 5.11. *Let G be a graph with $\Delta(G) = d$ and let $k \geq 2$ be an integer. Then it is possible to colour G $(2k + 1)$ -colour acyclically (and therefore $2k$ -colour acyclically) with at most $6(k + 2)(2k)^{k+1}d^k = O(d^k)$ colours.*

PROOF. If $d = 1$, then G contains only isolated edges and only two colours are needed, so assume that $d \geq 2$.

Call a pair of vertices $\{u, v\}$ *special* if there are at least d paths of length $k + 1$ from u to v .

Let $f : V(G) \rightarrow [x]$ be a random function where each vertex receives a colour from $[x]$ uniformly at random.

The proof uses the Erdős-Lovász local lemma. The bad events to be considered are as follows:

- (1) For each u, v with $1 \leq d(u, v) \leq k$, let $W_{u,v} = \{f(u) = f(v)\}$.

- (2) For each special pair u, v , let $X_{u,v} = \{f(u) = f(v)\}$.
- (3) For each $(2k+2)$ -cycle $C = v_0v_1 \dots v_{2k+1}$ containing a non-special pair of opposite vertices, where $v_0, v_1, \dots, v_{2k+1}$ are distinct vertices, let Y_C be the event that C receives at most $2k$ colours.
- (4) For each path $P = v_0v_1 \dots v_{2k+2}$, where $v_0, v_1, \dots, v_{2k+2}$ are distinct vertices, let Z_P be the event that P receives at most $2k$ colours.

If none of the events of type $W_{u,v}$ occurs then f is a proper colouring and every cycle of length at most $2k+1$ is multicoloured. In this case, the only vertices in a $(2k+2)$ -cycle that can be the same colour are pairs of opposite vertices, so if in addition none of the events $X_{u,v}$ and Y_C occurs then every cycle of length $2k+2$ receives at least $2k+1$ colours, whether or not it contains a non-special pair of opposite vertices. If none of the events Z_P occurs then every cycle of length at least $2k+3$ receives at least $2k+1$ colours.

CLAIM 1. *Each vertex v is in at most*

- (1) $2d^k$ pairs of vertices distance at most k apart,
- (2) d^k special pairs,
- (3) $(k+1)d^{k+2}$ cycles containing a non-special pair of opposite vertices, and
- (4) $(k+2)d^{2k+2}$ paths of length $2k+2$.

PROOF OF CLAIM. (1) By the same argument as in the proof of Theorem 5.4, there are at most d vertices distance 1 from v , at most d^2 vertices distance 2 from v and so on, and so at most

$$d + d^2 + \dots + d^k = \frac{d}{d-1}(d^k - 1) < 2d^k$$

vertices distance at most k from v , the inequality holding since $d \geq 2$.

- (2) There are at most d^{k+1} paths of length $k+1$ starting at v , and each special pair that v is in contributes d such paths, so v is in at most

$$\frac{d^{k+1}}{d} = d^k$$

special pairs.

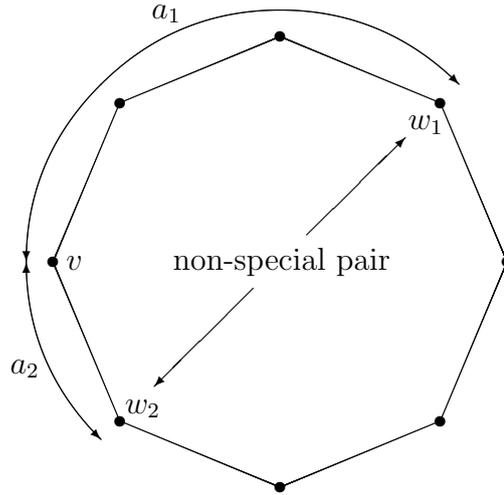


FIGURE 5.5. A cycle with a non-special opposite pair.

- (3) The aim is to count the $(2k + 2)$ -cycles containing v and a non-special pair of opposite vertices in groups depending on where the closest non-special pair to v is. Moving clockwise from v , let w_1 be the first vertex in a non-special pair and let a_1 be its distance from v , and let w_2 be the vertex opposite w_1 at distance a_2 anticlockwise from v , so that $a_1 + a_2 = k + 1$ (see Figure 5.5). It is possible that $w_1 = v$ and $a_1 = 0$. There are $k + 1$ choices for a_1 and a_2 . Given a_1 and a_2 , there are at most d^{a_1} paths of length a_1 and d^{a_2} paths of length a_2 starting at v , and so at most

$$d^{a_1+a_2} = d^{k+1}$$

choices for the arc from w_1 to w_2 containing v . Since $\{w_1, w_2\}$ is a non-special pair, there are at most d choices for the arc from w_1 to w_2 not containing v . In total, there are therefore at most

$$(k + 1)d^{k+1}d = (k + 1)d^{k+2}$$

$(2k + 2)$ -cycles containing v and a non-special pair of opposite vertices.

- (4) Without loss of generality, assume that v plays the role of one of v_0, v_1, \dots, v_{k+1} in $P = v_0v_1 \dots v_{2k+2}$, which gives $k + 2$ possibilities. For each of these there are at most d^{2k+2} paths P , which means that v is in at most $(k + 2)d^{2k+2}$ paths of length $2k + 2$.

□

By multiplying the number of vertices in a pair, cycle or path by the number of paths of length at most k , special pairs, cycles with a non-special pair and paths that each vertex can be in, the number of edges in the dependency graph from an event of each type to events of each type is at most that shown in the table below.

		to			
		$W_{u,v}$	$X_{u,v}$	Y_C	Z_P
from	$W_{u,v}$	$4d^k$	$2d^k$	$2(k+1)d^{k+2}$	$2(k+2)d^{2k+2}$
	$X_{u,v}$	$4d^k$	$2d^k$	$2(k+1)d^{k+2}$	$2(k+2)d^{2k+2}$
	Y_C	$4(k+1)d^k$	$2(k+1)d^k$	$2(k+1)^2d^{k+2}$	$2(k+1)(k+2)d^{2k+2}$
	Z_P	$2(2k+3)d^k$	$(2k+3)d^k$	$(2k+3)(k+1)d^{k+2}$	$(2k+3)(k+2)d^{2k+2}$

CLAIM 2. *The probabilities of the events of each type satisfy the following equalities and inequalities:*

- (1) $\mathbb{P}(W_{u,v}) = \frac{1}{x}$,
- (2) $\mathbb{P}(X_{u,v}) = \frac{1}{x}$,
- (3) $\mathbb{P}(Y_C) \leq \frac{(2k)^{2k+2}}{x^2}$, and
- (4) $\mathbb{P}(Z_P) \leq \frac{(2k)^{2k+3}}{x^3}$.

PROOF OF CLAIM. (1) This is true since f is a uniformly random colouring with colours taken from $[x]$.

(2) See (1).

(3) For each $S \subset [x]$ with $|S| = 2k$, let Y_C^S be the event that C is coloured with colours taken from S , so that

$$Y_C = \bigcup_{\substack{S \subset [x] \\ |S|=2k}} Y_C^S.$$

(However, this is not a partition, since if a colouring uses strictly fewer than $2k$ colours, then more than one event Y_C^S holds.) Given a choice of S , each of the $2k+2$ vertices of C must receive one of the colours in S , which means that

$$\mathbb{P}(Y_C^S) = \left(\frac{2k}{x}\right)^{2k+2},$$

and hence

$$\mathbb{P}(Y_C) \leq \sum_{\substack{S \subseteq [x] \\ |S|=2k}} \mathbb{P}(Y_C^S) \leq \binom{x}{2k} \left(\frac{2k}{x}\right)^{2k+2} \leq x^{2k} \left(\frac{2k}{x}\right)^{2k+2} = \frac{(2k)^{2k+2}}{x^2}.$$

(4) As in (3),

$$\mathbb{P}(Z_P) \leq \binom{x}{2k} \left(\frac{2k}{x}\right)^{2k+3} \leq \frac{(2k)^{2k+3}}{x^3}.$$

□

As in the proof of Theorem 5.4, the weightings used are double the probabilities or bounds for the probabilities given in Claim 2; it is enough to prove the condition in the Erdős-Lovász local lemma for the events Z_P , that is,

$$\begin{aligned} \left(1 - \frac{2}{x}\right)^{2(2k+3)d^k} & \left(1 - \frac{2}{x}\right)^{(2k+3)d^k} \left(1 - \frac{2(2k)^{2k+2}}{x^2}\right)^{(2k+3)(k+1)d^{k+2}} \\ & \left(1 - \frac{2(2k)^{2k+3}}{x^3}\right)^{(2k+3)(k+2)d^{2k+2}} \geq \frac{1}{2}. \end{aligned}$$

It suffices to find x such that all four factors are at least $\frac{6}{7}$, for which any

$$x \geq \max \left\{ 28(2k+3)d^k, 14(2k+3)d^k, \sqrt{14(2k+3)(k+1)(2k)^{k+1}d^{\frac{k}{2}+1}}, \sqrt[3]{14(2k)^{2k+3}(2k+3)(k+2)d^{\frac{2k+2}{3}}} \right\}$$

will do. Since $k \geq 2$, it is enough to take $x = 6(k+2)(2k)^{k+1}d^k$ (or even $x = 28(2k+3)d^k$ once d is reasonably large). □

When choosing the order of magnitude of x at the end of the proof, the critical condition was that the colouring was a distance- k colouring, not that special pairs and $(2k+2)$ -cycles were coloured in a particular way. This was because there was plenty of room to manoeuvre when choosing the threshold function for a pair to be considered special.

Indeed, suppose that a pair of vertices were called special if there are least d^s paths of length $k+1$ between them. Then each vertex would be in at most $O(d^{k+1-s})$ special pairs and at most $O(d^{k+1+s})$ $(2k+2)$ -cycles containing a non-special pair of opposite vertices. Since the probability that a special pair is monochromatic is $\frac{1}{x}$ and the probability that a

$(2k + 2)$ -cycle is coloured with at most $2k$ colours is $O(\frac{1}{x^2})$, the number of colours needed to avoid these events is $O(d^{k+1-s})$ and $O(d^{\frac{k+1+s}{2}})$. Since $k \geq 2$, these are both $O(d^k)$ for any choice of s provided that $1 \leq s \leq k - 1$.

In the case of 3-colour acyclic colouring studied by Alon, McDiarmid and Reed [4], $k = 1$ and so this is not possible. Instead, s is chosen so that $k + 1 - s = \frac{k+1+s}{2}$, that is, $s = \frac{k+1}{3} = \frac{2}{3}$, and the number of colours needed is $O(d^{\frac{4}{3}})$ rather than $O(d)$.

5.4. Open problems

In the results in this chapter, I have found the orders of magnitude of both the length- l -acyclic chromatic number (up to a logarithmic factor) and the c -colour acyclic chromatic number of a graph as functions of its maximum degree, assuming that l or c is fixed. As a result, the bounds are nowhere near best possible as functions of l or c . It would be worth trying to improve these and to remove the logarithmic factor in the lower bound for the length- l -acyclic chromatic number.

The two concepts in this chapter can be combined to give a *c-colour length-l-acyclic colouring*. If $l \geq c$ then this is a proper colouring such that every cycle of length at least l receives at least c colours. If $l < c$ then it is a proper colouring such that every cycle of length at least l but less than c is multicoloured and every cycle of length at least c receives at least c colours.

In this chapter I have proved results for $c = 3$ and general l , and for $l = 3$ and general c . What happens for general c and l ?

Small minimum degree colourings of graphs and hypergraphs

6.1. Introduction

Recall that a degree- r colouring of a u -uniform hypergraph is a colouring such that every edge is multicoloured and every subhypergraph with minimum degree at least r receives at least $u + 1$ colours. In this chapter, I shall bound the maximum possible degree- r chromatic number for u -uniform hypergraphs of a given maximum degree d , denoted by $D_r^{(u)}(d)$. The bounds are of a different nature when $u = 2$ and when $u \geq 3$, so I shall treat them separately. In Section 6.2, I shall bound $D_r^{(2)}(d)$, which for convenience I shall denote simply by $D_r(d)$, and in Section 6.3 I shall bound $D_r^{(u)}(d)$ for $u \geq 3$. In Section 6.4, I shall suggest some natural open problems following on from these.

6.2. The degree- r chromatic number of a graph

The main theorem that I shall prove in this section is the following.

THEOREM 6.1. *For integers $r, d \geq 2$, it is the case that*

$$\frac{\left(\frac{d}{6}\right)^{\frac{r^2}{r^2-1}}}{r \left(\frac{r^2}{r^2-1} \log d\right)^{\frac{1}{r^2-1}}} < D_r(d) < 2^{r^3+r+1} r^4 d^{\frac{r^2}{r^2-1}}.$$

This has the following immediate corollary.

COROLLARY 6.2. *For a fixed integer $r \geq 2$, it is the case that*

$$\Omega \left(\frac{d^{\frac{r^2}{r^2-1}}}{(\log d)^{\frac{1}{r^2-1}}} \right) \leq D_r(d) \leq O \left(d^{\frac{r^2}{r^2-1}} \right)$$

as $d \rightarrow \infty$. □

The probabilistic method for finding a lower bound is by now familiar from Theorems 5.3, 5.7 and 5.9.

THEOREM 6.3. *For n sufficiently large there is a graph G on n vertices with maximum degree $\Delta(G) = d$ satisfying*

$$\frac{n}{r} \geq \frac{\left(\frac{d}{6}\right)^{\frac{r^2}{r^2-1}}}{r \left(\frac{r^2}{r^2-1} \log d\right)^{\frac{1}{r^2-1}}} = \Omega \left(\frac{d^{\frac{r^2}{r^2-1}}}{(\log d)^{\frac{1}{r^2-1}}} \right)$$

such that there is no degree- r colouring with at most $\frac{n}{r}$ colours.

PROOF. Let

$$p = 3 \left(\frac{\log n}{n} \right)^{\frac{1}{r^2}},$$

let n be large and divisible by r^2 and let $G \in \mathcal{G}_{n,p}$.

Suppose that $f : V(G) \rightarrow \left[\frac{n}{r}\right]$ is a colouring that uses at most $\frac{n}{r}$ colours. By discarding at most $r - 1$ vertices from each colour class, at least $\frac{n}{r^2}$ monochromatic r -tuples can be found. Choose $\frac{n}{r^2}$ of these and call them $T_1, \dots, T_{\frac{n}{r^2}}$.

Let $K_{i,j}$ be the event that the graph induced by the vertices in $T_i \cup T_j$ contains the complete bipartite graph with vertex classes T_i and T_j . If any event $K_{i,j}$ occurs then this colouring cannot be a degree- r colouring of G . Let A_f be the event that f is a degree- r colouring. Then

$$\begin{aligned} \mathbb{P}(A_f) &\leq \mathbb{P} \left(\bigcap_{i < j} K_{i,j}^c \right) \\ &\leq \left(1 - p^{r^2}\right)^{\binom{n/r^2}{2}}, \end{aligned}$$

since the events $K_{i,j}$ are independent. Writing E for the event that G can be degree- r coloured with at most $\frac{n}{r}$ colours,

$$\begin{aligned} \mathbb{P}(E) &\leq \sum_{f:V(G)\rightarrow\left[\frac{n}{r}\right]} \mathbb{P}(A_f) \\ &\leq n^n \left(1 - p^{r^2}\right)^{\binom{n/r^2}{2}} \\ &\leq \exp \left(n \log n - \binom{n/r^2}{2} p^{r^2} \right) \\ &= o(1), \end{aligned}$$

which means that there exists a graph G that cannot be degree- r coloured with at most $\frac{n}{r}$ colours, as required. By Lemma 4.9, the graph G can be chosen so that, in addition, the bound in the theorem is satisfied. \square

THEOREM 6.4. *Let G be a graph with $\Delta(G) = d$ and let $r \geq 2$ be an integer. Then it is possible to degree- r colour G with at most $2^{r^3+r+1}r^4d^{\frac{r^2}{r^2-1}} = O\left(d^{\frac{r^2}{r^2-1}}\right)$ colours.*

PROOF. Call the tree on $r+1$ vertices, r of which are leaves, the r -claw (see Figure 6.1); call the r -tuple $\{v_1, \dots, v_r\}$ *special* if it is the set of leaves of at least $d^{\frac{r}{r+1}}$ r -claws, that is, if there are at least $d^{\frac{r}{r+1}}$ vertices adjacent to all of v_1, \dots, v_r .

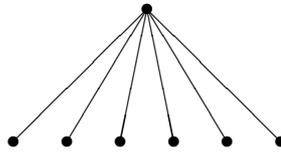


FIGURE 6.1. The 6-claw.

Let $f : V(G) \rightarrow [x]$ be a random function where each vertex receives a colour from $[x]$ uniformly at random.

The proof uses the Erdős-Lovász local lemma. The bad events to be considered are as follows:

- (1) For each edge uv , let $W_{uv} = \{f(u) = f(v)\}$.
- (2) For each special r -tuple v_1, \dots, v_r , let X_{v_1, \dots, v_r} be the event that $f(v_1) = \dots = f(v_r)$.
- (3) For each connected bipartite subgraph $B \subset G$ with $\delta(B) \geq r$ and $|V(B)| = t \leq r^2$ not containing a special r -tuple, let Y_B^t be the event that B receives at most two colours.
- (4) For each connected bipartite subgraph $B \subset G$ with $|V(B)| = r^2 + 1$, let Z_B be the event that B receives at most two colours.

If none of the events of type W_{uv} occurs then f is a proper colouring and every non-bipartite subgraph receives at least three colours. If in addition none of the events X_{v_1, \dots, v_r} and Y_B^t occurs then every bipartite subgraph with minimum degree at least r on at most r^2 vertices receives at least three colours. If none of the events Z_B occurs then every

bipartite subgraph with minimum degree at least r on at least $r^2 + 1$ vertices receives at least three colours.

CLAIM 1. *Each vertex is in at most*

- (1) d edges,
- (2) $d^{\frac{r^2}{r+1}}$ special r -tuples,
- (3) $t2^{\frac{t^2}{2}} d^{t-\frac{2r}{r+1}}$ connected bipartite subgraphs B with $\delta(B) \geq r$ and $|V(B)| = t \leq r^2$ not containing a special r -tuple, and
- (4) $(r^2 + 1)2^{r^4} d^{r^2}$ bipartite subgraphs B with $|V(B)| = r^2 + 1$.

PROOF OF CLAIM. (1) The maximum degree $\Delta(G) = d$.

- (2) The vertex v is a leaf in a most d^r r -claws: there are at most d choices for the root and then at most d choices for each of the other $r - 1$ leaves. Each special r -tuple that v is in provides at least $d^{\frac{r}{r+1}}$ r -claws, and so v is in at most

$$\frac{d^r}{d^{\frac{r}{r+1}}} = d^{\frac{r^2+r-r}{r+1}} = d^{\frac{r^2}{r+1}}$$

special r -tuples.

- (3) There are at most

$$2^{\binom{t}{2}} \leq 2^{\frac{t^2}{2}}$$

graphs on t vertices, so certainly at most that many connected bipartite graphs B on t vertices with $\delta(B) \geq r$. For each vertex $w \in V(B)$, I shall show that v is in at most $d^{t-\frac{2r}{r+1}}$ copies of B in the position of w , which is enough since there are at most t choices for w .

In the bipartite graph B , let V_1 be the smaller vertex class and let V_2 be the larger. (If they are equally sized then either labelling will do.) Let $s_1 = |V_1|$ and $s_2 = |V_2|$. I shall first construct a set $U \subset V_2$ with $|U| \leq s_2 - r + 1$ such that $w \in V_1 \cup U$ and the graph induced by the vertices in $V_1 \cup U$ is connected (see Figure 6.2).

If $w \in V_2$ then let $u_1 = w$; otherwise, let u_1 be any neighbour of w , so that in either case $u_1 \in V_2$. Let $X_1 = \Gamma(u_1)$, so that $w \in \{u_1\} \cup X_1$ and $|X_1| \geq r$. Since B is a connected graph, there exists some $u_2 \in V_2$ with neighbours in both X_1

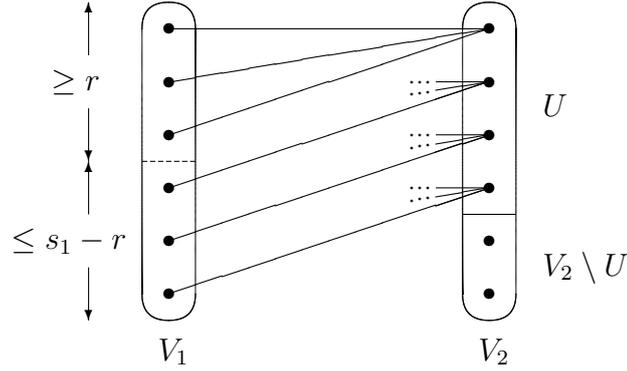


FIGURE 6.2. A set $U \subset V_2$ such that the graph induced by the vertices in $V_1 \cup U$ is connected, and $|U| \leq s_2 - r + 1$ and $|V_2 \setminus U| \geq r - 1$.

and $V_1 \setminus X_1$. Let $X_2 = X_1 \cup \Gamma(u_2)$. Since u_2 has neighbours in $V_1 \setminus X_1$, it is the case that $|X_2| \geq r + 1$, and since u_2 has neighbours in X_1 , the graph induced by the vertices in $\{u_1, u_2\} \cup X_2$ is connected.

Repeat this process, at each step picking a vertex u_{i+1} with neighbours in both X_i and $V_1 \setminus X_i$ and then defining $X_{i+1} = X_i \cup \Gamma(u_{i+1})$. By the choice of u_{i+1} , it is always the case that $|X_{i+1}| \geq r + i$ and that the graph induced by $\{u_1, \dots, u_{i+1}\} \cup X_{i+1}$ is connected.

The process terminates after j steps when $X_j = V_1$. Let $U = \{u_1, \dots, u_j\}$. Since $s_2 \geq s_1 = |V_1| = |X_j| \geq r + j - 1$, the inequality $|U| = j \leq s_2 - r + 1$ holds, as required.

Given w , there are at most $d^{s_1 - 1 + j}$ ways of choosing the $s_1 - 1 + j$ vertices in $(V_1 \cup U) \setminus \{w\}$. Since all the vertices in V_1 have now been chosen and there are no special r -tuples, there are then at most $\left(d^{\frac{r}{r+1}}\right)^{s_2 - j}$ ways of picking the $s_2 - j$ vertices in $V_2 \setminus U$, which gives a total of at most

$$d^{s_1 - 1 + j + \frac{(s_2 - j)r}{r+1}} = d^{\frac{r(s_1 + s_2) + s_1 + j - r - 1}{r+1}}$$

ways. Since $j \leq s_2 - r + 1$, this gives at most

$$d^{\frac{r(s_1 + s_2) + s_1 + s_2 - 2r}{r+1}} = d^{\frac{(r+1)t - 2r}{r+1}} = d^{t - \frac{2r}{r+1}}$$

ways, as required.

(4) There are at most

$$2^{\binom{r^2 + 1}{2}} \leq 2^{r^4}$$

graphs on $r^2 + 1$ vertices, so certainly at most that many connected bipartite graphs B on $r^2 + 1$ vertices. For each such B , there are at most $r^2 + 1$ choices for where to put v , up to isomorphism. Working through the remaining vertices one by one, each time picking a vertex that is adjacent to one of those already considered, there are then at most d^{r^2} choices for the other r^2 vertices, which means that v is in at most $(r^2 + 1)2^{r^4} d^{r^2}$ such graphs. \square

By multiplying the number of vertices in an edge, r -tuple or bipartite graph by the number of edges, special r -tuples, and bipartite graphs that each vertex can be in, the number of edges in the dependency graph from an event of each type to events of each type is at most that shown in the table below.

		to			
		W_{uv}	X_{v_1, \dots, v_r}	Y_B^t	Z_B
from	W_{uv}	$2d$	$2d^{\frac{r^2}{r+1}}$	$t2^{\frac{t^2}{2}+1}d^{t-\frac{2r}{r+1}}$	$(r^2 + 1)2^{r^4+1}d^{r^2}$
	X_{v_1, \dots, v_r}	rd	$rd^{\frac{r^2}{r+1}}$	$rt2^{\frac{t^2}{2}}d^{t-\frac{2r}{r+1}}$	$r(r^2 + 1)2^{r^4}d^{r^2}$
	Y_B^t	r^2d	$r^2d^{\frac{r^2}{r+1}}$	$r^2t2^{\frac{t^2}{2}}d^{t-\frac{2r}{r+1}}$	$r^2(r^2 + 1)2^{r^4}d^{r^2}$
	Z_B	$(r^2 + 1)d$	$(r^2 + 1)d^{\frac{r^2}{r+1}}$	$(r^2 + 1)t2^{\frac{t^2}{2}}d^{t-\frac{2r}{r+1}}$	$(r^2 + 1)^22^{r^4}d^{r^2}$

CLAIM 2. *The probabilities of the events of each type satisfy the following equalities and inequalities:*

- (1) $\mathbb{P}(W_{uv}) = \frac{1}{x}$,
- (2) $\mathbb{P}(X_{v_1, \dots, v_r}) = \frac{1}{x^{r-1}}$,
- (3) $\mathbb{P}(Y_B^t) \leq \frac{2^t}{x^{t-2}}$, and
- (4) $\mathbb{P}(Z_B) \leq \frac{2^{r^2}}{x^{r^2-1}}$.

PROOF OF CLAIM. This is straightforward, and proved in the same way as Claim 2 in the proof of Theorem 5.11. \square

As in the proofs of Theorems 5.4 and 5.11, the weightings used are double the probabilities or bounds for the probabilities given in Claim 2, and it is enough to prove the

condition in the Erdős-Lovász local lemma for the events Z_B , that is,

$$\left(1 - \frac{2}{x}\right)^{(r^2+1)d} \left(1 - \frac{2}{x^{r-1}}\right)^{(r^2+1)d^{\frac{r^2}{r+1}}} \prod_{t=2r}^{r^2} \left(1 - \frac{2^{t+1}}{x^{t-2}}\right)^{(r^2+1)t2^{\frac{t^2}{2}}d^{t-\frac{2r}{r+1}}} \left(1 - \frac{2^{r^2+1}}{x^{r^2-1}}\right)^{(r^2+1)2^{r^4}d^{r^2}} \geq \frac{1}{2}.$$

Since there are at most r^2 factors in this product, it suffices to find x such that each factor is at least $1 - \frac{1}{2r^2}$. The power of d in the order of magnitude of x necessitated by each term

$$\left(1 - \frac{2^{t+1}}{x^{t-2}}\right)^{(r^2+1)t2^{\frac{t^2}{2}}d^{t-\frac{2r}{r+1}}}$$

is $\frac{t-\frac{2r}{r+1}}{t-2}$. This is maximized in the range $2r \leq t \leq r^2$ when $t = 2r$, when it is equal to

$$\frac{2r - \frac{2r}{r+1}}{2r - 2} = \frac{2r^2 + 2r - 2r}{2(r-1)(r+1)} = \frac{r^2}{r^2 - 1}.$$

It therefore suffices to choose any

$$x \geq \max \left\{ 4r^2(r^2 + 1)d, \sqrt[r-1]{4r^2(r^2 + 1)d^{\frac{r^2}{r^2-1}}}, \sqrt[2r-2]{2^{\frac{r^4}{2}+r^2+2}r^4(r^2 + 1)r^{2r^3+2}d^{\frac{r^2}{r^2-1}}}, \sqrt[r^2-1]{2^{r^4+r^2+2}r^2(r^2 + 1)^2d^{\frac{r^2}{r^2-1}}} \right\}.$$

Since $r \geq 2$, it is enough to take $x = 2^{r^3+r+1}r^4d^{\frac{r^2}{r^2-1}}$. \square

6.3. The degree- r chromatic number of a hypergraph

There are (at least) two possible definitions of what constitutes a proper vertex-colouring of a uniform hypergraph: one where each edge must be multicoloured and one where each edge cannot be monochromatic. I shall consider the first case, and prove that degree- r colouring a u -uniform hypergraph for $u \geq 3$ is essentially no harder than ensuring that all of its edges are multicoloured. Colouring the vertices greedily in any order shows that the vertices of a u -uniform hypergraph H with maximum degree $\Delta(H) = d$ can be coloured with at most $(u-1)d + 1$ colours, and I shall show that a number of colours linear in d is sufficient to degree-2 colour and therefore to degree- r colour H for all r .

Recall that a degree- r colouring of a u -uniform hypergraph is a colouring such that every edge is multicoloured and every subhypergraph with minimum degree at least r receives at least $u + 1$ colours. In this section, I shall bound the maximum possible degree- r chromatic number of hypergraphs of a given maximum degree d , denoted by $D_r^{(u)}(d)$.

THEOREM 6.5. *For integers $u \geq 3$, $r \geq 2$ and $d \geq 2$, it is the case that*

$$\left(1 - \frac{1}{u}\right) d < D_r^{(u)}(d) \leq D_2^{(u)}(d) < 3du^u.$$

This has the following immediate corollary.

COROLLARY 6.6. *For fixed integers $u \geq 3$, $r \geq 2$, it is the case that*

$$D_r^{(u)}(d) = \Theta(d)$$

as $d \rightarrow \infty$.

□

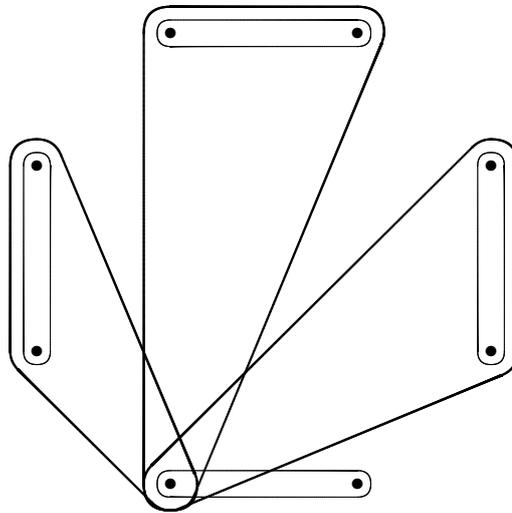


FIGURE 6.3. The desired hypergraph is formed by superimposing one copy of this hypergraph for each vertex.

For every edge to be multicoloured, it might be necessary to use at least $(1 - \frac{1}{u})d$ colours. Indeed, for any $m \geq 2$, let H be a u -uniform hypergraph with $V(H) = [m(u-1)]$, for $0 \leq j \leq m-1$ let

$$I_j = \{j(u-1) + 1, j(u-1) + 2, \dots, (j+1)(u-1)\}$$

and let

$$E(H) = \{\{i\} \cup I_j : i \notin I_j\}$$

(see Figure 6.3). Then each vertex i is in an edge for each of the $m - 1$ intervals I_j not containing i and in an edge for each of the $(m - 1)(u - 1)$ vertices not in the interval containing i . Therefore,

$$d = m - 1 + (m - 1)(u - 1) = (m - 1)u$$

and every pair of vertices is in at least one common edge, so the number of colours needed to ensure that every edge is multicoloured is

$$(u - 1)m = (m - 1)(u - 1) + u - 1 = \left(1 - \frac{1}{u}\right) d + u - 1.$$

The proof of the upper bound using the Erdős-Lovász local lemma is easier than those up to this point, in the sense that there are only two types of bad event.

THEOREM 6.7. *Let H be a u -uniform hypergraph with $\Delta(G) = d$. Then it is possible to degree-2 colour H with at most $3du^u$ colours.*

PROOF. Let $f : V(G) \rightarrow [x]$ be a random function where each vertex receives a colour from $[x]$ uniformly at random.

The proof uses the Erdős-Lovász local lemma. The bad events to be considered are as follows:

- (1) For all v_1, v_2 with v_1 and v_2 in some common edge, let $W_{v_1, v_2} = \{f(v_1) = f(v_2)\}$.
- (2) For every pair of edges e_1, e_2 with $1 \leq |e_1 \cap e_2| \leq u - 2$, let X_{e_1, e_2} be the event that the two edges receive the same multiset of colours.

If none of the events W_{v_1, v_2} occurs then f is a multicolouring and every non- u -partite subhypergraph receives at least $u + 1$ colours. In the claim below, I shall show that every connected u -partite u -uniform hypergraph with minimum degree at least 2 contains two edges e_1, e_2 with $1 \leq |e_1 \cap e_2| \leq u - 2$. Therefore, if in addition none of the events X_{e_1, e_2} occurs then every subhypergraph with minimum degree at least 2 receives at least $u + 1$ colours.

CLAIM 1. *Let K be a connected u -partite u -uniform hypergraph with $\delta(K) \geq 2$, where $u \geq 3$. Then there exist $e_1, e_2 \in E(K)$ such that $1 \leq |e_1 \cap e_2| \leq u - 2$.*

PROOF OF CLAIM. Suppose that there are no two edges $e_1, e_2 \in E(K)$ such that $1 \leq |e_1 \cap e_2| \leq u - 2$. Then for any two distinct edges $e_1 \neq e_2$, either $e_1 \cap e_2 = \emptyset$ or $|e_1 \cap e_2| = u - 1$.

Pick any edge $e_1 \in E(K)$. Since K is connected, there is an edge $e_2 \in E(K)$ with $|e_1 \cap e_2| = u - 1$. Let v_1 be the vertex in $e_1 \setminus e_2$, let v_2 be the vertex in $e_2 \setminus e_1$ and let $S = e_1 \cap e_2$. Since K is u -partite, v_1 and v_2 cannot be in any common edge.

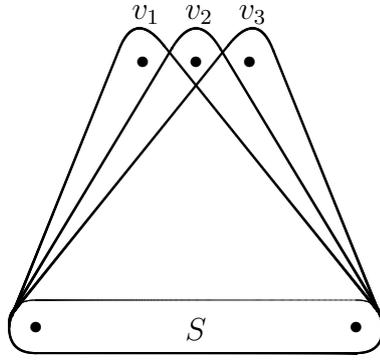


FIGURE 6.4. The condition on the sizes of the intersections forces K to have this structure, which contradicts $\delta(K) \geq 2$.

There must be at least three edges, as otherwise $d(v_1) = d(v_2) = 1$, which contradicts $\delta(K) \geq 2$. Thus there is a third edge e_3 that intersects at least one of e_1 and e_2 , so without loss of generality $|e_1 \cap e_3| = u - 1$. Since $u < 2(u - 1)$ for $u \geq 3$, this means that $e_2 \cap e_3 \neq \emptyset$ and so $|e_2 \cap e_3| = u - 1$. Since $|e_1 \cap e_3| = |e_2 \cap e_3|$, the edge e_3 must contain both v_1 and v_2 or neither. The former is not possible as K is u -partite, so it must be that $e_1 \cap e_2 = e_1 \cap e_3 = e_2 \cap e_3 = S$, and $e_3 = S \cup \{v_3\}$ for some v_3 (see Figure 6.4).

Since K is connected, the argument can be repeated and must eventually include all edges, so that each edge $e_i = S \cup \{v_i\}$ for some vertex v_i . But then the degree $d(v_i) = 1$ for all i , which contradicts $\delta(K) \geq 2$. \square

CLAIM 2. *Each vertex v is in at most*

- (1) $(u - 1)d$ pairs of vertices in a common edge, and
- (2) ud^2 pairs of edges e_1, e_2 with $1 \leq |e_1 \cap e_2| \leq u - 2$.

PROOF OF CLAIM. (1) Since $\Delta(H) = d$, the vertex v is in at most d edges, each of which contains $u - 1$ more vertices (not necessarily different from those in other edges).

(2) Since $\Delta(H) = d$, the vertex v is in at most $\binom{d}{2} < d^2$ pairs of edges whose intersection contains v . By choosing one of the edges containing v , one of the other vertices in that edge and then one of the other edges containing that vertex, the vertex v is in at most $(u - 1)d^2$ pairs of intersecting edges whose intersection does not contain v .

□

Note that the bounds on $|e_1 \cap e_2|$ were not needed to obtain the bound in (2); this condition is in the statement of the claim only as a reminder of the type of event being considered.

By multiplying the number of vertices in a pair of vertices or edges by the number of pairs of vertices and edges that each vertex can be in, the number of edges in the dependency graph from an event of each type to events of each type is at most that shown in the table below.

		to	
		W_{v_1, v_2}	X_{e_1, e_2}
from	W_{v_1, v_2}	$2(u - 1)d$	$2ud^2$
	X_{e_1, e_2}	$(2u - 1)(u - 1)d$	$(2u - 1)ud^2$

CLAIM 3. *The probabilities of the events of each type satisfy the following equality and inequality:*

- (1) $\mathbb{P}(W_{v_1, v_2}) = \frac{1}{x}$ and
- (2) $\mathbb{P}(X_{e_1, e_2}) \leq \frac{(u-1)!}{x^2}$.

PROOF OF CLAIM. (1) This is true since f is a uniformly random colouring with colours taken from $[x]$.

(2) Let $q = |e_1 \cap e_2|$, so that $1 \leq q \leq u - 2$. For each bijection $\pi : e_1 \setminus e_2 \rightarrow e_2 \setminus e_1$, let

$$A_\pi = \left\{ f(v) = f(\pi(v)) \text{ for all } v \in e_1 \setminus e_2 \right\}$$

be the event that the colourings of the edges e_1 and e_2 match up in this order, and let E be the event that the colourings of the edges e_1 and e_2 are the same in any order, so that

$$E = \bigcup_{\pi: e_1 \setminus e_2 \rightarrow e_2 \setminus e_1} A_\pi.$$

(However, this is not a partition, since if f is not a multicolouring, then more than one event A_π might hold.) Then

$$\mathbb{P}(E) \leq \sum_{\pi: e_1 \setminus e_2 \rightarrow e_2 \setminus e_1} \mathbb{P}(A_\pi) = \frac{(u-q)!}{x^{(u-q)}} \leq \frac{(u-1)!}{x^2}.$$

□

As in the proofs of Theorems 5.4, 5.11 and 6.4, the weightings used are double the probabilities or bounds for the probabilities given in Claim 2, and it is enough to prove the condition in the Erdős-Lovász local lemma for the events X_{e_1, e_2} , that is,

$$\left(1 - \frac{2}{x}\right)^{(2u-1)(u-1)d} \left(1 - \frac{2(u-1)!}{x^2}\right)^{(2u-1)ud^2} \geq \frac{1}{2}.$$

It suffices to find x for which both factors are at least $\frac{3}{4}$, for which any

$$x \geq \max \left\{ 8(2u-1)(u-1)d, \sqrt{8(u-1)!(2u-1)ud} \right\}$$

will do. Since $u \geq 3$, it is enough to take $x = 3du^u$.

□

6.4. Open problems

As in Chapter 5, I have only tried to find the order of magnitude of the degree- r chromatic numbers of graphs and u -uniform hypergraphs as functions of their maximum degree d , and it would be nice to have tighter bounds as functions of r and u as well.

An alternative way of describing a degree- r colouring is that any 2-coloured subgraph has minimum degree at most r . What if the graph induced by any two classes has *maximum* degree at most r ?

Over the course of the last two chapters, I have introduced an analogue of the acyclic chromatic number by varying the following one at a time:

- (1) the minimum length l of the cycles that we care about (or more generally the minimum order l of the subgraphs we care about),
- (2) the minimum number of colours c that these subgraphs should receive,
- (3) the minimum degree r of these subgraphs, and
- (4) the number of vertices u in an edge.

In addition to these, for $u \geq 3$ one could relax the condition from every edge being multicoloured to no edge being monochromatic.

The ultimate goal is a result including all five of these factors, but for now even a result taking more than one into account would constitute progress. As well as uniting some or all of these, one could merge them with many of the ideas from Chapter 4. For example, there are obvious list and edge analogues. Finding exact best possible bounds for the various quantities as functions of the graph's maximum degree d for small d would also be good.

Beyond that, there is still plenty of scope for varying the problems posed; in this part of this dissertation, I have only attempted to bound them as functions of the graph's maximum degree. As with the original work on acyclic colourings, it would be interesting to find bounds for graphs drawn on different surfaces, for graphs with large girth, and for random graphs, for example.

Conclusion

In this dissertation, I proved eight main results, four in each part.

Part 1: Problems of optimal choice on posets

In Chapter 2, I found an optimal strategy for the secretary problem when the candidates are m pairs of identical twins, its probability of success and the asymptotic behaviour of both of these.

RESULT 1 (Theorem 2.1). *For $m \in \mathbb{N}$, let*

$$k_m = \min \left\{ k : \frac{2m}{k} + \sum_{j=k}^{m-1} \frac{1}{j} \leq 5 \right\}.$$

An optimal strategy for the secretary problem on m pairs of identical twins is to wait until candidates have been seen from at least k_m of the pairs and then to pick the next candidate who is the best so far and whose twin has already been seen. Asymptotically,

$$\lim_{m \rightarrow \infty} \frac{k_m}{m} = \frac{1}{x_0} \approx 0.4709,$$

where x_0 is the unique solution to $2x + \log x = 5$, and the probability of success tends to

$$\frac{1}{x_0} + \frac{4(x_0 - 1)^2}{3x_0} \left(\left(\frac{x_0}{x_0 - 1} \right)^{\frac{1}{2}} - 1 \right) \approx 0.7680.$$

For its extension to c -tuplets, I found an optimal strategy and its probability of success, and bounded the behaviour of these.

RESULT 2 (Theorem 2.2). *For $c, m \in \mathbb{N}$ with $c \geq 2$, let*

$$k_m^{(c)} = \min \left\{ k : \sum_{j=1}^{m-k} \left[\frac{\binom{m-j-1}{k-1}}{\binom{m-1}{k-1}} \prod_{i=2}^j \left(1 - \frac{1}{\binom{ci}{c}} \right) \right] \leq 1 \right\}.$$

An optimal strategy for the secretary problem on m sets of identical c -tuplets is to wait until candidates have been seen from at least $k_m^{(c)}$ of the c -tuples and then to pick the next

candidate who is the best so far and all of whose c -tuplets have already been seen. For all m ,

$$\left(\frac{1}{2} - \frac{c+1}{2(2^{c+1}(c-1) - c - 1)} \right) m < k_m^{(c)} \leq \left\lceil \frac{m}{2} \right\rceil,$$

and the probability of success is at least

$$1 - \frac{c+1}{2^c(c-1)}.$$

In Chapter 3, I found a strategy for the secretary problem on a poset where the number of elements n and the number of which are maximal k are both known and the width is equal to k , and showed that it gave the best possible probability of success for this problem.

RESULT 3 (Theorem 3.1). *Let (P, \prec) be a poset with k maximal elements and of width k . Then there is an algorithm for the secretary problem on (P, \prec) depending only on $|P|$ and k that is successful with probability at least p_k , where*

$$p_k = \begin{cases} \frac{1}{e} & \text{if } k = 1, \\ \sqrt[k-1]{\frac{1}{k}} & \text{if } k > 1, \end{cases}$$

and these are the best possible such bounds.

For a poset whose width is not necessarily equal to its number of maximal elements k , I found a strategy that is successful with probability $\frac{1}{e}$, which is the best possible probability of success for the classical secretary problem and therefore best possible in this situation.

RESULT 4 (Theorem 3.2). *Let (P, \prec) be a poset with k maximal elements. Then there is an algorithm for the secretary problem on (P, \prec) depending only on $|P|$ and k that is successful with probability at least $\frac{1}{e}$, and this is the best possible such bound.*

Part 2: Generalizations of acyclic colourings

In Chapter 5, I found an upper bound for the number of colours needed to colour a graph G properly in such a way that every cycle of length at least l receives at least three

colours, as a function of its maximum degree d , and showed that it was best possible up to a logarithmic factor.

RESULT 5 (Corollary 5.2). *For a fixed integer $m \geq 2$, it is the case that*

$$\Omega\left(\frac{d^{\frac{2m}{2m-1}}}{(\log d)^{\frac{1}{2m-1}}}\right) \leq A^{(2m)}(d) = A^{(2m-1)}(d) \leq O\left(d^{\frac{2m}{2m-1}}\right)$$

as $d \rightarrow \infty$.

I found an upper bound for the number of colours needed to colour a graph G properly in such a way that every cycle receives at least c colours, as a function of its maximum degree d , and showed that it was best possible up to a constant factor.

RESULT 6 (Corollary 5.6). *For a fixed integer $k \geq 2$, it is the case that*

$$A_{2k}(d), A_{2k+1}(d) = \Theta(d^k)$$

as $d \rightarrow \infty$.

In Chapter 6, I found an upper bound for the number of colours needed to colour a graph G properly in such a way that every subgraph with minimum degree at least r receives at least three colours, as a function of its maximum degree d , and showed that it was best possible up to a logarithmic factor.

RESULT 7 (Corollary 6.2). *For a fixed integer $r \geq 2$, it is the case that*

$$\Omega\left(\frac{d^{\frac{r^2}{r^2-1}}}{(\log d)^{\frac{1}{r^2-1}}}\right) \leq D_r(d) \leq O\left(d^{\frac{r^2}{r^2-1}}\right)$$

as $d \rightarrow \infty$.

I found an upper bound for the number of colours needed to colour a u -uniform hypergraph H in such a way that every edge is multicoloured and every subhypergraph of minimum degree at least r receives at least three colours, as a function of its maximum degree d , and showed that it was best possible up to a constant factor.

RESULT 8 (Corollary 6.6). *For fixed integers $u \geq 3, r \geq 2$, it is the case that*

$$D_r^{(u)}(d) = \Theta(d)$$

as $d \rightarrow \infty$.

These results are part of a natural progression from earlier work, and suggest possible directions of future research.

Bibliography

1. M. O. Albertson and D. M. Berman, *The acyclic chromatic number*, Proceedings of the seventh southeastern conference on combinatorics, graph theory, and computing, 1976, pp. 51–69.
2. ———, *Every planar graph has an acyclic 7-coloring*, Israel J. Math. **28** (1977), no. 1, 169–174.
3. ———, *An acyclic analogue to Heawood’s theorem*, Glasg. Math. J. **19** (1978), no. 2, 163–166.
4. N. Alon, C. McDiarmid, and B. Reed, *Acyclic coloring of graphs*, Random Structures & Algorithms **2** (1991), no. 3, 277–288.
5. N. Alon and J. H. Spencer, *The probabilistic method*, Wiley-Interscience, 2000.
6. N. Alon, B. Sudakov, and A. Zaks, *Acyclic edge colorings of graphs*, J. Graph Theory **37** (2001), no. 3, 157–167.
7. K. Appel and W. Haken, *Every planar map is four colorable*, Bull. Amer. Math. Soc. **82** (1976), no. 5, 711–712.
8. ———, *Every planar map is four colorable. Part I: discharging*, Illinois J. Math. **21** (1977), no. 3, 429–490.
9. K. Appel, W. Haken, and J. Koch, *Every planar map is four colorable. Part II: reducibility*, Illinois J. Math. **21** (1977), no. 3, 491–567.
10. B. Bollobás, *The diameter of random graphs*, Trans. Amer. Math. Soc. **267** (1981), no. 1, 41–52.
11. ———, *Modern graph theory*, Springer Verlag, 1998.
12. ———, *Random graphs*, Cambridge University Press, 2001.
13. O. V. Borodin, *On acyclic colorings of planar graphs*, Discrete Math. **25** (1979), no. 3, 211–236.
14. O. V. Borodin, D. G. Fon-Der Flaass, A. V. Kostochka, A. Raspaud, and E. Sopena, *Acyclic list 7-coloring of planar graphs*, J. Graph Theory **40** (2002), no. 2, 83–90.
15. O. V. Borodin, A. V. Kostochka, and D. R. Woodall, *Acyclic colourings of planar graphs with large girth*, J. Lond. Math. Soc. **60** (1999), no. 2, 344–352.
16. R. L. Brooks, *On colouring the nodes of a network*, Math. Proc. Cambridge Philos. Soc. **37** (1941), no. 2, 194–197.
17. M. I. Burnstein, *Every 4-valent graph has an acyclic 5-coloring*, Soobšč Akad. Nauk Gruzin. SSR **93** (1979), no. 1, 21–24 (Russian).
18. A. Cayley, *Mathematical questions with their solutions*, The Educational Times **23** (1875), 18–19.
19. G. Chartrand, D. Geller, and S. Hedetniemi, *Graphs with forbidden subgraphs*, J. Combin. Theory Ser. B **10** (1971), no. 1, 12–41.

20. G. Chartrand, H. Kronk, and C. Wall, *The point-arboricity of a graph*, Israel J. Math. **6** (1968), no. 2, 169–175.
21. Y. S. Chow, S. Moriguti, H. Robbins, and S. M. Samuels, *Optimal selection based on relative rank (the “secretary problem”)*, Israel J. Math. **2** (1964), no. 2, 81–90.
22. Y. S. Chow, H. Robbins, and D. Siegmund, *The theory of optimal stopping*, Dover, 1991.
23. K. L. Chung, *A course in probability theory*, Academic Press, 2001.
24. R. P. Dilworth, *A decomposition theorem for partially ordered sets*, Ann. of Math. **51** (1950), no. 1, 161–166.
25. P. Erdős and L. Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions*, Infinite and finite sets, 1975, pp. 609–627.
26. T. S. Ferguson, *Who solved the secretary problem?*, Statist. Sci. **4** (1989), no. 3, 282–289.
27. A. Q. Frank and S. M. Samuels, *On an optimal stopping problem of Gusein-Zade*, Stochastic Process. Appl. **10** (1980), no. 3, 299–311.
28. P. R. Freeman, *The secretary problem and its extensions: a review*, Int. Stat. Rev. **51** (1983), no. 2, 189–206.
29. R. Freij and J. Wästlund, *Partially ordered secretaries*, Electron. Commun. Probab. **15** (2010), 504–507.
30. J. Galambos, *Advanced probability theory*, CRC, 1995.
31. F. Galvin, *The list chromatic index of a bipartite multigraph*, J. Combin. Theory Ser. B **63** (1995), no. 1, 153–158.
32. M. Gardner, *Mathematical games*, Sci. Amer. **202** (1960), no. 2, 152.
33. ———, *Mathematical games*, Sci. Amer. **202** (1960), no. 3, 178–179.
34. B. J. Garrod, G. Kubicki, and M. Morayne, *How to choose the best twins*, SIAM J. Discrete Math., to appear.
35. B. J. Garrod and R. D. Morris, *The secretary problem on an unknown poset*, submitted.
36. N. Georgiou, M. Kuchta, M. Morayne, and J. Niemiec, *On a universal best choice algorithm for partially ordered sets*, Random Structures & Algorithms **32** (2008), no. 3, 263–273.
37. S. Gerke, C. Greenhill, and N. Wormald, *The generalized acyclic edge chromatic number of random regular graphs*, J. Graph Theory **53** (2006), no. 2, 101–125.
38. N. Ghoussoub, *An integral representation of randomized probabilities and its applications*, Séminaire de Probabilités XVI 1980/81, 1982, pp. 519–543.
39. J. Gianini-Pettitt, *Optimal selection based on relative ranks with a random number of individuals*, Adv. in Appl. Probab. **11** (1979), no. 4, 720–736.
40. J. P. Gilbert and F. Mosteller, *Recognizing the maximum of a sequence*, J. Amer. Statist. Assoc. **61** (1966), no. 313, 35–73.

41. A. V. Gnedin, *A multicriteria problem of optimal stopping of a selection process*, Autom. Remote Control **42** (1981), 981–986.
42. ———, *Multicriteria extensions of the best choice problem: sequential selection without linear order*, Strategies for sequential search and selection in real time, 1992, pp. 153–172.
43. C. Greenhill and O. Pikhurko, *Bounds on the generalised acyclic chromatic numbers of bounded degree graphs*, Graphs Combin. **21** (2005), no. 4, 407–419.
44. B. Grünbaum, *Acyclic colorings of planar graphs*, Israel J. Math. **14** (1973), no. 4, 390–408.
45. S. M. Gusein-Zade, *The problem of choice and the optimal stopping rule for a sequence of independent trials*, Theory Probab. Appl. **11** (1966), no. 3, 472–476.
46. I. Guttman, *On a problem of L. Moser*, Canad. Math. Bull. **3** (1960), 35–39.
47. P. J. Heawood, *Map colour theorem*, Q. J. Math. **24** (1890), 332–338.
48. J. Kahn, *Asymptotics of the list-chromatic index for multigraphs*, Random Structures & Algorithms **17** (2000), no. 2, 117–156.
49. W. Kaźmierczak, *An algorithm that chooses the maximum element of a partially ordered set*, Master’s thesis, Politechnika Wroclawska, 2006 (Polish).
50. A. V. Kostochka, *Acyclic 6-coloring of planar graphs*, Diskret. Analiz. **28** (1976), 40 (Russian).
51. ———, *Upper bounds of chromatic functions of graphs*, 1978.
52. K. Kothapalli, S. Varagani, V. C. Venkaiah, and K. Yadav, *Acyclic vertex coloring of graphs of maximum degree 5*, International conference on graph theory and its applications, 2008.
53. ———, *Acyclic vertex coloring of graphs of maximum degree 6*, LAGOS: Latin-American algorithms, graphs and optimization symposium, 2009.
54. J. Kozik, *Dynamic threshold strategy for universal best choice problem*, 21st international meeting on probabilistic, combinatorial, and asymptotic methods in the analysis of algorithms, 2010, pp. 439–452.
55. G. Kubicki and M. Morayne, *Graph-theoretic generalization of the secretary problem: the directed path case*, SIAM J. Discrete Math. **19** (2005), no. 3, 622–632.
56. M. Kuchta and M. Morayne, *A secretary problem with many lives*, submitted.
57. D. V. Lindley, *Dynamic programming and decision theory*, J. R. Stat. Soc. Ser. C. Appl. Stat. **10** (1961), no. 1, 39–51.
58. J. Mitchem, *Every planar graph has an acyclic 8-coloring*, Duke Math. J. **41** (1974), no. 1, 177–181.
59. M. Molloy and B. Reed, *Further algorithmic aspects of the local lemma*, Proceedings of the thirtieth annual ACM symposium on theory of computing, 1998, pp. 524–529.
60. M. Morayne, *Partial-order analogue of the secretary problem: the binary tree case*, Discrete Math. **184** (1998), no. 1–3, 165–181.
61. M. Morayne and M. Sulowska, *Graph-theoretic generalization of the best choice problem; randomized analysis of a simple effective algorithm for k -ary trees*, submitted.

62. L. Moser, *On a problem of Cayley*, Scripta Mathematica **22** (1956), no. 5, 289–292.
63. R. Muthu, N. Narayanan, and C. R. Subramanian, *Improved bounds on acyclic edge colouring*, Discrete Math. **307** (2007), no. 23, 3063–3069.
64. C. St.J. A. Nash-Williams, *Edge-disjoint spanning trees of finite graphs*, J. Lond. Math. Soc. **36** (1961), no. 1, 445–450.
65. ———, *Decomposition of finite graphs into forests*, J. Lond. Math. Soc. **39** (1964), no. 1, 12.
66. J. Nešetřil and N. C. Wormald, *The acyclic edge chromatic number of a random d -regular graph is $d + 1$* , J. Graph Theory **49** (2005), no. 1, 69–74.
67. J. D. Petrucci, *Best-choice problems involving uncertainty of selection and recall of observations*, J. Appl. Probab. **18** (1981), no. 2, 415–425.
68. J. Preater, *The best-choice problem for partially ordered objects*, Oper. Res. Lett. **25** (1999), no. 4, 187–190.
69. E. L. Presman and I. M. Sonin, *The best choice problem for a random number of objects*, Theory Probab. Appl. **17** (1973), no. 4, 657–668.
70. M. Przykucki, *Optimal stopping in a search for a vertex with full degree in a random graph*, submitted.
71. M. Przykucki and M. Sulkowska, *Gusein-Zade problem for directed path*, Discrete Optim. **7** (2010), no. 1–2, 13–20.
72. G. Ringel and J. W. T. Youngs, *Solution of the Heawood map-coloring problem*, Proc. Natl. Acad. Sci. USA **60** (1968), no. 2, 438–445.
73. S. Skulrattanakulchai, *Acyclic colorings of subcubic graphs*, Inform. Process. Lett. **92** (2004), no. 4, 161–167.
74. M. H. Smith, *A secretary problem with uncertain employment*, J. Appl. Probab. **12** (1975), no. 3, 620–624.
75. W. Stadje, *Efficient stopping of a random series of partially ordered points*, Multiple criteria decision making theory and applications, 1980, pp. 430–447.
76. C. Thomassen, *Every planar graph is 5-choosable*, J. Combin. Theory Ser. B **62** (1994), no. 1, 180–181.
77. J. Tkocz, *The best choice problem for almost linear orders*, submitted.
78. V. G. Vizing, *On an estimate of the chromatic class of a p -graph*, Diskret. Analiz. **3** (1964), 25–30 (Russian).
79. M. Voigt, *List colourings of planar graphs*, Discrete Math. **120** (1993), no. 1–3, 215–219.
80. M. C. K. Yang, *Recognizing the maximum of a random sequence based on relative rank with backward solicitation*, J. Appl. Probab. **11** (1974), no. 3, 504–512.