Order Theory and Nonparametric Analysis for Interval Censored Data

Alain C. Vandal

A thesis submitted in partial fulfulment of the requirements
for the degree of Doctor of Philosophy in Statistics,
The University of Auckland, 1998

Abstract

Interval censored data arises when individuals can be subjected to periodic inspection at random moments, and their status (e.g. failed or functioning) is ascertained at each inspection. We exploit the order theoretic properties of interval orders to develop a new language describing interval censored data. We propose a method by which the set of linear extensions of an interval order may be partitioned into sets of linear extensions of weak orders, using so-called marked configurations of the interval order. The technique relies heavily on the natural linear ordering of maximal antichains in interval orders. We also propose a method whereby sets from this partition can be generated with known probability so as to permit efficient cluster or staged sampling. These techniques, among other uses, may be applied to generate sampling estimates of average rank score statistics for interval censored data similar in construction to that proposed by Prentice (1978) for right-censored data. In order to address the above problem we must determine all sets which form minimal covers of maximal antichains for interval orders. Finding minimal covers generalizes the minimum clique cover problem. We produce an algorithm enumerating all minimal covers using the minimal elements of the interval order and also characterize maximal removable sets, which are the complements of minimal covers We use this characterization to provide bounds on the maximum number of minimal covers for an interval order with a given number of maximal antichains. Finally, we determine nonparametric maximum likelihood estimators (NPMLE) of the cumulative distribution function (CDF) on the set of maximal antichains $\mathcal{M}$ of the data rather than the real line, extending the reasoning of Peto (1973) and Turnbull (1976). We discuss some properties of self-consistent estimators of the CDF in light of the structure of $\mathcal{M}$. We show the identity between self-consistency augmented by Kuhn-Tucker conditions and Fenchel duality, which characterize the NPMLE on $\mathcal{M}$. We port to $\mathcal{M}$ recently developed isotonic regression techniques to estimate the NPMLE. We correct some misapprehensions which have gained currency in recent literature on interval censored data.

Keywords: interval censored data; survival analysis; interval order; nonparametric maximum likelihood; maximal antichains; self-consistency; isotonic regression; linear extensions; order partition.

Acknowledgments

The folklore of writing a doctoral dissertation is rich in stories of frustration and tends to fuel fond complaints for years after its completion. Through good luck, or perhaps bad luck for my stock of anecdotes, the story of this dissertation is one of active research and discovery, and most of all the story of an ideal research collaboration. As we were attacking a sinkful of dirty dishes, Robert Gentleman described a problem which had been nagging him for a few years. How to find the average rank score statistics over all the possible rank vectors compatible with a set of observed intervals? Because the solution for right-censored data was painless yet elegant, it seemed that interval censored data should yield in the same way. It turned out that it would not. The problem, however, morphed from interesting puzzle to elaborate construction to panoramic landscape as I became acquainted with the work of order and graph theorists. This work presents a few snapshots of this landscape, taken from the two different angles of linear extension (or rank vector) sampling and of nonparametric estimation of the distribution function for interval censored data.

The Department of Statistics of the University of Auckland was in every way the supportive environment that I recalled from a first visit to Auckland in 1992, and everyone there deserves my gratitude. Some people deserve particular thanks for the material contribution they have made to the thesis, through discussion, critical comment or proofreading. I would like to thank on that front Dr. Andrew P. Balemi for critical comments on the structure of Chapter 4, Prof. Marston D. E. Conder for enlightening and lively discussions concerning minimal covers, Dr. Geoffrey Nicholls for sharing his knowledge of Monte Carlo methods, my lovely fiancée Ms. Sally Ross for patient proofing and reviewing, and Dr. Mohammad Salehi for useful conversations concerning sampling estimates. I would like to reserve special thanks to Prof. Alastair J. Scott for making possible these doctoral studies in such a congenial environment, and who as adviser has been supportive of the original tack this research took. The lion's share of my thanks has to go to Dr. Robert Gentleman, my

friend and supervisor and the man with the original idea, who believed from day one that statistics and order theory could be brought between the same set of covers. His enthusiasm for the topic, his flair for research, his unfailing personal support way beyond the call of supervisory duties and his immense availability in the midst of a demanding schedule have pulverized the myth of the supervisor-as-obstacle, and replaced it with a tale of collaboration and enabling. As a result, I close this work with the firm intention of pursuing this collaboration and continuing this line of research so long as it keeps supplying such striking results.

In spite of the collaborative component of the supervision process, the first person plural used in this dissertation is meant to be the "we of modesty" intended to include the reader and the sole writer.

Lastly, I wish to acknowledge the crucial financial support of three institutions: the New Zealand Marsden Fund, the Department of Statistics of the University of Auckland and the Fonds pour la formation de chercheurs et l'aide à la recherche du Québec.

Contents

List of Figures

List of abbreviations

CADLAG Continuité à droite, limite à gauche (continuity on the right, limit on the left)

CDF Cumulative distribution function.

MLE Maximum likelihood estimate.

NPMLE Nonparametric maximum likelihood estimate.

PF Probability function.

Introduction

Both in medical research and in industry, the analysis of failure time data lies at the heart of decisions of great importance. Unsurprisingly, statisticians and other practitioners in these fields go to great lengths to avoid unwarranted assumptions about the data with which they work. Given that frequentist models still hold great sway in scientific and technological fields, such efforts are worthwhile practically as they leave conclusions to the data rather than to analytical convenience, and foundationally as they avoid the informal mechanism of hidden assumptions which so often underlies – and undermines – frequentist data analysis.

As a result, nonparametric techniques are the most common techniques used in survival analysis. Survival analysis concerns primarily CDF (or survival function) estimation, rank-based statistics and hazard modelling. Data censoring often prevents the translation of conceptually simple survival models into simple analysis techniques. Censoring causes information loss by replacing exact data with intervals, bounded or not, known to contain the exact times of failure. Broadly speaking, it is this type of censoring which we call "interval censoring". Censored data generally represents imperfect knowledge about the times at which events have occurred, and as such will be described by intervals on the non-negative real half-line.

Different types of observed intervals may arise from different conditions on the observational process. Medical follow-up is a typical process by which data may be interval censored: a patient is monitored at preset periodic intervals for the development of a condition. The patient may miss some visits. The resulting data consists only of a time interval during which the condition has developed. If the patient has not developed the condition by the time of the last inspection, the condition is presumed to develop in a time interval unbounded above and bounded below by the time of the last inspection: this is the case of right-censoring. The patient may already have developed the condition by the time of the first inspection, in which case the condition is presumed to have developed between some plausible origin zero (e.g. birth) and the time of the first inspection: this is the case of left-censoring. Interval censored data includes as particular cases: exact data; right-censored data, where data consists of exact observations and intervals unbounded above; current status data, consisting of intervals either bounded below by zero or unbounded above; and doubly censored data, consisting of either exact observations, intervals bounded below by zero or intervals unbounded above. There are of course many cases of interval censored data which belong to none of these categories.

The loss of information caused by censoring prevents the use of plug-in estimators to perform analysis on the data. Plug-in estimators are the natural nonparametric estimators for exact data. What information is left in the observations by censoring must be coaxed, using different techniques, to yield statistics somehow similar to plug-in estimators while introducing a minimal number of assumptions.

We will show in this work that rank tests and nonparametric CDF estimation on interval censored data can be performed on an important invariant of the data: the underlying interval order. An interval order (Chapter 1, Definition 1.1) is a partially ordered set, members of which can be identified with intervals on the real line of the form (l_i, u_i), with the order relation $\prec$ given by $(l_i, u_i) \prec (l_j, u_j)$ whenever $u_i < l_j$. Interval orders already have a long history in the study of scheduling and preference models, but we believe this work represents the first instance of interval order theory being applied to the study of interval censored data.

Order theory arises as an important player in interval censored data analysis as a result of the following question: what is left of the data after censoring? Nonparametric techniques on exact data typically rely on the ordering of the data. Rank-based statistics are an obvious example, seeing as they rely strictly on the ranks of the observations. Order is a less obvious but important invariant as far as CDF estimation on exact data is concerned. To see this, note that we can separate the form of an empirical distribution function from the actual location of its jumps on the real line. So long as we know how many observations belong to each tie group and how these tie groups are ordered, we can safely predict how high and in what order jumps must occur in the estimate; porting the estimate to the real line is quite a separate activity. The information concerning the tie groups is exactly the partial order structure of the data, in the sense explained in Chapter 1. In much the same way, ordering properties are fundamental for both rank-based tests and nonparametric CDF estimation with interval censored data.

A general discussion of some fundamental assumptions of nonparametric statistical analysis for censored data is perhaps in order. Rank-based statistics are conceptually simple and provide a good illustration for this discussion. Let us assume initially that we are dealing with a two-sample setting $X = \{x_1, \ldots, x_{n_1}\}$ and $Y = \{y_1, \ldots, y_{n_2}\}$ with no two observations tied. We consider all pairs of observations of the form (x, y), $x \in X$, $y \in Y$, and assign a score to each of them. For definiteness, let us assume that we are dealing with the score $U(x, y)$, which takes on the value 1 if $x < y$ and -1 otherwise. We call these $U(x, y)$'s *U-scores.* Other scoring schemes are of course possible and well documented; however the simple U-score will serve our expository purposes. The Wilcoxon statistic is defined as the sum of the U-scores over all the pairs of observations as described above; under a null hypothesis of identical distributions for samples X and Y, its distribution does not depend on that of the data, and has an expected value of zero and a readily computable variance. It can be used to test against appropriate alternatives, which typically involve the location of the underlying unknown distributions. Lehmann (1975), Chapter 1, provides a survey of the Wilcoxon and related tests.

Ties in exact data may occur when the underlying distribution for failure times has atoms, but can also be viewed as a simple case of censoring of observations from a continuous distribution, where limited precision makes discrimination between close measurements impossible. Such precision-dependent censoring occurs within small intervals which overlap with one other only within equivalence classes, each such class consisting in a set of tied observations. The Wilcoxon statistic can be adapted for the case of exact data with ties by allowing $U(x, y) = 0$ if $x = y$. Generalizing from this technique, Gehan (1965b) proposed a generalized Wilcoxon statistic for right-censored data, almost immediately generalized further in Gehan (1965a) to deal with doubly censored data. The rationale he implemented was that if two observations $x \in X$ and $y \in Y$ overlap, then they are incomparable and should be assigned a U-score of zero. Such overlap is opposed to comparability, where two observations, censored or not, are known to be ordered in some definite way. Comparability occurs when the intersection of the intervals or points which form the observation is empty. We then have complete knowledge about the ordering of the event times. For the right-censored and doubly censored cases, Gehan (1965a,1965b) provided expressions for the expected value and variance of his generalized Wilcoxon statistic and showed its asymptotic normality under certain regularity conditions, conditionally on the observed pattern of censoring. The essential purpose of these conditions is to allow randomized assignment of observations to one sample or the other under the null hypothesis and the observed pattern of censoring, thus allowing the use of permutation arguments to determine the null distribution of the statistic. Such permutation arguments do not always hold, and in particular may be inapplicable when censoring patterns differ between the samples (Mantel, 1967). The basic idea of assigning a non-zero score only to comparable pairs of observations made the further generalization of the Wilcoxon statistic to interval censored data rather straightforward. This task was completed by, among others, Schemper (1984,1991) and Abel (1986).

Prentice (1978) approached rank tests for censored data from a different angle. Dealing with right-censored data, he considered all possible complete orderings of the data which are compatible with the observations. Prentice's basic assumption was that every rank vector compatible with the data has an equal probability of being the true rank vector of the data. Under this assumption, he proposed a general linear rank statistic consisting of the unweighted average of a rank score statistic (such as the Wilcoxon statistic) over the set of all rank vectors compatible with the data. Some amenable features of right-censored data allowed Prentice to produce closed form expressions for the expected value and variance of such statistics.

Gehan and his successors took the view of that incomparable observations can be ordered arbitrarily and that no other information concerning that ordering is available. Hence, for instance, an incomparable pair gets assigned a score of zero. We call this the *pure incomparability assumption.* By contrast, Prentice posited that every linear ordering consistent with the data had an equal chance of correctly representing the data's true ordering. We call this assumption the *underlying order assumption.* Both assumptions have merit as nonparametric foundations: simplicity of implementation for pure incomparability, and ease of interpretation for underlying order. In the case of exact data, with or without ties, the assumptions are in fact equivalent: a tied (incomparable) group of observations will contribute an equal number of compatible rank vectors for each possible pairwise ordering of its members. In more complex censored cases, these assumptions cease to be equivalent.

It is not our purpose in this work to critique or compare these assumptions in any definitive way. This reticence is attributable to the absence of a generally applicable methodology based on the underlying order assumption. Closed form expressions for expectation and variance of rank-based statistics, for instance, are unavailable in the more general case of interval censored data. Another problem exists in that the set of linear orderings compatible with the data gets very large very quickly with the number of observations, making enumeration impracticable. Self & Grossman (1986) pursued Prentice's reasoning and proposed to produce sample estimates of rank score statistics for interval censored data. Sample estimation requires, within this scheme, uniform random generation of rank vectors from the population of rank vectors compatible with the data. Though Self & Grossman (1986) proposed several algorithms to generate rank vectors uniformly, their algorithms do not achieve uniformity.

Thus our first goal in this work will be to propose a general methodology to deal with statistics on the set of linear orderings compatible with the data. From a combinatorial point of view, this is a larger problem than that of finding sampling estimates for rank-based statistics. Our second goal will be to uncover how current nonparametric estimation paradigms for censored data, namely self-consistency and maximum likelihood, extend the view on the CDF plug-in estimator which we described above.

In order to deal with both of these problems, we had to borrow terminology and tools from order theory, a particular flavour of combinatorial mathematics. We not only avoided unnecessary bushwhacking in doing so, but also rode on the crest of some striking mathematical constructions. The aspects of order theory we refer to in this work have been grouped in § 1 of Chapter 1. The notion of partial order is basic to our investigation; in particular, interval orders will serve to abstract interval censored data from the real line and identify important properties which are independent of their real representation. Linear extensions (§ 1.2) and maximal antichains (§ 1.3) and their structures are the first and possibly most important objects to be borne out of order theory for our purposes. It could be argued that the fundamental nonparametric assumption chosen to perform a statistical analysis will dictate which of these two objects will provide the entry point to the analysis: linear extensions for the underlying order assumption, and maximal antichains for the pure incomparability assumption.

Chapter 1 also includes a few original extensions of interval order theory, such as duals and double duals (Definition 1.7); minimal elements (Definition 1.8) a simple proof that the Petrie matrix of an interval order has full row-rank (Lemma 1.5); the notions of cover and minimal cover of maximal antichains (§ 1.5); and an extension of the characteristic matrix of interval orders (§ 1.7). Interval censored data is defined in § 2 of the same Chapter, along with a primordial motivation for the abstract interval order as a tool for statistical analysis. The central result of § 2 is Theorem 1.10, which couches a result of Peto (1973) and Turnbull (1976) identifying the form and location of CDF nonparametric maximum likelihood estimates in order theoretic terms.

As we have mentioned, there is currently no practical method to uniformly generate linear orderings compatible with an observed interval censored data set, which we will simply call linear extensions for the remainder of this Introduction. A general methodology to do so for any partially ordered set was proposed by Matthews (1991); it has not been implemented as computer software and is, in the words of the author "probably too slow for practical implementation". We had to delve into the structure of interval orders as defined by Fishburn (1973), and ultimately had to produce a small body of theory concerning covers and minimal covers of maximal antichains for interval orders. This theory is expounded in Chapter 2. Chapter 3 is devoted to the construction of a partition of the set of linear extensions into simple sets for which linear extensions or statistics on them are simple to generate. A pseudo-random generation method for this partition which relies on the theory of Chapter 2 is also proposed. Together, partition and generation method make sampling estimates based on the set of linear extensions easy to produce. These estimates include, but are not limited to, rank-based statistics.

Chapter 4 veers away from the heavily combinatorial topics of Chapters 2 and 3, and extends the work initiated in § 2 of Chapter 1 on estimation of the CDF for interval censored data. We consider self-consistency and nonparametric likelihood approaches to estimation. The fundamental nonparametric assumption behind these approaches is the pure incomparability assumption. Our first task in Chapter 4 is to define precisely what we mean by a CDF on the set of maximal antichains, and to show how it relates to the equivalence class of distribution functions which are covered by the NPMLE of the CDF. Uniqueness of the estimate on the maximal antichains ensues from properties of the maximal antichain structure of interval orders. In § 2, we revisit the notion of self-consistency in the light of the order structure of interval censored data. A relationship between covers of maximal antichains, self-consistency and nonparametric maximum likelihood estimation is expounded. Bounds on self-consistent estimates depending only on the maximal antichain structure of the data are also produced. Finally, we review statements from two works in the recent literature on interval censored data in the light of our analysis, and show how the order theoretic viewpoint helps in clarifying some issues. Section 3 in Chapter 4 aims at carrying over recently developed isotonic regression techniques for nonparametric estimation to the space of maximal antichains. Isotonic regression techniques are also known as convex minorant techniques. The application of isotonic regression on maximal antichains is explained for some cases of interval censored data.

Finally, this work closes with a brief discussion of research perspectives in the field of order theory and nonparametric analysis of interval censored data.

A note on the numbering system throughout this work: within a chapter, sections are identified by a single number and subsections by a number preceded by the section number. Equations, theorems, examples, figures and such, however, are referenced using the chapter number and an ordinal, to account for the fairly high frequency of inter-chapter references to these headings. This hybridized system was settled upon to avoid triple nesting while maintaining ease of reference. A section or subsection number not accompanied by a chapter number is a reference within the current chapter.

Chapter 1

Interval orders and interval censored data

Before describing interval censored data in § 2 and the problems we propose to address in this work, we will expose some aspects of order theory in § 1. Order theory is an active branch of combinatorial mathematics, which traces its theoretical foundations to the works of Szpilrajn (1930) and Dushnik & Miller (1941). Statisticians are rarely exposed to order theoretic concepts; nevertheless, we have decided to introduce material relevant to order theory at the outset of this work to mark clearly the break with classical interval censored data analysis, to hint at the richness of the approach and to provide the reader with a convenient reference to some definitions and theorems. The main connection between interval orders and interval censored data will be made explicit in § 2.

The reader is encouraged to refer to Appendix A, where some notational conventions are exposed, before proceeding.

1 Interval orders

The main object of our attention will be interval orders. Interval orders are partially ordered sets defined in such a way as to embody the natural ordering properties of sets of intervals on the real line. The abstract definition of interval orders will enable us to move beyond the real line to conceptually separate the ordering properties of the data from their real-valued representation. Some groundwork from the theory of partially ordered sets needs to be laid out before we can motivate the abstract representation exposed in Theorem 1.1. Definitions and discussions of linear extensions and maximal antichains will follow in § 1.2 and § 1.3. These two objects are basic to our approach.

The material in this section is partly adapted and extended from Fishburn (1985). Some of the statements taken from this source have been simplified inasmuch as we only deal with finite non-empty sets of real-valued intervals, and therefore with finite non-trivial interval orders. The theory of infinite interval orders is of obvious interest to the study of consistency of estimators for the CDF; however, we do not deal with the issue of consistency in the present study. Readers may wish to refer to Figures 1.4 and 4.1 to test the concepts of Chapter 1.

1.1 Partially ordered sets

Let $\underline{X} = (X, \prec)$ denote a partially ordered set, or *poset*. That is, there is a binary relation $\prec \subset X \times X$ which is both

- irreflexive: *if* $(x, y) \in \prec$ *then* $(y, x) \notin \prec$; and
- transitive: *if* $(x, y) \in \prec$ and $(y, z) \in \prec$, *then* $(x, z) \in \prec$.

We call $\prec$ a *precedence* or *ordering* relation. In the sequel, we will denote $(x, y) \in \prec$ by $x \prec y$, which is read "x precedes y".

We shall use $\sim$ to denote the *incomparability* relation for poset $(X, \prec)$ defined by

$$x \sim y \text{ if and only if } not(x \prec y) \text{ and } not(y \prec x).$$

The incomparability relation is the symmetric complement of $\prec$ with respect to $X \times X$. To indicate the relationship between $\prec$ and $\sim$, we will sometimes write $\sim = sc(\prec)$.

The graph $(X, \sim)$ is called the *interval graph* of $\underline{X}$. Under our definition, $x \sim x$ is always true. The term "incomparability relation" is sometimes reserved to denote the relation between distinct unordered elements.

In a real interval mapping of an interval order, as described later in Theorem 1.1, the relation $\sim$ can be thought of as relating two overlapping intervals, while the relation $\prec$ indicates the direction of the precedence relationship between two intervals which do not overlap.

We will be concerned with three types of partial orders.

Definition 1.1 *A linear or complete order is an ordered set* $(X, <)$ *such that* $x \not\sim y$ *for all* $x, y \in X$ *with* $x \neq y$.

A weak order is an ordered set $(X, \prec_0)$ *such that* $x \prec_0 y$ *implies that either* $x \prec_0 z$ *or* $z \prec_0 y$ *for all* $x, y, z \in X$.

An interval order is an ordered set $(X, \prec)$ *such that*

$$(a \prec x, b \prec y) \Rightarrow (a \prec y \text{ or } b \prec x), \qquad \text{for all } a, b, x, y \in X.$$

For completeness, it is worth setting down the definition of a fourth partial order.

Definition 1.2 *A semi-order is an interval order* $\prec$ *on* X *with the added property that whenever* $a \prec b \prec c$, *then either* $a \prec x$ *or* $x \prec b$ *for all* $a, b, c, x \in X$.

Definition 1.1 is more easily interpreted in terms of forbidden suborders. The Hasse diagrams of Figure 1.1 display the forbidden suborders for the above three orders. Hasse diagrams indicate precedence relationships $x \prec y$ by linking x to y with an arc or a path, with y appearing above x. Elements of ordered sets are indicated by circles.

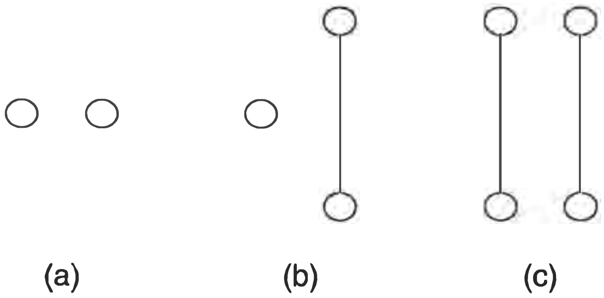


Figure 1.1: *Forbidden suborders in (a) linear orders (ties not allowed), (b) weak orders (ties allowed only under transitive incomparability) and (c) interval orders.*

Though we are primarily concerned with interval orders, linear and weak orders play an important role throughout our study. Linear orders induce a distinct rank on each of their elements, while weak orders induce a ranking with ties on their elements. That is to say, weak orders consist of linearly ordered subsets of X, each subset containing only mutually incomparable elements.

In a weak order, incomparability is an equivalence relation, and will consequently often be denoted by $\approx$ instead of $\sim$. If $(X, \prec)$ is a weak order, we will denote by $(X/\approx, \prec)$ the linear order induced on its equivalence classes. It should be noted that weak orders are a special case of interval orders, as linear orders are a special case of weak orders.

Unless otherwise specified, $\underline{X} = (X, \prec)$ will denote an interval order from this point on, and we will let $n = |X|$ (the cardinality of X) and $X = \{x_1, \ldots, x_n\}$, with arbitrarily ordered subscripts.

Interval orders embody the ordering properties of sets of intervals on the real line equipped with the usual linear order relation $<$, indeed of sets of intervals in any linear order. The following theorem, adapted from Theorems 2.6 and 2.8 in Fishburn (1985), will be sufficient for our purposes.

Theorem 1.1 *$(X, \prec)$ is an interval order if and only if there is a mapping H from X into open intervals in $\mathbb{R}$ such that $x \prec y \Leftrightarrow \sup H(x) \leq \inf H(y)$.*

Theorem 1.1 provides a simple interpretation for precedence in a real representation. The equivalent interpretation for incomparability is:

$$x \sim y \Leftrightarrow \sup H(x) > \inf H(y) \text{and} \sup H(y) > \inf H(x).$$

Alternatively, we can write: $x \sim y \Leftrightarrow H(x) \cap H(y) \neq \emptyset$. So long as the interval order is finite, the open intervals of Theorem 1.1 may be replaced by semi-open or closed intervals. These may include points, that is, intervals of the form $[a, a]$. Order theoretic results can therefore be applied to sets consisting of real intervals and real numbers, and thus in particular to interval censored data, as we will see in § 2.

The mapping H can be represented using the pair (h, ρ), where $h : X \to \mathbb{R}$ is the location function and $\rho : X \to \mathbb{R}^+ \cup \{0\}$ is the length function. We can then set $H(x) = (h(x), h(x) + \rho(x))$. Theorem 1.2 below shows that the interval order representation of a given set of real-valued intervals is invariant under a set of transformations which is larger than that of monotonic transformations.

Theorem 1.2 (Fishburn, 1985, Chapter 7) *Let $(X, \prec)$ be a finite interval ordered set. Define the relation $\sqsubset$ on non-empty, disjoint subsets of X by*

$$A \sqsubset B \text{ if } \sum_{a \in A} \rho(a) < \sum_{b \in B} \rho(b)$$

for all real-valued interval representations (h, ρ) of X. Then for every $\rho_* : X \to \mathbb{R}^+$ there is a function $h_* : X \to \mathbb{R}$ such that (h_*, ρ_*) is a representation of $(X, \prec)$ if and only if $\sum_{a \in A} \rho_*(a) < \sum_{b \in B} \rho_*(b)$ for all $A, B \subset X$ such that $A \sqsubset B$.

The invariant set of transformations of the real-valued intervals, however, depends upon the interval order being considered. An open question is whether the invariant set of transformations is in fact larger than that of monotonic transformations across all finite interval orders.

The representation theorem for semi-orders (Fishburn, 1973) characterizes them as having a real interval representation where all intervals are of the same length.

1.2 Linear extensions

Definition 1.3 *A linear extension of a partially ordered set* $(X, \prec)$ *consists of* $(X, <_\lambda)$*, where* $<_\lambda$ *is a linear order relation on* X *and* $\prec \subset <_\lambda$.

We will often treat linear extensions as ordered lists λ of a set rather than as an order relation $<_\lambda$. The two views are equivalent, with $\lambda = (x_{i_1}, x_{i_2}, \ldots, x_{i_n}) \Leftrightarrow x_{i_1} <_\lambda x_{i_2} <_\lambda \cdots <_\lambda x_{i_n}$.

Definition 1.4 *Let* $(X, \prec)$ *be a poset. We denote by* $\mathcal{L}(X, \prec)$ *the set of linear extensions given by*

$$\lambda \in \mathcal{L}(X, \prec) \Leftrightarrow \prec \subset <_\lambda,$$

where $<_\lambda$ *is the linear order relation induced by linear extension* λ.

Linear extension sets and partial orders thus form a simple concept lattice in the sense of Wille (1981). In a convenient abuse of terminology, we will say that λ belongs to, or is in, a partial order if λ belongs to its set of linear extensions.

Szpilrajn (1930) laid an early cornerstone of order theory:

Theorem 1.3 (Szpilrajn) *Every poset has a linear extension.*

Dushnik & Miller (1941) built on this foundation to show the deep relation between posets and their linear extension. Theorem 1.4 paraphrases their result.

Theorem 1.4 (Dushnik and Miller) *If* $\underline{X} = (X, \prec)$ *is a poset, then*

$$(\prec) = \bigcap_{\lambda \in \mathcal{L}(\underline{X})} (<_\lambda).$$

Rank vectors are objects which are equivalent to linear extensions but hold greater currency in statistics than in combinatorial mathematics. Given an arbitrary linear order of the elements of

X (e.g. subscript ordering from some index set), the rank of an element x in linear extension λ will consist in $|\{z; z <_\lambda x\}| + 1$. Though our focus will be on linear extensions, in keeping with the terminology of order theory, the reader should bear in mind that linear extensions and rank vectors embody the same ordering information.

In general, one can talk of extensions of a poset other than linear. We have mentioned the underlying order assumption in the Introduction as an assumption that places equal weight a priori on each linear extension, in a modern version of the principle of insufficient reason. Though we will not address the matter in this study, a weak as opposed to linear order may underlie the interval order and would certainly represent a more general model. In this case, determining weak extensions may be more appropriate.

We will reserve further discussion of the set of linear extensions of interval orders until Chapter 3, where we deal with sampling statistics over this set.

1.3 Maximal antichains

Chains in partially ordered sets play only an ancillary role in the present study. They are often the starting point from which combinatorists approach partially ordered sets, and we therefore set down their definition for completeness.

Definition 1.5 *A chain in a partially ordered set* $(X, \prec)$ *is a subset* $K \subset X$ *such that* $(K, \prec \cap (K \times K))$ *is a linear order.*

A maximal chain is a chain not properly contained in any other chain.

A maximum chain is a maximal chain of largest possible cardinality.

Maximal antichains, by contrast, are the most convenient tool to describe the ordering properties of interval order elements as they lend themselves to statistical analysis.

Definition 1.6 *An antichain in a partially ordered set* $(X, \prec)$ *is a subset* $M \subset X$ *such that* $x \sim y$ *for all* $x, y \in M$*.*

A maximal antichain is an antichain not properly contained in any other antichain.

A maximum antichain is a maximal antichain of largest possible cardinality.

The set of maximal antichains of a partial order $\underline{X}$ is usually denoted $MA(\underline{X})$, but we will express it by $\mathcal{M}$ to lighten the notation.

A crucial characterization of interval orders is that there exists a natural linear ordering on their set of maximal antichains (see Behrendt, 1988, and Fishburn, 1985, § 3·3). Specifically, if $M_a, M_b \in \mathcal{M}$ and we define the relation $<$ over $\mathcal{M}$ by

$$M_a < M_b \Leftrightarrow (M_a \setminus M_b) \prec (M_b \setminus M_a),$$

where the relation $\prec$ is extended to subsets of X, then $(\mathcal{M}, <)$ is a linear order. It will therefore make sense to use terminology such as the largest or smallest element in a set of maximal antichains. A fact of note which follows from definitions is that the maximal antichains of $\underline{X}$ are also the maximal cliques of the interval graph $(X, \sim)$. A clique in a graph is a complete subgraph, that is, a subgraph where every pair of nodes is linked by an arc.

Setting $m = |\mathcal{M}|$, called the *magnitude* of X, we will assign subscripts $i = 1, \ldots, m$ to the elements of $\mathcal{M}$ according to their linear ordering, that is with $M_i < M_j \Leftrightarrow i < j$. Minima and maxima are thus well-defined elements over subsets of $\mathcal{M}$.

It is useful to consider ordering properties of a partial order from the point of view of maximal antichains. To do so, consider, for a partial order $(X, \prec)$, the equivalence relation $\approx$ induced by $\sim = sc(\prec)$:

$$x \approx y \Leftrightarrow \{z; x \sim z\} = \{z; y \sim z\}.$$

The corresponding equivalence classes are denoted $X/\sim$ (Fishburn, 1985, Chapter 1). It is a trivial observation that if $x \approx y$ in a partial order, then x and y must belong to exactly the same maximal antichains. This observation yields alternative characterizations of weak and linear orders. A weak order on X is such that $\approx = \sim$, so that its maximal antichains are precisely the equivalence classes $X/\sim = X/\approx$, while a linear order is such that the equivalence classes $X/\sim$ are all singletons.

1.4 Petrie representation

The *Petrie matrix* of an interval order is an indicator matrix relating the elements of X to the maximal antichains of $\mathcal{M}$. Specifically, the Petrie matrix of $\underline{X}$ is given by $\mathbf{A} = [\alpha_{ij}] \in \{0,1\}^{m \times n}$, where

$$\alpha_{ij} = \mathbf{1}[x_j \in M_i] \tag{1.1}$$

See Example 1.1. This definition along with our ordered subscripting convention on $\mathcal{M}$ imply that the rows of **A** are ordered similarly to the elements of $\mathcal{M}$. Under this ordering, all interval orders will have a unique Petrie matrix representation up to the subscript ordering of X or, equivalently, up to ordering of the columns of **A**. In the order theoretic representations of interval censored data seen in Theorem 1.10, our α_{ij} notation is the transpose of the one which is more usual in the statistical literature.

Elements of X have properties which depend on the maximal antichains of $\underline{X}$ and are most easily assessed via the Petrie matrix of X. The basis of these properties is the element's dual.

Definition 1.7 *The dual of $x \in X$ is given by*

$$x^* = \{M \in \mathcal{M}; x \in M\}\,.$$

The cardinality of x^* is called the length of x.

The double dual of a maximal antichain $M \in \mathcal{M}$ is given by

$$M^{**} = \bigcup_{x \in M} x^*.$$

The linear ordering of $\mathcal{M}$ causes the maximal antichains forming the dual of any given element to be consecutive. This property is commonly called the *consecutive-1's property* when applied to the columns of matrix **A** (Golumbic, 1980, Chapter 8), whence the name of Petrie matrix (Fishburn, 1985, § 3.3). The Petrie matrix is also known as the *clique matrix* of the interval graph $(X, \sim)$ (Golumbic, 1980, Chapter 3), with the proviso that we require the rows of the clique matrix to be ordered according to the maximal antichain or clique linear ordering, thus explicitly preserving the consecutive-1's property.

The propositions $x \in M$ and $M \in x^*$ equivalently express the fact that x is covered by or contained within maximal antichain M. The dual of an element – and therefore the element itself – is isomorphic to its corresponding column in the Petrie matrix of X. The consecutive-1's property of the Petrie matrix allows an alternative compact representation for the elements of X in terms of the first and last maximal antichains in their duals. Specifically, we can say that, under known m, x_j^* is isomorphic to the integer pair (l_j, u_j), $j = 1, \ldots, n$, where $l_j = \text{argmin}_i\,[\alpha_{ij} = 1]$ and $u_j = \text{argmax}_i\,[\alpha_{ij} = 1]$. We will sometimes call this sequence of integer pairs the *Petrie pairs* of $\underline{X}$.

Under the linear ordering of the maximal antichains of an interval order, we will write for simplicity $M < x^*$ for $M < \min x^*$ and $x^* < M$ for $\max x^* < M$.

The structure of an element's dual determines some of its fundamental properties.

Definition 1.8 *An element of X is called simplicial if it belongs to a single maximal antichain, i.e. if $|x^*| = 1$. A maximal antichain containing a simplicial element is called essential.*

An element of X is called universal if it belongs to every maximal antichain, i.e. if $|x^*| = m$.

An element of X is called minimal if its dual properly contains the dual of no element in X.

The term "simplicial" derives from the fact that, in the interval graph $(X, \sim)$, the neighborhood of a simplicial element is complete (Fishburn, 1985, Chapter 3).

Universal elements are incomparable with every element of X. Minimal elements are discussed in greater detail in Chapter 2.

Example 1.1 *Consider the following real open intervals on an arbitrary scale:*

$$x_1 = (0.50, 2.00),\ x_2 = (1.50, 4.50),\ x_3 = (1.75, 3.50),$$

$$x_4 = (2.75, 7.00),\ x_5 = (5.50, 8.00),\ x_6 = (4.00, 6.50).$$

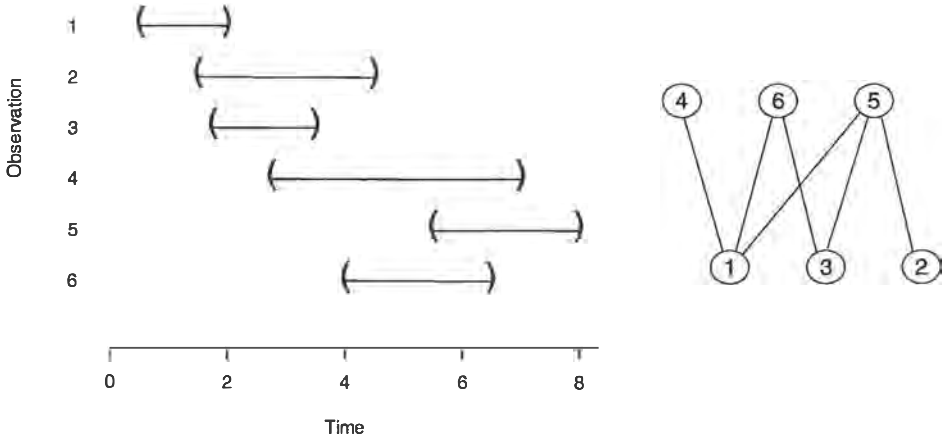


Figure 1.2: *Example intervals and Hasse diagram of corresponding interval order.*

There are 4 maximal antichains in the interval order induced by these intervals: $M_1 = \{x_1, x_2, x_3\}$, $M_2 = \{x_2, x_3, x_4\}$, $M_3 = \{x_2, x_4, x_6\}$ *and* $M_4 = \{x_4, x_5, x_6\}$.

The Petrie matrix for these intervals is given by

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix},$$

where row i *corresponds to maximal antichain* M_i, $i = 1, 2, 3, 4$, *and column* j *corresponds to* x_j, $j = 1, \ldots, 6$. *The only minimal elements in this set of intervals are* x_1 *and* x_5. *Note that minimality does not relate to the length of the interval in the real representation.*

The definition of maximal antichains and their linear ordering cause every maximal antichain to be the smallest member of some dual and the largest member of some (possibly different) dual. We will refer to this property as the starting/ending property of interval orders. As a result of this property, the Petrie matrix of an interval order will have at least one sequence of ones start on every row, and at least one sequence of ones end on every row. If a single sequence is involved, then its corresponding element must be first and last of its sequence: it is therefore simplicial and its maximal antichain is essential.

The starting/ending property forces the existence of at least two simplicial elements in the interval order, with respective duals $\{M_1\}$ and $\{M_m\}$, making M_1 and M_m essential in any interval order of magnitude m. This observation leads in turn to the following Lemma, to which we will return in the context of nonparametric maximum likelihood theory.

Lemma 1.5 *If* $\mathbf{A} \in \{0,1\}^{m \times n}$ *is the Petrie matrix of an interval order with cardinality* n *and magnitude* m, *then* $\operatorname{rank}(\mathbf{A}) = m$.

Proof. Consider the equation

$$\mathbf{A}'\beta = \mathbf{0}, \tag{1.2}$$

where $\beta = [\beta_1, \ldots, \beta_m]$. There are columns of $\mathbf{A}$ which correspond to the simplicial elements in M_m, say r of them, with $r \geq 1$. Each of these columns is equal to $[0, 0, \ldots, 0, 1]'$, and therefore $\beta_m = 0$. Assume, since the ordering of the columns of $\mathbf{A}$ can be arbitrary, that the columns of $\mathbf{A}$

described above correspond to elements $x_{n-r+1}, x_{n-r+2}, \ldots, x_n$ of X and occur in that same order in **A**. Then we can rewrite (1.2) as

$$\left[\begin{array}{c|c} \mathbf{A}_\circ' & \mathbf{a}_{n-r} \\ \hline \mathbf{0}_{(n-r)\times(m-1)} & \mathbf{e}_r \end{array}\right] \left[\begin{array}{c} \beta_\circ \\ \hline 0 \end{array}\right] = \mathbf{0}_m \Leftrightarrow \mathbf{A}_\circ'\beta_\circ = \mathbf{0}_{m-1}$$

for some arbitrary $\mathbf{a}_{n-r} \in \{0,1\}^{n-r}$, and where $\mathbf{e}$ is a vector consisting all of one's. Putting $X_\circ = X \setminus \{x_{n-r+1}, x_{n-r+2}, \ldots, x_n\}$, it is easy to see that $\mathbf{A}_\circ \in \{0,1\}^{(m-1)\times(n-r)}$ is the Petrie matrix of interval order $(X_\circ, \prec \cap [X_\circ \times X_\circ])$. It retains the consecutive-ones property, the linear ordering of its rows, and the essentiality of its last maximal antichain $M_{\circ,m-1}$, since the dual of some element of X must start at M_{m-1} by the starting/ending property.

Repeating the above process $m-1$ more times on the Petrie matrices of successively reduced interval orders shows that $\beta_m = \beta_{m-1} = \cdots = \beta_1 = 0$, which in turn shows **A** to be of full row-rank.

□

1.5 Covers of maximal antichains

We will often refer in this work to covers of maximal antichains, or simply covers. They are an important concept, both in the context of rank score tests and related statistics and in that of nonparametric CDF estimation.

Definition 1.9 *For* $\underline{X} = (X, \prec)$ *an interval order, we call* $\mathcal{W} \subset \mathcal{M}$ *a cover of* X *if* $X = \bigcup_{M \in \mathcal{W}} M$. *We will call* $\mathcal{W}$ *a minimal cover of* X *if no proper subset of* $\mathcal{W}$ *is a cover of* X. *A minimal cover is a minimum cover if it has lowest possible cardinality.*

For a given interval order, an essential maximal antichain has an alternative definition as a maximal antichain which belongs to every cover. Equivalently, it can be characterized as belonging to every minimal cover.

Covers of maximal antichains for interval orders possess an important property of invariance with respect to interval ordering.

Lemma 1.6 *Let* $(X, \prec)$ *be an interval order. Then if* $\mathcal{W} \subset MA(\underline{X})$ *(properly) is a cover for* X, *there exists* $\prec_{\mathcal{W}}$ *such that*

- $(\prec) \subset (\prec_{\mathcal{W}})$ *properly,*

- $(X, \prec_{\mathcal{W}})$ *is an interval order and*
- $\mathcal{W} = MA\,(X, \prec_{\mathcal{W}})$.

Proof. It is enough to construct $\prec_{\mathcal{W}}$ for $\mathcal{W} = MA\,(\underline{X}) \setminus \{M\}$, for some non-essential $M \in MA\,(\underline{X})$. Essential maximal antichains can evidently not be removed, since the resulting set would not be a cover. Every cover of X can be generated by repeating the process of removing appropriately chosen non-essential maximal antichains.

Define then $\circ = \{(x, y) \in X \times X; \max x^* = M \text{ and } \min y^* = M\}$, and set $\prec_{\mathcal{W}} = \prec \cup \circ$. It is easily verified that $\prec_{\mathcal{W}}$ satisfies the statement of the Lemma. □

Lemma 1.6 states that covers induce interval orders by creating new precedence relations where maximal antichains forced incomparability in the original interval order. A simple corollary to Lemma 1.6 is that in an interval order induced by a minimal cover, every maximal antichain is essential. This result will be used in discussions in Chapters 2 and 4. Referring to the proof of Lemma 1.5, we note that removing an essential maximal antichain also induces an interval order, though on a proper subset of X.

The general theory of minimal covers forms the substance of Chapter 2.

1.6 Endpoint ordering of interval orders

Define the composition of two binary relations $\circ_1$ and $\circ_2$ by

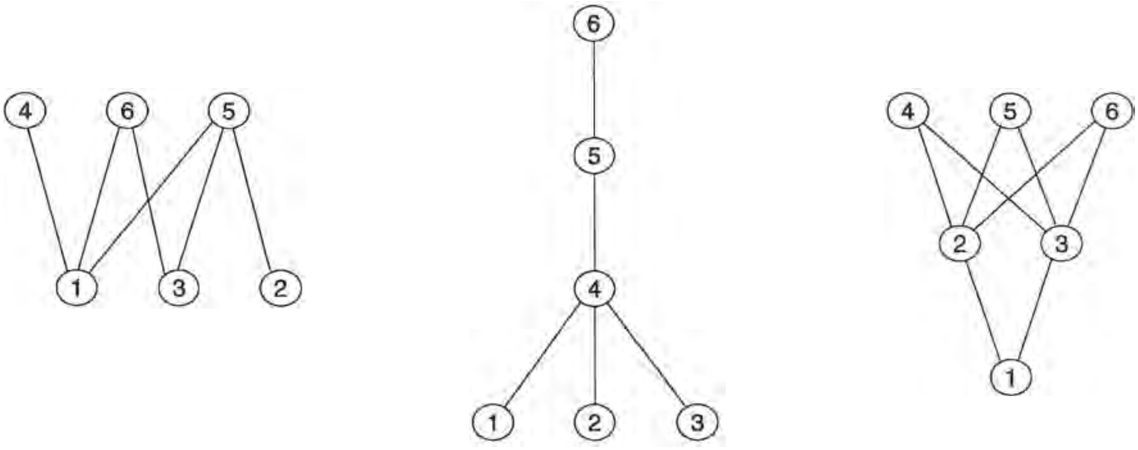
$$(\circ_1)(\circ_2) = \{(a, b); a \circ_1 z \text{ and } z \circ_2 b \text{ for some } z\}.$$

Fishburn (1985) (§ 2.2) uses composition of relations to construct a weak order on the abstract version of interval endpoints in an interval order.

For $(X, \prec)$ an interval order and $\sim$ the symmetric complement of $\prec$, let $\prec^- = (\sim)(\prec)$ and $\prec^+ = (\prec)(\sim)$. In an open real representation of an interval order, $\prec^-$ and $\prec^+$ are weak orderings on the left and right endpoints respectively (Fishburn, 1985, § 2.2). In this weak ordering, left (right) endpoints belonging to the same equivalence class are incomparable, even though a precedence relation may exist between them in a real-valued representation.

Example 1.2 *Consider the open intervals of Example 1.1. The real endpoints of these intervals are linearly ordered. Specifically, let* $<^-$ *and* $<^+$ *denote these linear orders, and* i *represent interval* x_i, $i = 1, \ldots, 6$. *Then* $1 <^- 2 <^- 3 <^- 4 <^- 6 <^- 5$ *and* $1 <^+ 3 <^+ 2 <^+ 6 <^+ 4 <^+ 5$.

However, the $\prec^-$ *and* $\prec^+$ *orderings are the weak orderings illustrated in Figure 1.3.*



(a) (b) (c)

Figure 1.3: *Hasse diagrams for (a) interval ordering* $\prec$, *(b) left-endpoint equivalence class weak ordering* $\prec^- = (\sim)(\prec)$ *and (c) right-endpoint equivalence class weak ordering* $\prec^+ = (\prec)(\sim)$.

We write X^- and X^+ to distinguish the elements of X as they are associated with their left or their right endpoint respectively. We distinguish elements of these two sets using the same superscripts.

We can now define a weak order on $X^- \bigcup X^+$, denoted $\prec_0$, and called the *conjoint weak order* (Fishburn 1985, § 2.2). This relation orders both the left and the right endpoints of the intervals in a natural way, namely, for $x, y \in X$,

1. $x^- \prec^0 y^- \quad \Leftrightarrow \quad x \prec^- y$
2. $x^+ \prec^0 y^+ \quad \Leftrightarrow \quad x \prec^+ y$
3. $x^+ \prec^0 y^- \quad \Leftrightarrow \quad x \prec y$
4. $x^- \prec^0 y^+ \quad \Leftrightarrow \quad x \prec y$ or $x \sim y$

We let $\approx^a = \sim^a$ be the incomparability relations for $(X, \prec^a)$, $a = -, +, 0$. Theorem 1.7 spells out the structure of $(X^- \bigcup X^+)/\approx^0$ as it relates to $X^-/\approx^-$ and $X^+/\approx^+$.

Theorem 1.7 (Fishburn, 1985, § 2.3) *Suppose* $(X, \prec)$ *is a finite interval order of magnitude* m. *Then* $X^-/\approx^-$ *and* $X^+/\approx^+$ *both have cardinality* m. *Moreover if we let* $(X^-/\approx^-, \prec^0) = \{L_1^- \prec^0 L_2^- \prec^0 \cdots \prec^0 L_m^-\}$ *and* $(X^+/\approx^+, \prec^0) = \{R_1^+ \prec^0 R_2^+ \prec^0 \cdots \prec^0 R_m^+\}$, *then*

$$L_1^- \prec^0 R_1^+ \prec^0 L_2^- \prec^0 R_2^+ \prec^0 \cdots \prec^0 L_m^- \prec^0 R_m^+.$$

We put $L_i = \{x; x^- \in L_i^-\}$ and $R_i = \{x; x^+ \in R_i^+\}$, for $i = 1, \ldots, m$. The fact that $|(X^-/\approx^-)| = |(X^+/\approx^+)| = m$ can be deduced from

$$\mathcal{M} = \left\{ \bigcup_{i=1}^{j} L_i \setminus \bigcup_{i=1}^{j-1} R_i; j = 1, \ldots, m \right\} \tag{1.3}$$

In applications, it is useful to recognize that (1.3) embodies a method for computing the set of maximal antichains of an interval ordered set. The reader familiar with Turnbull's treatment of nonparametric maximum likelihood estimation for interval censored data may wish, at this point, to compare Theorems 1.9 and 1.10 with (1.3).

1.7 Characteristic matrix of an interval order

Using the equivalence classes $L_i, R_i, i = 1, \ldots, m$ derived from Theorem 1.7, we define the $m \times m$ *characteristic matrix* χ of $(X, \prec)$ in the following manner:

$$\chi_{ij} = \left| L_i \bigcap R_j \right|. \tag{1.4}$$

Since the L_i and R_i both partition X, matrix χ will have at least one non-zero entry in every row and column. This requirement can be shown to be equivalent to the starting/ending property of maximal antichains in an interval order. Theorem 1.7 guarantees that χ will be upper-triangular.

Our definition of the characteristic matrix differs from Fishburn (1985, § 2.3), where it is defined as the indicator matrix $\mathbf{1}[L_i \cap R_j \neq \emptyset]$. This simpler version of the characteristic matrix plays a role in extremization problems involving the cardinality of the interval order and the number of distinct interval lengths required in its real-valued representations (Fishburn, 1985, Chapter 8). It plays no obvious role in the analysis of interval censored data.

The characteristic matrix of an interval order is easily related to its Petrie matrix.

Lemma 1.8 *Let* $\mathbf{K} = [c_{ik}] = [\mathbf{1}(i \geq k)]$, $i, k = 1, \ldots, m$, *(the cumulative sum transformation) so that* $\mathbf{K}^{-1} = \mathbf{1}[i = k] - \mathbf{1}[i - 1 = k]$ *(the backward difference transformation).*

Then the characteristic matrix of an interval order with Petrie matrix A is given by

$$\chi = \mathbf{K}' \circ \mathbf{K}^{-1}\mathbf{A}\mathbf{A}'\mathbf{K}^{-1} \tag{1.5}$$

that is, the upper triangular part of $\mathbf{K}^{-1}\mathbf{A}\mathbf{A}'\mathbf{K}^{-1}$.

Moreover, the characteristic and Petrie matrices of an interval order are in 1-1 correspondence.

Proof. Let $n_{ik} = (\mathbf{A}\mathbf{A}')_{ik}$, and define also $n_{0k} = n_{i,m+1} = n_{0,m+1} = 0$, $i, k = 1, \dots, m$. Obviously, $n_{ik} = \#(M_i \cap M_k)$, $i, k = 1, \dots, m$.

Using the definition of maximal antichains provided in (1.3), it is a simple matter to verify that, for $i \leq k$, $L_i \cap R_k = (M_i \cap M_k) \setminus [(M_{i-1} \cap M_k) \bigcup (M_i \cap M_{k+1})]$. Since the subtracted set is a subset of $M_i \cap M_k$ (a consequence of $i \leq k$) and since $|(M_{i-1} \cap M_k) \bigcup (M_i \cap M_{k+1})| = |M_{i-1} \cap M_k| + |M_i \cap M_{k+1}| - |M_i \cap M_k|$, we get

$$\chi_{ik} = |L_i \cap R_k| = n_{ik} - n_{(i-1)k} - n_{i(k+1)} + n_{(i-1)(k+1)} \tag{1.6}$$

for $i, k = 1, \dots, m$. Equation (1.5) is now easily verified.

To show that $\mathbf{A}$ and χ are in 1-1 correspondence, first note that, since $\mathbf{A}$ has full row-rank by Lemma 1.5, a simple application of the singular value decomposition (see, for instance, Horn & Johnson, 1985, § 7.3) shows that $\mathbf{A}$ is fully determined by $\mathbf{A}\mathbf{A}'$. Then it is enough, heuristically, to recognize that the information contained in $\chi^* = \mathbf{K}^{-1}\mathbf{A}\mathbf{A}'\mathbf{K}^{-1}$ is also contained in χ, since $\mathbf{K}$ is nonsingular. To see this, note that if $j = i - 1$, the expression for χ_{ij} given by (1.6) corresponds to the negative of the cardinality of the symmetric difference of M_j and M_{j+1}, while if $i > j + 1$, the same expression once again corresponds to $|M_i \cap M_j|$. Hence the lower left-hand $(m-1) \times (m-1)$ submatrix of χ^* is symmetric about the subdiagonal of χ^*, with the subdiagonal itself given by

$$\chi^*_{i,i-1} = -|M_{i-1} \Delta M_i| = -\left(\sum_{k=1}^{i-1} \chi_{k,i-1} + \sum_{k=i}^{m} \chi_{i,m} \right)$$

for $i = 2, \dots, m$, wholly defined in terms of elements of χ.

More formally, simple manipulations show that with vector $\mathbf{k} = \left[\left(\mathbf{I} - \mathbf{K}^{-1}\right)^{-1} \chi + \left(\mathbf{I} - \mathbf{K}^{-1}\right) \chi' \right] \mathbf{e}$, then we have $\mathbf{A}\mathbf{A}' = \chi + (\mathbf{I} - \mathbf{K}^{-1})(\mathbf{D}_{\mathbf{k}} + \chi')(\mathbf{I} - \mathbf{K}^{-1})$. □

A simple interpretation of the characteristic matrix brought to light by the proof of Lemma 1.8 is that, for $i \leq k$,

$$\chi_{ik} = \left| \left\{ x \in X; x^* = \{M_i, M_{i+1}, \dots, M_k\} \right\} \right|. \tag{1.7}$$

The characteristic matrix is an invariant of the interval order under all its real-valued interval representations.

Example 1.3 *The characteristic matrix of the interval order in Example 1.1 is*

$$\chi = \begin{bmatrix} 1 & 1 & 1 & 0 \\ & 0 & 0 & 1 \\ & & 0 & 1 \\ & & & 1 \end{bmatrix}.$$

2 Interval censored data and interval orders

We show in this section how the main features of interval censored data are captured by its order theoretic representation. We first state the form of the data in § 2.1. In § 2.2, we briefly introduce the nonparametric likelihood function for the CDF of interval censored data. Our only purpose in doing so is to motivate the order theoretic approach by showing that the classical result of Peto (1973) and Turnbull (1976) concerning the support of the CDF NPMLE in fact shows that estimation of the CDF can be performed on the set of maximal antichains of the data. A fuller discussion of nonparametric estimates of the CDF is reserved for Chapter 4. We also describe some special, well-studied cases of interval censored data in terms of their maximal antichain structure.

2.1 The data model

We are interested in a random event time T from a distribution F having support on $[0, +\infty)$. A realization of event time t_j is observed exactly with probability γ_j, $0 \leq \gamma_j \leq 1$, for $j = 1, \ldots, n$, corresponding to a continuous inspection process. Otherwise the event time realization is not observed exactly; under inspection point process Q_j, the observed data will consist of the last inspection time prior to the event and the first inspection time after the event. The process Q_j consists of inspection times $q_{j,k}$, $k = 1, \ldots, K_j$, to which initial and final inspection times $q_{j,0} = 0$ and $q_{j,K_j+1} = +\infty$ are added for convenience. With probability $1 - \gamma_j$, the observation for individual j will therefore be the open interval $x_j = (l_j, r_j) = (q_{j,k_0}, q_{j,k_0+1})$, for some k_0 depending on j.

An exact observation can be viewed as a very small open interval, and indeed usually is, since so-called exact measurements are in fact of finite precision. Intervals themselves do not need to be open for Theorem 1.1 to apply. The data description remains similar on an order theoretic level, and cumbersome notation is avoided.

We make two important assumptions which ensure that the censoring remains uninformative.

Assumption I *The inspection processes Q_j are all independent of F.*

Assumption II *No event occurs with positive probability at an atom in any of the inspection processes.*

Interval censoring as described above includes several common data models as special cases:

- Exact data;
- Right-censored data (see, e.g., Kaplan & Meier (1958), consisting of exact data points along with intervals of the form $(l_j, +\infty)$.
- Current status data, also called interval censored data case I (see, e.g., Groeneboom & Wellner (1992), consisting entirely of intervals of the form $(0, r_j)$ and $(l_j, +\infty)$.
- Doubly censored data (see, e.g., Turnbull, 1974, Gu & Zhang, 1993, Mykland & Ren, 1996, Wellner & Zhan, 1997, consisting of exact data points along with intervals of the form $(0, r_j)$ and $(l_j, +\infty)$. Readers familiar with the above authors will recognize that our approach to doubly censored data differs from that usually used in the literature.

2.2 Order theoretic representation of interval censored data

Assumptions I and II above ensure that estimation and inference can be conditioned on the inspection times at no cost in terms of information. The censoring information and the observed data will be split into two components: the ordering structure, on which most of the estimation and inference can be done; and the mapping $H_{\mathcal{M}}$ defined in Theorem 1.10, which enables us to port our results to the real line.

We can specify the likelihood function for the CDF F as follows:

$$L(F) = \prod_{j=1}^{n} \left[F(r_j) - F\left(l_j^-\right)\right].$$

Let $\{b_s\}_{s=0}^{2(n+1)}$ denote the ordered elements of $0, \{l_j\}_{j=1}^{n}, \{r_j\}_{j=1}^{n}$ and $+\infty$. Peto (1973) and Turnbull (1976) both showed the following result.

Theorem 1.9 *Let* (l_j, r_j), $j = 1, \ldots, n$, *be an interval data set, and consider the subset* $\{b_a\}_{a=1}^{2m}$ *of these points such that*

$$b_a \in \begin{cases} \{l_j\}_{j=1}^{n} \bigcup \{0\} & \text{if } a \text{ is odd, and} \\ \{r_j\}_{j=1}^{n} \bigcup \{+\infty\} & \text{if } a \text{ is even,} \end{cases} \tag{1.8}$$

and there are no other elements of $\{0, \{l_j\}_{j=1}^{n}, \{r_j\}_{j=1}^{n}\}$ *between* b_a *and* b_{a+1} *for any other* a.

Then outside of the intervals (b_{2a-1}, b_{2a}), $a = 1, \ldots, m$, *the NPMLE of* F *must remain constant and within these intervals, it is unidentifiable.*

The following theorem shows how the Petrie representation captures the combinatorial structure of interval censored data which is relevant to nonparametric likelihood estimation. The link between the maximal antichain structure of interval orders and the support of the NPMLE of F shows how the estimation problem, conditioned on the censoring pattern, can be viewed as the estimation of a probability measure on a discrete set.

Theorem 1.10 *Let* $\{(l_j, r_j); j = 1, \ldots, n\}$ *be a set of positive real valued open intervals, and let* $\underline{X} = (X, \prec)$ *be its interval order representation, putting* $H(x_j) = (l_j, r_j)$ *as in Theorem 1.1. Define the mapping* $H_{\mathcal{M}}$ *from* $\mathcal{M}$ *to positive real-valued open intervals by*

$$H_{\mathcal{M}}(M) = \bigcap_{x \in M} H(x) \tag{1.9}$$

Then $H_{\mathcal{M}}$ *is one-to-one, that is, the set of intervals identified in Theorem 1.8 and* $\mathcal{M}$ *are in one-to-one correspondence.*

Proof. (The following argument relies implicitly on Theorem 1.1 and the short discussion which accompanies it.) We first show that to every maximal antichain of $\underline{X}$ there corresponds an interval of the form (1.8) in Theorem 1.9. Let $M \in \mathcal{M}$ be a maximal antichain. Let $u = \text{argmax}_j \{l_j : x_j \in M\}$ and $v = \text{argmin}_j \{r_j : x_j \in M\}$. Since $x_u \sim x_v$, we have $l_u < r_v$ and $l_v < r_u$, and therefore $l_v \leq l_u < r_v \leq r_u$. Suppose some element x_w has a corresponding left or right endpoint lying strictly between l_u and r_v. If $l_u \leq l_w \leq r_v$, then $x_w \prec x_u$ by definition of u, so $r_u < l_w$. But then $r_v \leq r_u < l_w$, a contradiction. Similarly, no right endpoint may lie between l_u and r_v. Hence for every maximal antichain there is an interval of the form (1.8).

We next show the converse statement. Consider an interval (b_{2i-1}, b_{2i}) of the form (1.8) and let K be the set of observations such that $x_j \in K$ if and only if $(l_j, r_j) \cap (b_{2i-1}, b_{2i}) \neq \emptyset$. Since there are no endpoints between b_{2i-1} and b_{2i}, it must be that $(b_{2i-1}, b_{2i}) \subset (l_j, r_j)$ for every $x_j \in K$; hence all (l_j, r_j) overlap, so that K is an antichain. Suppose that K is not a maximal antichain. Then there exists a maximal antichain, M, such that K is a proper subset of M. Let $x^* \in M \setminus K$, with corresponding real representation (l^*, r^*). Since $x^* \notin K$, $(l^*, r^*) \cap (b_{2i-1}, b_{2i}) = \emptyset$. Let x_u denote the element of K with left endpoint equal to b_{2i-1} and let x_v denote the element of K with right endpoint equal to b_{2i}. Then either $x^* < x_u$ or $x^* > x_v$; but all three must be in M, and therefore all pairwise incomparable. Therefore there is no $x^* \in M \setminus K$, so that K is in fact maximal. □

It is customary, when dealing with the product-limit estimator of the survival function derived by Kaplan & Meier (1958), to state that the estimate is undefined on the right of the last observation if that observation is (right-)censored. Theorems 1.9 and 1.10 show that this is a particular case of the unidentifiability of the CDF within the intervals which compose $H_{\mathcal{M}}(\mathcal{M})$.

Example 1.4 *Let the intervals of Example 1.1 represent interval censored data. Each observation is an open interval, on an arbitrary time scale, within which the event of interest is known to have occurred. The 4 maximal antichains in the interval order induced by these intervals are illustrated in Figure 1.4 by the shaded boxes. The maximal overlap between the intervals in a single maximal antichain is the mapping of that maximal antichain to a real interval via $H_{\mathcal{M}}$, viz.*

$$H_{\mathcal{M}}(M_1) = (1.75, 2) \quad H_{\mathcal{M}}(M_2) = (2.75, 3.5)$$
$$H_{\mathcal{M}}(M_3) = (4.0, 4.5) \quad H_{\mathcal{M}}(M_4) = (5.5, 6.5)$$

The linear ordering of the maximal antichains as well as the result of Theorem 1.10 are evident in Figure 1.4. The set $H_{\mathcal{M}}(\mathcal{M})$ consists in the intervals where the NPMLE can put mass (see Chapter 4).

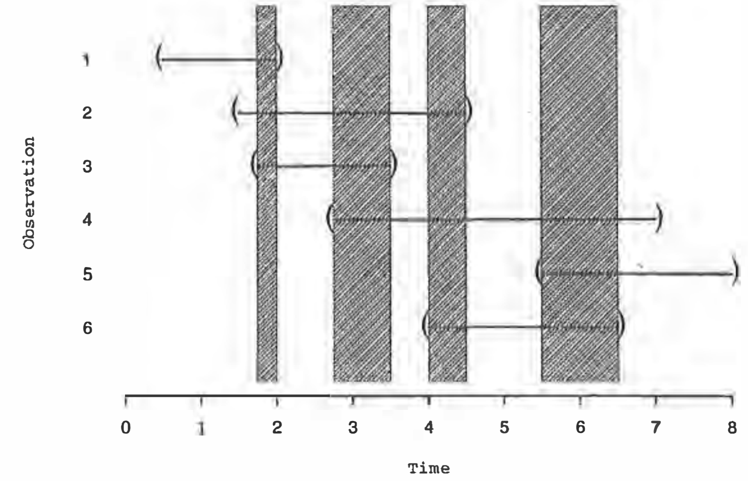


Figure 1.4: *Real representation of interval censored data and box representation of* $H_{\mathcal{M}}(\mathcal{M})$.

Referring back to the two competing fundamental assumptions of nonparametric analysis for censored data mentioned on page 4, it is clear that the likelihood function does not involve the linear extensions compatible with the data, depends on the cardinality and order of the maximal antichains, and is determined wholly by the partial order relationships of the interval order. In that sense, nonparametric maximum likelihood and self-consistent estimates, as we shall see in Chapter 4, participate of the pure incomparability assumption.

We can now describe special cases of interval censored data in terms of the maximal antichain structure of their underlying interval order. All statements below are made with probability one; we also assume that no element is universal in our description of the double duals.

Exact data. The interval order underlying exact data is linearly ordered under continuous F and weakly ordered if F has atoms. Every maximal antichain is essential and contains exactly one simplicial element under the continuous model, possibly more otherwise. The Petrie matrix is the $n \times n$ identity matrix under the continuous model. Under an F with atoms, the Petrie matrix is constrained to have a single 1 in every column. The characteristic matrix of such data is non-zero everywhere on the diagonal and only on the diagonal, i.e. $\chi_{ij} > 0$ if and only if $i = j, i, j = 1, \ldots, m$. In terms of double duals, $M_i^{**} = \{M_i\}$ for $i = 1, \ldots, m$.

Right-censored data. Every maximal antichain in an interval order representing right-censored data is essential, as maximal antichains are mapped onto the (small) intervals of the exact data points, except possibly for M_m, which may be unbounded above if the last observation is censored. This $\underline{i}$ corresponds to the first maximal antichain of the right-censored element with the longest dual. The characteristic matrix of such data is non-zero everywhere on the diagonal and can have non-zero values in the last column only on and after some row $\underline{i} > 1$, i.e. $\chi_{ii} > 0$ for all $i = 1, \dots, m$, and $\chi_{ij} > 0$ only if $j = m$ and $i \geq \underline{i}$. In terms of double duals, $M_i^{**} = \{M_i\}$ for $1 \leq i < \underline{i}$, and $M_i^{**} = \left\{M_{\underline{i}}, \dots, M_m\right\}$ for $\underline{i} \leq i \leq m$.

Current status data. Only M_1 and M_m are essential. For all $i = 2, \dots, m-1$, $\cup_{x \in M_i} x_i^* = \mathcal{M}$. Moreover, $\chi_{ii} > 0$ if and only if $i = 1$ or $i = m$, and $\chi_{ij} > 0$ only if either $i = 1$ or $j = m$. In terms of double duals, $M_1^{**} = \{M_1, \dots, M_{m-1}\}$, $M_m^{**} = \{M_2, \dots, M_m\}$, and $M_i^{**} = \mathcal{M}$ for $i = 2, \dots, m$.

Doubly censored data. Doubly censored data has a more complex structure than the previous examples. As happens with exact and right-censored data, the exact observations induce essential maximal antichains; however, some non-essential maximal antichains may be created by the overlap of left- and right-censored data, similarly to current status data. It is a simple matter to show that, as for current status data, if $M \in \mathcal{M}$ is not essential, then $\cup_{x \in M} x^* = \mathcal{M}$; however some essential maximal antichains may lie in $\{M_2, \dots, M_{m-1}\}$. In doubly censored data, $\chi_{ij} > 0$ only if either $i = j$, $i = 1$ or $j = m$. The double dual structure can be generalized from right-censored data. There exist $\underline{i}$ and $\bar{i}$ with $1 < \underline{i} \leq \bar{i} < m$ such that $M_i^{**} = \{M_1, \dots, M_{\bar{i}}\}$ for $1 \leq i < \underline{i}$, $M_i^{**} = \left\{M_{\underline{i}}, \dots, M_m\right\}$ for $\bar{i} < i \leq m$, and $M_i^{**} = \mathcal{M}$ for $\underline{i} \leq i \leq \bar{i}$.

There seems to have been a trend in recent years to treat the above cases as being distinct from interval censored data. In the case of exact and right-censored data, the existence of a closed form for the NPMLE (see § 3.2) and the general tractability of the problem probably warrant the distinction. In general, however, the above special cases as well as the general case of interval censored data benefit greatly from the simplification afforded by the order theoretic approach. In particular, this approach shows that the ordering structure of some data sets can be that of one of the above special cases without the data nominally adhering to traditional real-line-based definitions. Example 1.4, for instance, is structurally a case of current status data since the only essential maximal antichains are M_1 and M_m, and since every dual extends to either M_1 or M_m. The techniques described in Chapter 4, § 3.2 can therefore be applied to nonparametric likelihood estimation of the CDF for these data in spite of the fact that they are not nominally current status data.

Chapter 2

Minimal covers of maximal antichains for interval orders

Having hinted at the relevance of an order theoretic representation for interval censored data, we now turn away from statistics for the space of this chapter and, largely, of the next, and describe in some detail the structure of minimal covers of maximal antichains for interval orders (Definition 1.9). Aside from being an interesting extension of the classical clique covering problem, minimal covers are a crucial cornerstone of weak order partitioning of interval orders. This generic method for pseudo-random generation of linear rank statistics on the set of linear extensions of an interval order is elaborated in Chapter 3.

In this chapter, we characterize minimal covers and present two main algorithms to enumerate these sets. The problem of covering the vertices of incomparability graphs, including interval graphs, by maximal cliques has been well investigated by such as Fulkerson & Gross (1965), Bertossi & Bonucelli (1987), Rhee & Liang (1996) and Felsner, Müller & Wernisch (1997). Because the maximal antichains of $(X, \prec)$ are also the maximal cliques of $(X, \sim)$ (Chapter 1, § 1.3), the minimal cover problem translates to an extension of the minimum clique cover problem for interval orders (see Golumbic, 1980, Chapter 4): that of finding sets of maximal cliques no subset of which forms a cover for an interval graph. To our knowledge, this question has not been considered before. For our purposes, discussions of minimal covers are essential, as they underlie the methods of Chapter 3. The special structure and pliability of interval orders have led us to adopt the point of view of the precedence relation $(X, \prec)$ dual to the interval graph $(X, \sim)$, which leads us to consider maximal antichains instead of maximal cliques.

We consider minimal covers from two points of view. The first, discussed in § 1, is a backtracking algorithm which constructs all minimal covers of an interval order. We show that this algorithm is a generalization of a simpler one which generates one maximum chain along with one minimum cover from an interval order. The second perspective is that of a characterization of the complements of minimal covers, termed *maximal removable sets.* We provide properties and algorithmic details concerning maximal removable sets. These, while of perhaps less immediate applicability than the generating algorithm of § 1, are valuable for the insight they provide on the structure of minimal covers and ultimately supply us with an upper bound on the maximum number of minimal covers achievable with a given number of maximal antichains. In § 3, we present the derivation of this upper bound, as well as some simulation results and a short discussion of some research avenues concerning minimal covers of interval orders.

1 Finding minimal covers

In this section, we first present a simple algorithm in Construction 2.2 which enables us to find a single minimum cover. This procedure is similar to that found in Gavril (1972); its expression is couched in the language of comparability rather than incomparability, and serves to introduce its generalization to Construction 2.3. These latter algorithms produce all minimal covers of an interval order.

The following result, found in Felsner (1992), Chapter 1, and closely related to Dilworth's Decomposition Theorem (Dilworth, 1950), concerns the cardinality of minimum covers, also called the *clique cover number* (Golumbic, 1980, Chapter 1).

Theorem 2.1 *Let $(X, \prec)$ be a partially ordered set. The height (length of the longest chain) of $(X, \prec)$ equals the minimum number of antichains required to cover the elements of X.*

Since the minimum number of antichains can be no smaller than the minimum number of maximal antichains which cover a poset, the theorem tells us that a minimum cover has cardinality equal to the length of the longest chain in the poset.

This result can be shown constructively for interval orders, with the added bonus that we produce a minimum cover and a maximum chain in the process. In essence, the following Construction and

Theorem state that there exists a minimum cover consisting of a set of maximal antichains, each of which uniquely contains a certain element. The set of these elements forms a maximum chain in the interval order.

Construction 2.2 *Let* $\{M_1, \ldots, M_m\}$ *be the linearly ordered set of maximal antichains of interval order* $\underline{X} = (X, \prec)$. *Let* y_1 *be a simplicial element of* X *belonging to the first maximal antichain* M_1, *and let* $\mu_1 = M_1$. *Form the sets* $Y_1' = \{y_1\}$ *and* $\mathcal{W}_1' = \{\mu_1\}$. *Then form the sets* Y_i' *and* $\mathcal{W}_i'$ *for* $i > 1$ *as follows:*

While i *is such that* $\{y; \mu_{i-1} < y^*\} \neq \emptyset$, *let* $Y_i' = Y_{i-1}' \bigcup \{y_i\}$, *where* y_i *is any selection of the set*

$$\left\{ y; y \text{ minimizes } \max_{\mu_{i-1} < w^*} w^* \right\},$$

and let $\mathcal{W}_i' = \mathcal{W}_{i-1}' \bigcup \{\mu_i\}$, *where*

$$\mu_i = \max y_i^*.$$

Theorem 2.2 *Under Construction 2.2, there exists* I *such that* $\{y; \mu_I < y^*\} = \emptyset$ *and* $\mathcal{W}' = \mathcal{W}_I'$ *is a minimum cover for* X.

Proof. The proofs of coverage and minimality for $\mathcal{W}'$ are specializations of the proofs of Claims 1 and 2 in Theorem 2.3, possible since Construction 2.2 is a special case of Construction 2.3. Proof that $\mathcal{W}'$ is of minimal cardinality is simple viewed from the context of Theorem 2.3. See also Gavril (1972). □

The initial inclusion of $\{M_1\}$ in the sequence of $\mathcal{W}_i'$'s in Construction 2.2 is legitimate and necessary since M_1 is always essential, and therefore required in any cover. We note without proof that $Y' = Y_I$ is a maximum chain in $\underline{X}$.

Example 2.1 *(In the following example as in subsequent ones, elements of* X *[columns] are identified by regular subscripts and maximal antichains of* $\underline{X}$ *[rows] by bolded subscripts.) Consider the*

Petrie matrix

$$\mathbf{A} = \begin{array}{c|cccccccccccccccccccccccc|}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \\
\mathbf{1} & \underline{\mathbf{1}} & \underline{1} & \underline{1} & \underline{1} & \underline{1} & \underline{1} & \underline{1} & \underline{1} & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{2} & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{3} & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{4} & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{5} & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{6} & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & \underline{1} & \underline{\mathbf{1}} & \underline{\mathbf{1}} & \underline{1} & \underline{1} & \underline{1} & \underline{1} & \underline{1} & \underline{1} & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{7} & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
\mathbf{8} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
\mathbf{9} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & \underline{1} & \underline{\mathbf{1}} & \underline{1} & \underline{1} & 0 & 0 \\
\mathbf{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & \underline{\mathbf{1}} & \underline{\mathbf{1}}
\end{array}$$

Construction 2.2 yields $\mathcal{W}' = \{\mu_1, \mu_6, \mu_9, \mu_{10}\}$, with $I = 4$. Members of these maximal antichains which do not *belong to maximal antichains previously chosen in the construction have their Petrie matrix entries underlined above. Y' could be any one of the sets $\{x_1, x_{11}, x_{20}, x_{23}\}$, $\{x_1, x_{11}, x_{20}, x_{24}\}$, $\{x_1, x_{12}, x_{20}, x_{23}\}$ or $\{x_1, x_{12}, x_{20}, x_{24}\}$. The occurrences of the y's in the maximal antichains for which they cause inclusion in the minimum cover are bolded in the matrix.*

Construction 2.2 will yield a single minimum cover. The natural generalization from this setting is to broaden the choice for μ_i without changing the property that no two elements of Y are covered by the same maximal antichain. This generalization leads to Construction 2.3. The $<$ precedence relation between maximal antichains is explained on page 16.

Construction 2.3 *Let y_1 be a simplicial element of X belonging to the first maximal antichain M_1, and let $\mu_1 = M_1$. Form the sets $Y_1 = \{y_1\}$ and $\mathcal{W}_1 = \{\mu_1\}$. Then form the sets Y_i and $\mathcal{W}_i$ for $i > 1$ as follows:*

While i is such that $\{y; \mu_{i-1} < y^*\} \neq \emptyset$, let $Y_i = Y_{i-1} \bigcup \{y_i\}$, where y_i is any selection of

$$\left\{y; y \textit{ minimizes } \max_{\mu_{i-1} < y^*} y^*\right\},$$

and let $\mathcal{W}_i = \mathcal{W}_{i-1} \bigcup \{\mu_i\}$, where

$$\mu_i \in y_i^* \setminus y_{i-1}^*$$

is a particular selection of μ_i.

Theorem 2.3 *For every sequence of pairs $(\mathcal{W}_i, Y_i)$, $i = 1, \ldots$ formed in Construction 2.3, there exists $I \leq m$ such that $\{y; \mu_I < y^*\} = \emptyset$. Defining $(\mathcal{W}, Y) = (\mathcal{W}_I, Y_I)$, the class of minimal covers of X is exactly the class of sets $\mathcal{W}$ which can be produced by Construction 2.3.*

Proof of Theorem 2.3. First we must show that I is well-defined for any selection of $y_i, \mu_i, i = 1, \ldots, I$ compatible with Construction 2.3. Since $\mu_{i-1} < y_i^*$ and since $\mu_i \in y_i^*$, it must be that $\mu_{i-1} < \mu_i$. Thus in at most m steps we will reach $\mu_I = M_m$. Since $M_m = \max \mathcal{M}$ we obtain $\{y; \mu_I < y^*\} = \emptyset$, and the construction terminates.

We now show the theorem's statement in three steps:

1. Every $\mathcal{W}$ produced by Construction 2.3 is a cover for X.
2. Every cover $\mathcal{W}$ produced by Construction 2.3 is minimal for X.
3. Every minimal cover for X is a set $\mathcal{W}$ compatible with the selection method of Construction 2.3.

- **Claim 1**: *Every $\mathcal{W}$ is a cover for X.*

Let $Y = Y_I$ and $\mathcal{W} = \mathcal{W}_I$ be particular realizations of Construction 2.3, and let $x \in X$ be arbitrary. We must show that $\mu \in x^*$ for some $\mu \in \mathcal{W}$. If $x \in Y$, x is covered by construction, so assume $x \in X \setminus Y$. Since $\mu_1 = M_1 < x^* < M_m = \mu_I$, there must be some i such that $\mu_i < x^*$ but $\mu_{i+1} \not< x^*$. Also, since we know y_{i+1} minimizes $\max\limits_{\mu_i < y^*} y^*$ and since $\mu_i < x^*$, we deduce that $\max y_{i+1}^* \leq \max x^*$.

We can now show that $\mu_{i+1} \in x^*$. Assume not; then we need $\max x^* < \mu_{i+1}$. But then $\max x^* < \mu_{i+1} \leq \max y_{i+1}^* \leq \max x^*$, a contradiction. Therefore x is contained in μ_{i+1}, so $\mathcal{W}$ is a cover.

- **Claim 2**: *Every cover $\mathcal{W}$ is minimal for X.*

Let Y and $\mathcal{W}$ be particular realizations of Construction 2.3 as before, and let $\mathcal{W}^{(i)} = \mathcal{W} \setminus \{\mu_i\}$ for some $i \in \{2, \ldots, I-1\}$, since neither $\mu_1 = M_1$ nor $\mu_I = M_m$ can be removed. We show that $\mathcal{W}^{(i)}$ is not a cover.

Since $\mu_{i-1} < y_i^*$, we know that $\mu_{i-1} \notin y_i^*$, which implies that $\{\mu_1, \ldots, \mu_{i-1}\} \bigcap y_i^* = \emptyset$, by the linear ordering of the maximal antichains. By construction, $\mu_{i+1} \notin y_i^*$, which again implies that $\{\mu_{i+1}, \ldots, \mu_t\} \bigcap y_i^* = \emptyset$. Hence $\mathcal{W}^{(i)}$ does not cover y_i, and thus $\mathcal{W}$ is a minimal cover.

- **Claim 3**: *Every minimal cover for X is produced by the algorithm of the theorem.*

Let $V = \{\nu_1 = M_1, \nu_2, \dots, \nu_{t-1}, \nu_t = M_m\}$ be a minimal cover. We need to supply a set $Y = \{y_1, \dots, y_t\}$ such that V and Y satisfy Construction 2.3. For this it is enough to show that

$$\text{if } y_i \in \left\{ y ; \underset{y}{\operatorname{argmin}} \max_{\nu_{i-1} < y^*} y^* \right\} \text{ for } i = 1, \dots, I, \text{ then } \nu_i \in y_i^* \setminus y_{i-1}^*. \tag{2.1}$$

Apply Construction 2.2 to V, which by Lemma 1.6 is the maximal antichain set of some interval order. Since V is minimal, it is its own minimum cover. The set Y' produced in the construction satisfies requirement 2.1.

□

Though Construction 2.3 involves an arbitrary selection from a set, the available set from which to select μ_i does not depend on the particular choices of y_{i-1} and y_i, since only $\max y_{i-1}^*$ and $\max y_i^*$ play a role in the determination of that set. Algorithm 2.4 **ListMinCovers** below explicitly recognizes this fact by retaining only this information from the elements of Y. On the other hand, the choice of μ_{i-1} does affect the set of available y_i's at every step.

The elements of the sets Y produced by Construction 2.3 can be circumscribed.

Definition 2.1 *An element of X is called a minimal element if its dual properly contains the dual of no other element of X.*

All sets Y compatible with Construction 2.3 will be subsets of the set of all minimal elements of X. This characterization of the elements of all sets Y from Theorem 2.3 means that the original set X may be reduced to just its minimal elements to solve the minimal cover problem. In view of this observation, we can translate Theorem 2.3 into the following algorithm, which returns the set of all minimal covers of X when called with arguments $(X_{\mathrm{min}}, \emptyset)$, with X_{min} the set of minimal elements of X. The second argument is $\emptyset$ only on the first call, with the convention that $\emptyset < M$ for all $M \in \mathcal{M}$; otherwise it is a maximal antichain, $\mu_\rfloor$, which corresponds to $\max y_{i-1}^*$ and contributes to defining the set $y_i^* \setminus y_{i-1}^*$ of Construction 2.3.

Algorithm 2.4

```
**ListMinCovers**$(Y, \mu_\rfloor)$

Arguments: $Y$, a set of minimal elements of an interval order;
           $\mu_\rfloor$, equal to $\emptyset$ on the first call, or to $\mu_\rfloor$, a maximal antichain, in any subsequent call.

If $Y = \emptyset$
    Return $\{\emptyset\}$
else
    $\mathcal{M} := \emptyset$
    $y_0 := \underset{y}{\operatorname{argmin}} \max_{y \in Y} y^*$
    For each $\mu_0 \in \{M; M \in y_0^*, \mu_\rfloor < M\}$
        $\mathcal{M}_0 :=$ ListMinCovers$(\{y; \mu_0 < y^*\}, \max y_0^*)$
    For each $\mathcal{W}_0 \in \mathcal{M}_0$
        $\mathcal{M} := \mathcal{M} \bigcup \{\{\mu_0\} \bigcup \mathcal{W}_0\}$
Return $\mathcal{M}$
```

The following example is accompanied by a sample recursion tree for the **ListMinCovers** algorithm at all stages of the recursion from each branch after the leaves are reached:

Example 2.2 *Consider the Petrie matrix*

$$A = \begin{array}{c} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \\ \mathbf{5} \\ \mathbf{6} \\ \mathbf{7} \\ \mathbf{8} \end{array} \begin{array}{c} \begin{array}{cccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

If we identify once again elements and maximal antichains by their subscripts, we can form the set of minimal X-elements $Y_{min} = \{1, 7, 10, 13, 16\}$, corresponding to Petrie pairs $\{(\mathbf{1}, \mathbf{1}), (\mathbf{2}, \mathbf{4}), (\mathbf{3}, \mathbf{5}), (\mathbf{6}, \mathbf{7}), (\mathbf{8}, \mathbf{8})\}$. The first tree (Figure 2.1) shows the arguments to each of the calls to **ListMinCovers,** *unfurled to the first returning call for each branch. The edges are labelled with the values of μ_0 and y_0 in the calling procedure.*

Figures 2.2 to 2.6 show how covers are built up upon returning from recursion.

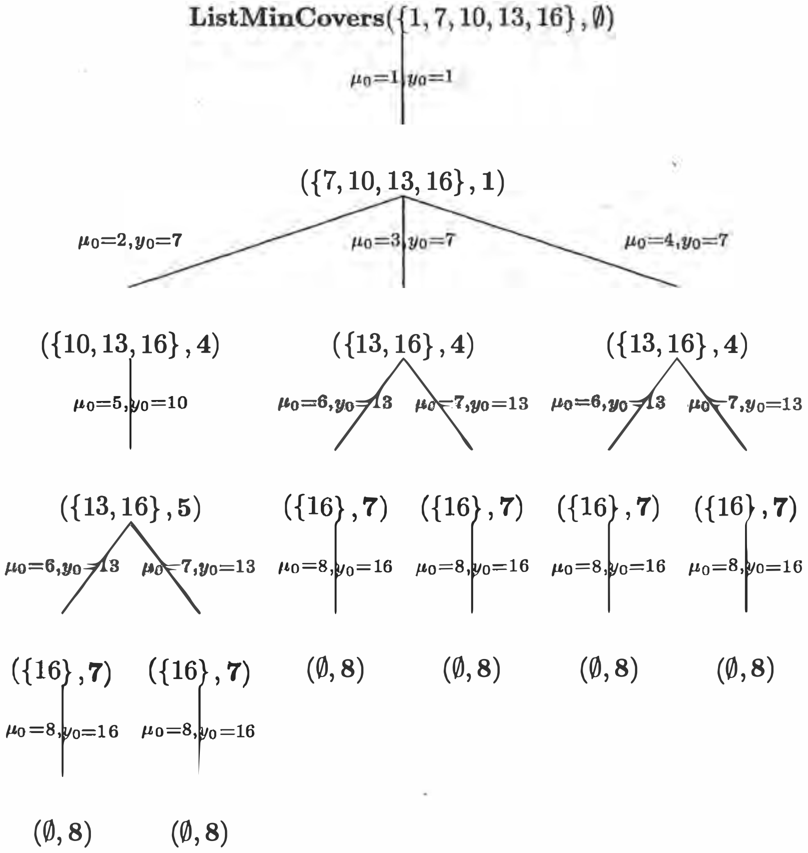


Figure 2.1: Argument tree for the recursive calls of **ListMinCovers**

Figure 2.2: Returned values for **ListMinCovers**, 5th level calls

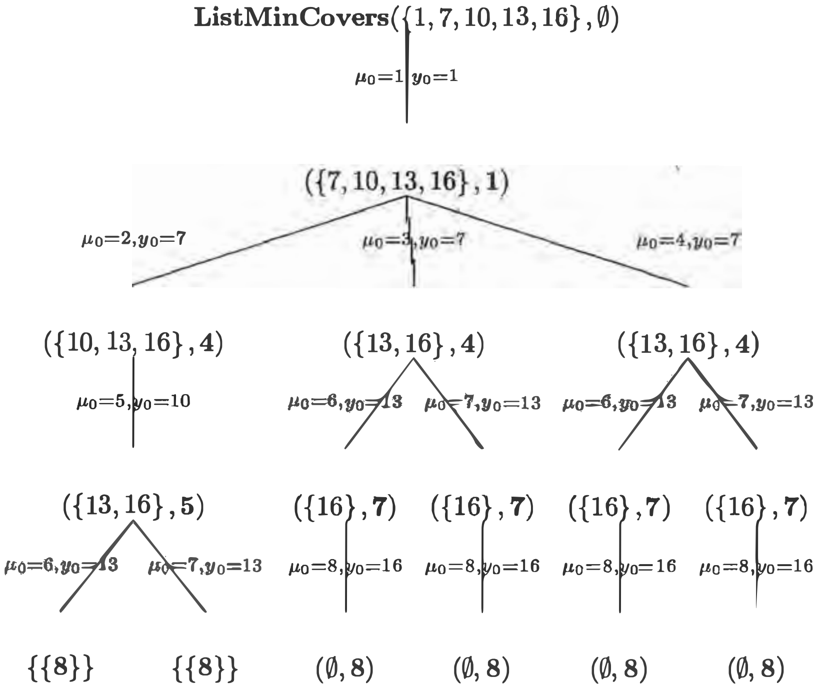


Figure 2.3: Returned values for **ListMinCovers**, 4th level calls

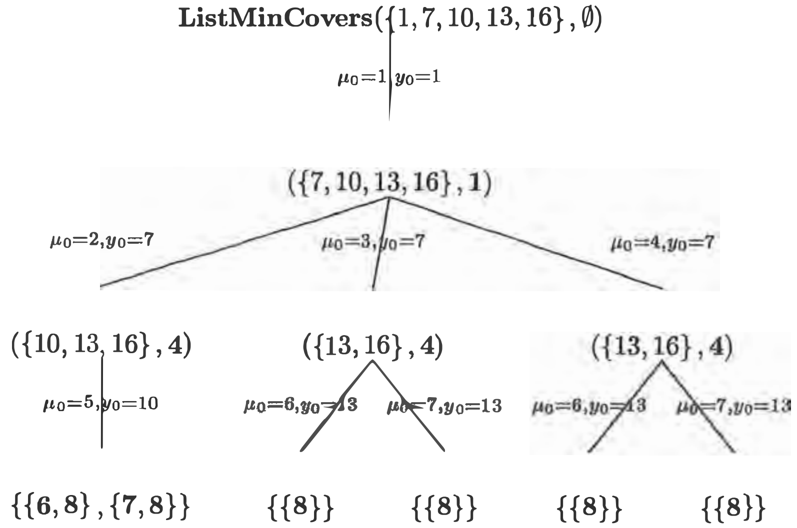


Figure 2.4: Returned values for **ListMinCovers**, 3rd level calls

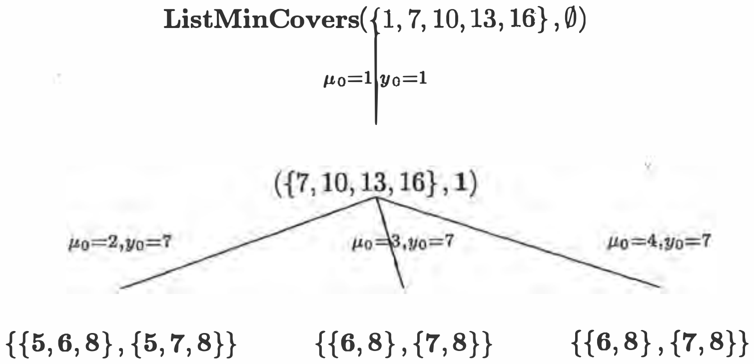


Figure 2.5: Returned values for **ListMinCovers**, 2nd level calls

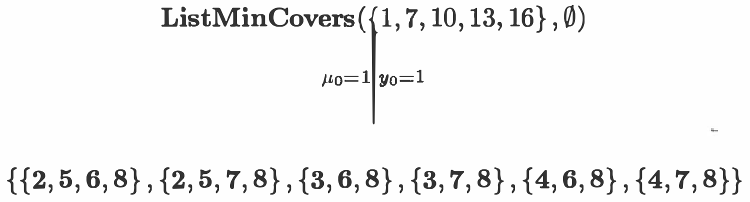


Figure 2.6: Returned values for **ListMinCovers**, initial call

$$\{\{\mathbf{1,2,5,6,8}\},\{\mathbf{1,2,5,7,8}\},\{\mathbf{1,3,6,8}\},\{\mathbf{1,3,7,8}\},\{\mathbf{1,4,6,8}\},\{\mathbf{1,4,7,8}\}\}$$

Apart from the reduction of the original X to its minimal elements, several other simplifications can be applied to the minimal cover enumeration procedure of Construction 2.3. These simplifications may provide substantial gain in practice, in particular if we modify Algorithm 2.5 to count, as opposed to list, minimal covers.

- Further reduction is accomplished by the prior inclusion of every essential maximal antichain in the minimal cover $\mathcal{W}$ and by correspondingly restricting the minimal element set X_{min} to non-simplicial elements, as these elements can only generate their own essential maximal antichain in Construction 2.3.
- If we order X_{min} according to the order of the initial maximal antichain of their dual, we need

only find minimal covers for each series of consecutively overlapping minimal elements, and combine them at the end of the procedure.

- Finally, maximal antichains which overlap exactly the same minimal elements will be mutually exclusive and interchangeable in any minimal cover produced by Construction 2.3. The classes of such elements can be kept track of and all but one element from each class deleted from the problem, to be reinstated after the covers are listed.

Example 2.3 *Consider Figure 2.1; the above simplifications and reductions take on the following forms:*

- *The first and last maximal antichains can be removed from the problem, and simply added to the sets produced at the end of the procedure.*
- *Once the leaves corresponding to the last essential maximal antichains are removed, we notice that a subtree of height 2 is repeated at every termination of the tree. This occurs because element 13 and the closest element above, 10, share no maximal antichain in common.*
- *Lastly, we see that the two right subtrees at the second level are identical, reflecting the fact that maximal antichains 2 and 3 overlap the same elements in the minimal cover.*

The above considerations yield the following algorithm from Construction 2.3:

Algorithm 2.5

```
SimplifiedListMinCovers($\underline{X}$)
Arguments: $\underline{X} = (X, \prec)$, an interval order.

1. Determine the set $Y'$ of all non-simplicial minimal elements of $X$.
2. Determine sets $K_1, \ldots, K_r$ of maximal antichains which overlap exactly the same elements in $Y'$.
3. For each set $K_j = \{\kappa_{j,1}, \ldots, \kappa_{j,r_j}\}, j = 1, \ldots, r$
        For each $y \in Y'$
            Let $y^* := y^* \setminus \{\kappa_{j,2}, \ldots, \kappa_{j,r_j}\}$
4. Form the disjoint subsets $Y'_1, \ldots, Y'_p$ of pairwise overlapping elements of $Y'$.
5. For each set $Y_k, k = 1, \ldots, p$, label its elements $Y_k = \{y_{k,1}, \ldots, y_{k,t_k}\}$
```

```
        so that $j_1 < j_2 \Leftrightarrow \min y^*_{k,j_1} < \min y_{k,j_2}$, $j_1, j_2 = 1, \ldots, t_k$.
6. For each set $Y_k$
        Let $\mathcal{M}_k :=$ ListMinCovers$(Y_k, \emptyset)$
7. Let $\mathcal{M} = \{\mathcal{W}; \mathcal{W} = \bigcup_{k=1}^{p} \mathcal{W}_k, \mathcal{W}_k \in \mathcal{M}_k, k = 1, \ldots, p\}$
8. Augment $\mathcal{M}$ by all combinations of its covers
        replacing element $\kappa_{j,1}$ by an element of $K_j \setminus \{\kappa_{j,1}\}$, $j = 1, \ldots, r$.
9. Add all essential maximal antichains to every cover in $\mathcal{M}$.
   Return $\mathcal{M}$
```

Example 2.4 *Using Algorithm 2.5 applied to the interval order of Example 2.2, we would get the following sequence of steps:*

1. $Y' := \{7, 10, 13\}$
2. $K_1 := \{2\}, K_2 := \{3, 4\}, K_3 := \{5\}, K_4 := \{6, 7\}$
3. $y^*_7 := \{2, 3\}, y^*_{10} := \{3, 5\}, y^*_{13} := \{6\}$ *(notice that this instruction reduces the number of steps which need to be performed by* **ListMinCovers***)*
4. $Y'_1 = \{7, 10\}, Y'_2 = \{13\}$
5. *As in preceding step.*
6. $\mathcal{M}_1 :=$ **ListMinCovers**$(\{7, 10\}, \emptyset) = \{\{2, 5\}, \{3\}\}$, $\mathcal{M}_2 =$ **ListMinCovers**$(\{13\}, \emptyset) = \{\{6\}\}$
7. $\mathcal{M} := \{\{2, 5, 6\}, \{3, 6\}\}$
8. $\mathcal{M} := \{\{2, 5, 6\}, \{2, 5, 7\}, \{3, 6\}, \{3, 7\}, \{4, 6\}, \{4, 7\}\}$
9. $\mathcal{M} := \{\{1, 2, 5, 6, 8\}, \{1, 2, 5, 7, 8\}, \{1, 3, 6, 8\}, \{1, 3, 7, 8\}, \{1, 4, 6, 8\}, \{1, 4, 7, 8\}\}$

2 Maximal removable sets

We now turn our attention to the generation of maximal removable sets. A set of maximal antichains is removable if no sequence of maximal antichains in the set is equal to x^* for any $x \in X$, i.e. if its complement is a cover. A removable set is maximal if adding any maximal antichain creates such a sequence. An alternative characterization of a maximal removable set is that it is a set of maximal antichains such that its complement with respect to $\mathcal{M}$ is a minimal cover. Hence, a maximal removable set is a removable set that is not a proper subset of any other removable set.

The algorithm discussed in this section is solely for exposition. It is much more efficient to use Algorithm 2.5 and in fact attempts to improve or analyze Algorithm 2.7 lead to consideration of minimal elements and a variant of Algorithm 2.5.

2.1 Simplifying Assumptions

For any minimal cover or maximal removable set problem, we can assume that the only essential maximal antichains are M_1 and M_m. If any other maximal antichain, say M_r, is essential, the problem can be split into two subproblems: one dealing with M_1 through M_r and the other with maximal antichains M_r through M_m. The solution to the large problem is simply the union of the solutions to the two subproblems.

For a minimal cover this statement would be negated if there existed a maximal antichain which belonged to the minimal cover of one subproblem but which could be removed once the union were formed. This would imply that the element covered by this maximal antichain had now been covered by a maximal antichain from the other subproblem. However, by the consecutive-1's property of the dual any such maximal antichain would be covered by M_r, and hence a contradiction would arise. Since every MRS is the complement of a minimal cover, the problems are equivalent and the same simplification obtains.

We also assume that any x such that x^* contains an essential maximal antichain has been removed from the problem, since it is covered by the essential maximal antichain. In other words, since the essential maximal antichain must be retained the element x imposes no other restrictions on the MRS.

2.2 MRS generation algorithm

Once the simplifying assumptions of § 2.1 have been applied we may assume that we are dealing with maximal antichains M_2 through M_{m-1} and that none of these maximal antichains are essential.

The following observation can be proven under these assumptions.

Observation 2.6 *For any i such that $2 \leq i \leq m-3$ both M_i and M_{i+2} can be removed together.*

Proof. Suppose not. Then there must exist an x such that x^* includes M_i and M_{i+2} but not M_{i+1}. For all interval orders x^* must have the consecutive-1's property; hence no such x can exist. □

The basis of our MRS generating algorithm is the two following properties.

Property I *If M_i is the largest maximal antichain in a removable set, then M_{i+2} can be removed for $i < m - 2$.*

Property II. *There can exist MRS 's containing both M_i and M_{i+3} and neither M_{i+1} nor M_{i+2} for $i < m - 2$.*

The proof of these properties follows. Property I is simply a special case of Observation 2.6 and hence holds. Property II can hold when there are $x, y \in X$ such that $x^* = \{M_a, \ldots, M_i, M_{i+1}\}$, $a \leq i$, and $y^* = \{M_{i+2}, M_{i+3}, \ldots, M_b\}$, $i + 3 \leq b$. In this case, if $\{M_a, \ldots, M_i\} \bigcup \{M_{i+3}, \ldots, M_b\}$ is removable and removed then neither M_{i+1} nor M_{i+2} can be removed.

In principle, our algorithm consists of enumerating all removable sets and then deleting from the enumeration those that are proper subsets of any other removable set. The efficiency of the algorithm can be maintained by performing the deletions while enumeration is taking place.

We first note the form of removable sets. We maintain the subscript ordering described in § 1.3. Consider the subscripts of any two adjacent elements of an MRS. They must follow one of three patterns: (M_j, M_{j+1}), (M_j, M_{j+2}), or (M_j, M_{j+3}). It is not possible for (M_j, M_{j+4}) to occur in sequence in an MRS because, by a variant of Property I, it is possible to remove M_{j+2} for any removable set which includes (M_j, M_{j+4}), and hence the removable set is not maximal.

This points to a further important property of MRS 's. Consider constructing an MRS starting with either M_2 or M_3 and subsequently adding maximal antichains with larger indices. The partially constructed list will be termed a candidate set. Define the *tail* of a candidate set as the set of maximal antichains with subscripts contiguous with the largest maximal antichain in that candidate set. Any two candidate sets with the same tail will evolve in exactly the same manner. A gap in the sequence of maximal antichains making up a removable set provides a kind of independence: the future evolution of a candidate set does not depend on those maximal antichains preceding the gap.

For example, suppose two candidate sets are $\{M_2, M_3, M_5\}$ and $\{M_3, M_5\}$, both with tail $\{M_5\}$. There is no point in pursuing $\{M_3, M_5\}$ since its evolution can be no different from that of $\{M_2, M_3, M_5\}$ and therefore it cannot be maximal. On the other hand if one candidate is $\{M_2, M_3, M_4\}$ and a second is $\{M_3, M_4\}$ then such elimination is not possible. While the smaller is a proper subset of the larger, they do not share the same tail.

From the preceding argument we can deduce that there are at most 3^{m-2} sequences to be examined. However, the number is actually much smaller. One needs to consider (M_j, M_{j+3}) only if (M_j, M_{j+1}) is not possible. Hence the true upper bound is 2^{m-2}. A further substantial saving comes from identifying and deleting candidates that will become proper subsets of other candidates.

Algorithm 2.7

```
ListMaxRemovableSets($S$)
Arguments: $S$ a list of candidates.

Begin
    If $S = \emptyset$ then
        Return ListMaxRemovableSets($\{\{M_2, \mathcal{M}\backslash M_2\}\}, \{M_3, \mathcal{M} \setminus \{M_2, M_3\}\}$)
    Else
        Begin
        $S^* = \emptyset$
        For each $w = \{s, \mathcal{M}\}$ in $S$
            Begin
                If isCovered ($s$, $S^*$, $S$) Break
                $S^*$ = append ( $S^*$, $\{(s \cup M_2), \mathcal{M}\backslash \{M_1, M_2\}\}$)
                If isRemovable ($s \cup M_1$)
                    $S^*$ = append ( $S^*$, $\{(s, M_1), \mathcal{M}\backslash M_1\}$)
                Else If isRemovable ($s \cup M_3$)
                    $S^*$ = append ($S^*$, $\{(s, M_3), \mathcal{M}\backslash \{M_1, M_2, M_3\}\}$)
            End
        If Finished ($S^*$)
            Return ($S^*$)
        Else
```

```
            Return ListMaxRemovableSets($S^*$)
        End
End
```

Our algorithm *grows* candidate lists starting at the left end. By Property I either M_2 or M_3 must be the first element of the MRS. From this point elements are added sequentially to the candidate lists according to Properties I and II and the observations made above.

Any candidate which is a proper subset of another candidate can be deleted if both have the same tail ; this is detected by the function **isCovered**. The function **isRemovable** determines whether a given candidate can be removed. Finally the function **Finished** determines whether all candidates have been pursued to M_m.

Consider the output of Algorithm 4.2 applied to the data of Example 2.2. In that example only M_1 and M_8 are essential.

Step 1 $S_0 = \{M_1, M_2\}$

$T_1 = \{M_2, M_3\}$ removable

$T_2 = \{M_2, M_4\}$ removable

$T_3 = \{M_3, M_4\}$ removable

$T_4 = \{M_3, M_5\}$ removable

Step 2 $S_1 = \{\{M_2, M_3\}, \{M_2, M_4\}, \{M_3, M_4\}, \{M_3, M_5\}\}$

$T_1 = \{M_2, M_3, M_4\}$ **not removable** $x_7^* \subset T_1$

$T_1' = \{M_2, M_3, M_6\}$ removable

$T_2 = \{M_2, M_3, M_5\}$ removable

$T_3 = \{M_2, M_4, M_5\}$ removable

$T_4 = \{M_2, M_4, M_6\}$ removable

$T_5 = \{M_3, M_4, M_5\}$ **not removable** $x_{10}* = T_5$

$T_5' = \{M_3, M_4, M_7\}$ removable

$T_6 = \{M_3, M_4, M_6\}$ removable

$\{M_3, M_5\}$ is not considered because it has the same tail as T_2.

Step 3 $S_2 = \{\{M_2, M_3, M_6\}, \{M_2,, M_3, M_5\}, \{M_2, M_4, M_5\},$

$\{M_2, M_4, M_6\}, \{M_3, M_4, M_7\}, \{M_3, M_4, M_6\}\}$

$T_1 = \{M_2, M_3, M_6, M_7\}$ **not removable** $x^*_{13} \subset T_1$

$T'_1 = \{M_2, M_3, M_5, M_6\}$ removable

$T_2 = \{M_2, M_3, M_5, M_7\}$ removable

$T_3 = \{M_2, M_4, M_5, M_6\}$ removable

$T_4 = \{M_2, M_4, M_5, M_7\}$ removable

$T_5 = \{M_2, M_4, M_6, M_7\}$ **not removable** $x^*_{13} \subset T_6$

$T'_5 = \{M_3, M_4, M_6\}$ removable

$T_6 = \{M_3, M_4, M_7\}$ removable

Done $S = \{\{M_2, M_3, M_5, M_6\}, \{M_2, M_3, M_5, M_7\}, \{M_2, M_4, M_5, M_6\},$

$\{M_2, M_4, M_5, M_7\}, \{M_3, M_4, M_6\}, \{M_3, M_4, M_7\}\}$

3 The number of minimal covers of an interval order

How many minimal covers does $\underline{X}$ have? The question is not only of theoretical interest. In applications, if this number grows too large with the size of $\underline{X}$, we might opt for random generation of minimal covers instead of enumeration.

The number of minimal covers will depend on four characteristics of the interval order:

- the number of maximal antichains;
- the number of minimal elements;
- the length of each minimal element;
- the degree of overlap between the minimal elements.

The problem of determining the number of minimal covers for a general interval order is thus fairly complex. The last item, degree of overlap, can be more formally described as the size of the intersection of the duals of minimal elements. Overlap between minimal elements can either increase or decrease the number of minimal covers for an interval order, as the next example shows.

Example 2.5 *In the following examples, partial Petrie matrices of elements which do not overlap with simplicial elements are displayed on the left of each table; minimal covers, one per column to the right, are denoted by dots (•) which indicate maximal antichain membership in the cover.*

The first pair of examples illustrates a case where overlap decreases the number of minimal covers:

$$\begin{bmatrix} \vdots & \vdots \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \end{bmatrix}$$

$\vdots$				
M_i	•	•		
M_{i+1}			•	•
M_{i+2}	•	•		
M_{i+3}		•		•
$\vdots$				

4 minimal covers

$$\begin{bmatrix} \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \end{bmatrix}$$

$\vdots$			
M_i	•	•	
M_{i+1}			•
M_{i+2}	•		
M_{i+3}		•	
$\vdots$			

3 minimal covers

The second pair of examples illustrates a situation where overlap increases the number of minimal covers:

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$\vdots$				
M_i	•	•		
M_{i+1}			•	•
M_{i+2}	•	•		
M_{i+3}	•		•	
M_{i+4}		•		•
M_{i+5}	•		•	
$\vdots$				

4 minimal covers

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$\vdots$					
M_i	•	•			
M_{i+1}			•	•	•
M_{i+2}	•	•	•		
M_{i+3}				•	
M_{i+4}	•				•
M_{i+5}		•	•	•	
$\vdots$					

5 minimal covers

We aim at bounding the maximal number of minimal covers $N_{\max}(m)$ for an interval order with m maximal antichains, and at establishing whether enumeration of the minimal covers is a useful approach for realistic interval data sets. With these views in mind, § 3.1 provides lower and upper bounds on $N_{\max}$, while § 3.2 following provides some simulation results designed to supply an empirical approximation for this number and to determine the applicability of Algorithm 2.5 in a realistic situation.

3.1 Bounds on $N_{\max}(m)$

Theorem 2.8 *For $m \geq 2$,*

$$\left\lfloor 3^{\frac{m-2}{3}} \right\rfloor \leq N_{\max}(m) \leq \left\lceil \alpha\rho^{m-1} + 0.5 \right\rceil , \tag{2.2}$$

where

$$\kappa_1 = \sqrt[3]{19 + 3\sqrt{33}},\ \kappa_2 = \sqrt[3]{19 - 3\sqrt{33}}$$

$$\alpha = \frac{1}{9\sqrt{33}}\left(\kappa_1^2 + \kappa_2^2\right) + \frac{5}{2}\left(\kappa_1 + \kappa_2\right) + \frac{1}{3} \approx 0.6184,\ \rho = \frac{1}{3}\left(\kappa_1 + \kappa_2 + 1\right) \approx 1.8393$$

Proof. An immediate lower bound on the maximum number of minimal covers $N_{\mathbf{max}}(m)$ for a given m can be determined if we assume no overlap between minimal elements. The number of minimal covers ${}^{\mathrm{no}}N_{\mathbf{max}}$ in the case of no overlap is then simply the product of the lengths of the minimal elements. Since M_1 and M_m are essential and both contain a minimal element of length 1, we need to partition $m-2$ elements into consecutive sequences such that the product of the lengths of these sequences is maximized. The problem can be reformulated as that of maximizing $\prod_{k=1}^{K} a_k$ subject to $\sum_{k=1}^{K} a_k = m-2$ for K and $a_1, \ldots, a_k$, for which a solution is exposed in Saaty (1970, Theorem 4-6). For all values of $m \geq 2$, ${}^{\mathrm{no}}N_{\mathbf{max}} \geq 3^{\frac{m-2}{3}}$.

For the upper bound, we bound the number of MRS 's compatible with m maximal antichains by using the properties of MRS 's described in § 2.2. Specifically, we determine the number B_m of ways in which a sequence of $m-1$ (ordered) items can be partitioned exactly in consecutive groups of 1, 2 or 3 items, such that a group of 3 items occurs neither at the beginning nor at the end of the sequence. If we take the items to be the linearly ordered maximal antichains of an interval order beginning at M_2, then a set formed by the final maximal antichain of each group, bar the last group, will form a set of maximal antichain of the requisite structure for an MRS. The number of such partitions thus forms an upper bound on the number of MRS 's achievable with m maximal antichains.

The reason for starting the sequence at M_2 and omitting to include the final maximal antichain from the last group is that M_1 and M_m are essential, and thus cannot be removed. The requirement that a group of 3 neither start nor end a sequence of groups formed in this manner ensures that at least one of maximal antichains M_2 and M_3, and at least one of M_{m-2} and M_{m-1}, will be included in the candidate set, the necessity of which is stated in § 2.2.

To determine the number of groupings of $m-1$ items in sequence which satisfy the above conditions, we start with the determination of the number G_m of general groupings in 1, 2 or 3 of $m-1$ items which do not necessarily satisfy the condition on the first and last group. Since each such sequence must start with a group of length $i = 1, 2$ or 3, and since the remainder of the sequence of $m - i - 1$ elements must satisfy the grouping requirements as well, it is clear that $G_m = G_{m-1} + G_{m-2} + G_{m-3}$ for $m > 4$. To determine the number B_m of group sequences of $m - 1$ items which satisfy the requirement of not starting or ending with a group of length 3, we simply subtract the number of unconstrained group sequences which start or end with a grouping of 3. There are $2G_{m-3} - G_{m-6}$ such groupings, so that $B_m = G_m - 2G_{m-3} + G_{m-6}$ for $m > 4$. A simple algebraic verification shows that $B_m = B_{m-1} + B_{m-2} + B_{m-3}$ for $m > 4$. Values for $2 \leq m \leq 4$ are depicted in the following diagrams. Only the endpoint of each group is identified by a circle (○).

M_2	○

$m = 2$, $B_2 = 1$

M_2	○	
M_3	○	○

$m = 3$, $B_3 = 2$

M_2	○	○	
M_3	○		○
M_4	○	○	○

$m = 4$, $B_4 = 3$

Since $B_2 = 1$, $B_3 = 2$, $B_4 = 3$, and $B_m = B_{m-1} + B_{m-2} + B_{m-3}$ for $m > 4$, $B_m = T_{m-1}$, where T_m is the m^{th} Tribonacci number (Feinberg, 1963). Spickerman (1982) showed that $T_m = \lceil \alpha\rho^m + 0.5 \rceil$, where α and ρ are as above, which completes the proof. □

3.2 Simulations and simulation results

Theorem 2.8 shows that the maximal number of minimal covers grows exponentially with m, lying somewhere between the curves $y = (0.48)1.44^m$ and $y = (0.33)1.84^m$ for $m > 1$. Having in mind practical applications for minimal covers, we wish to determine whether, in practice, the number of minimal covers tends to reach these large values. We adopted a simulation approach with this purpose in mind, with the specific objectives of assessing the bounds mentioned above and of determining whether it is feasible to enumerate the minimal covers of an interval order likely to occur as a real data set. The simulation results we present are based on two pseudo-random interval order generation mechanisms designed to provide answers to both of these questions.

The generation mechanisms we present warrant a preliminary explanation. Recall from Chapter 1, § 1.1 that all interval orders $\underline{X}$ can be represented as sets of intervals on the real line

$$\{[h(x), h(x) + \rho(x)]\, ; x \in X, h : X \mapsto \mathbb{R}, \rho : X \mapsto \mathbb{R}^+\},$$

characterized by left endpoint function h and non-negative length function ρ, and such that for $x, y \in X$, $x \prec y$ if and only if $h(x) + \rho(x) < f(y)$. Thus, to generate random finite interval orders, it is enough to generate left and right endpoints defining real intervals.

In simulation series A, we systematically varied the number n of elements in X between 10 and 115, and the ratio of expected left-endpoint placement to interval length in the real representation. We produced real representations of interval orders by generating left endpoints according to an Exponential distribution with mean 1, then generating lengths according to an Exponential distribution with mean $\mu = 1/\lambda$, where λ took on the values $0.4, 0.6, \ldots, 3.0$. A simple calculation shows that the probability of overlap between any two intervals in such a setup is $1/(1+\lambda)$, and so varied between 0.25 and 0.72 in the course of our simulation. By way of comparison, the proportion of pairs of overlapping intervals in the breast cosmesis data presented in Finkelstein & Wolfe (1985), was approximately 0.45. For each pair (n, λ), 50 interval orders were generated, thus yielding 14,300 interval orders in total. Simulation series A was designed to produce a large variety of overlapping patterns which would help assess the bounds of Equation (2.2) (see Figure 2.7).

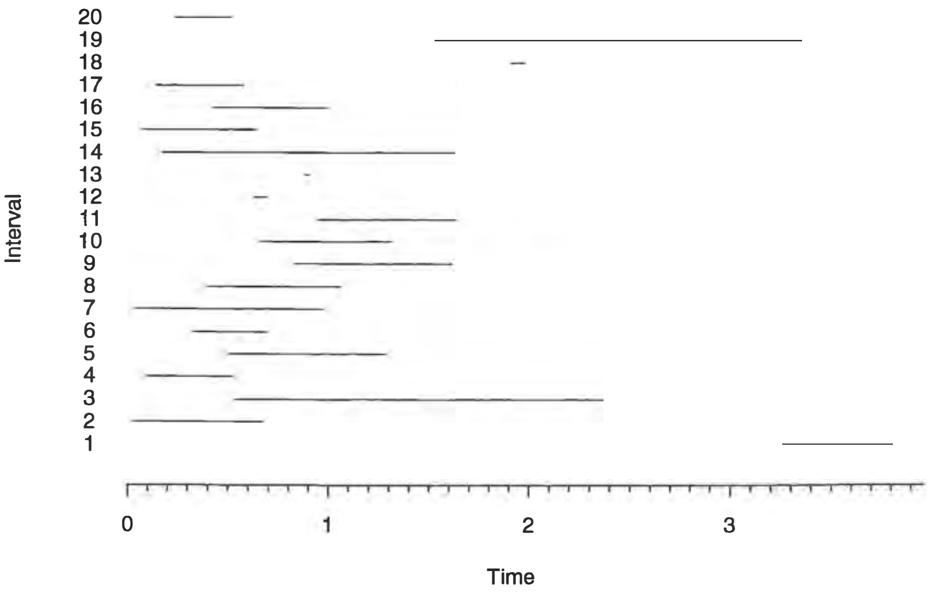


Figure 2.7: Series A sample interval set; $n = 20, \lambda = 0.5$.

For simulation series B, we generated intervals by setting potential inspection times at $t = 0, 1, 2, \ldots, 30$. For each interval , the inspection times were retained with a probability of 1 for $t = 0$, a probability of 0.4 for $t = 2, \ldots, 6$, and a probability of 0.1 for $t = 7, \ldots, 30$. A number n of event times were generated according to an Exponential distribution with mean μ, $\mu = 2, 2\times(1.25) = 3.5, 2 \times (1.25)^2 = 3.125, \ldots, 2 \times (1.25)^{15} = 56.8$, for values of $n = 10, 20, \ldots, 150$. Intervals were formed by using the largest inspection time smaller than the event time as the left endpoint, and the smallest inspection time larger than the event time as the right endpoint. This setup mimics a long-term prospective study in which a condition is monitored at fixed inspection times which may be missed; event time corresponds to the moment of change in condition. This simulation setup can produce right-censored cases (i.e. intervals without a finite right endpoint) with a probability of $\exp(-30/\mu)$. This probability ranged over $0.0667, 0.0833, \ldots, 0.53$. Sixty simulations were run for each pair (n, μ), thus yielding a total of 13,500 simulated interval orders. Simulation series B was designed to mimic typical data from a long term prospective study where a condition is periodically monitored for change (see Figure 2.8).

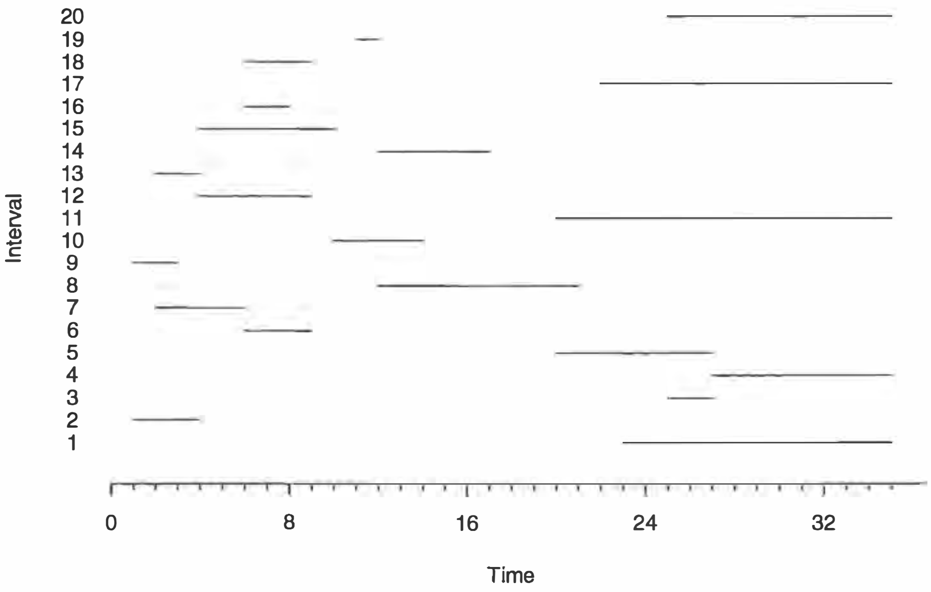


Figure 2.8: Series B sample interval set; $n = 20, \mu = 15$. (Intervals 1, 4, 11, 17 and 20 are right censored)

Simulation results are illustrated graphically using boxplots. The conventions we adhered to in the use of boxplots correspond to the standard S-Plus software conventions (Becker et al., 1988). The range represented by the inner shaded box corresponds to values of N lying between the first and third quartile of the data, with the line within this box indicating the median of the values.

The results of simulation series A are shown in Figures 2.9 and 2.10. The exponential growth of the number of minimal covers N with the cardinality n of $\underline{X}$ is manifest in both the average case and the maximal case, placing an exponential lower bound on this growth for the general interval order. An approximate value of ${}^{\mathrm{A}}N_{\mathbf{max}}(m)$, the maximum number of minimal covers in terms of m for simulation series A, can be determined by a Poisson regression of m on n with identity link. Such a regression yields $m \approx 0.366n + 0.805$, from which we can derive the approximation ${}^{\mathrm{A}}N_{\mathbf{max}}(m) \approx (0.485)1.023^{m}$. Figure 2.10 shows, as was expected, that this growth is more strongly associated with the increase in the number k of minimal elements in X rather than with the increase in n, though both quantities are positively correlated. Because simulations were not run an equal number of times for each value of k, the range of values of N as a function of k should not be interpreted as meaningful on the boxplots.

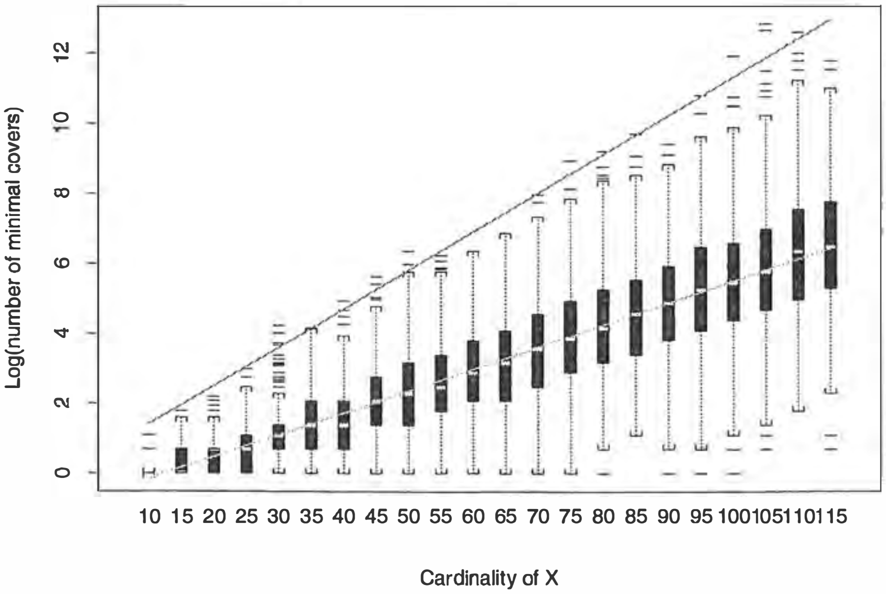


Figure 2.9: N as a function of n for simulation series A. Lower solid line is a least-squares regression line of $\log N$ on n; upper solid line is a least-squares regression line of $\log \max_n N$ on n.

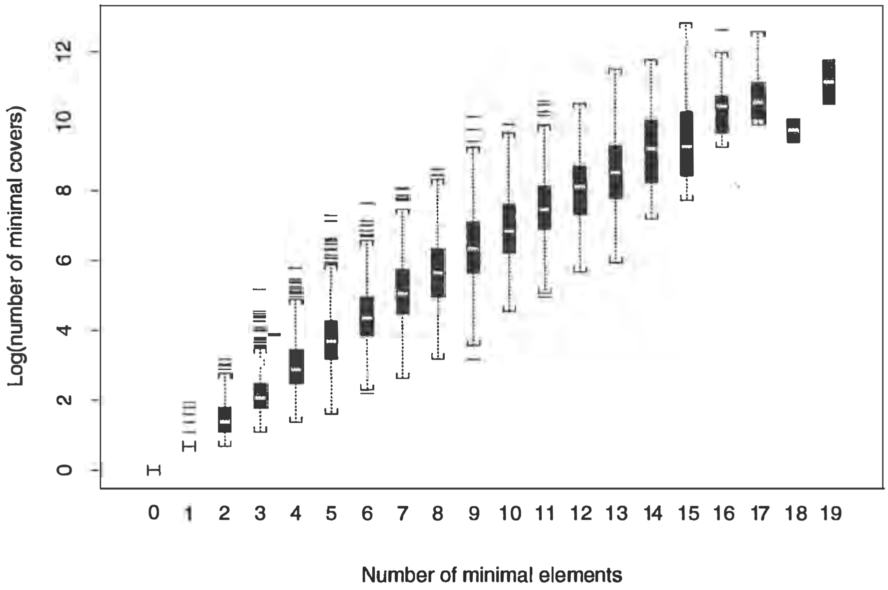


Figure 2.10: N as a function of k for simulation series A.

By contrast, simulation series B shows that the rate of increase of N with respect to n diminishes with increasing n in the average case, while the maximum value of N itself remains more or less constant for $n \geq 80$ (Figure 2.11). These results are explained by the fact that the simulation series B setup creates an expected proportion of right censored values which increases with μ; this increase causes the number of maximal antichains to converge in probability to 1 as μ grows larger. In the limit, all intervals overlap, forming a single maximal antichain and a single minimal cover. Thus N tends to 1 in probability. Figure 2.12, however, shows that the relationship between the number of minimal elements and the number of minimal covers remains roughly exponential, which indicates that right-censoring curbs the number of minimal covers by preventing the creation of large numbers of minimal elements.

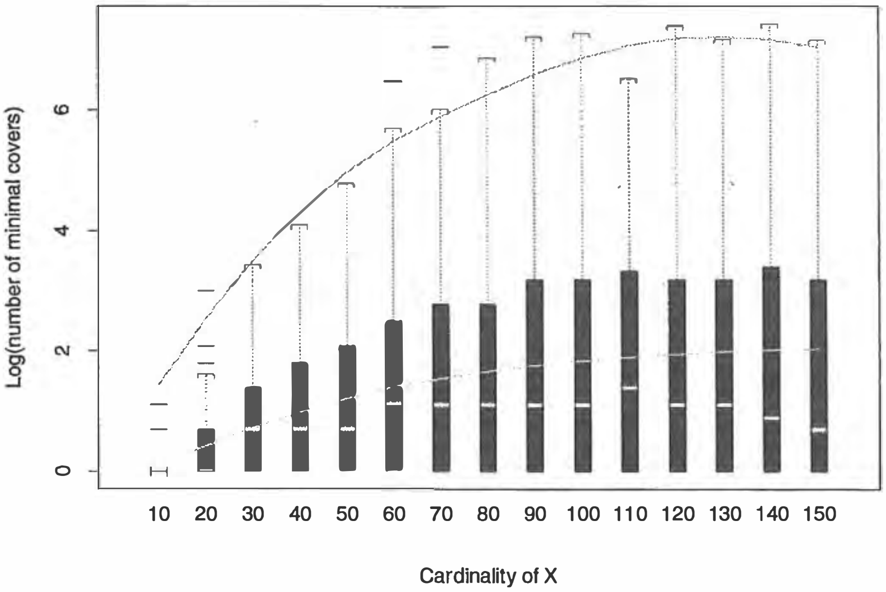


Figure 2.11: N as a function of n for simulation series B. Smooth curves were computed using local regression (*loess*).

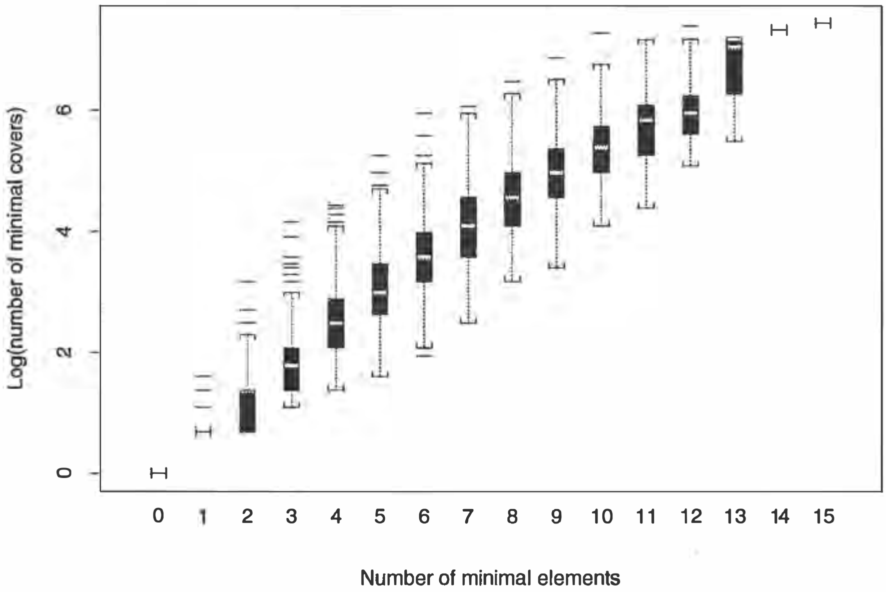


Figure 2.12: N as a function of k for simulation series B.

The results of simulation series A confirm a rapid exponential growth in the number of minimal covers, which could preclude their enumeration even for modestly sized interval orders. However, simulation series B provides an indication that we can expect the number of minimal covers in a realistic data set to remain manageable. The censoring mechanism itself, based as it is on fixed inspection times, forces the number of maximal antichains, and thus the number of minimal elements, to be at most the number of inspection times. We can expect this phenomenon to keep N manageable in applications.

Refining the bound on $N_{\mathbf{max}}(m)$ remains an open problem of theoretical interest. Of immediate practical interest is the determination of the expected number of minimal covers under a specific interval generation mechanism. Embedded semi-orders may provide an avenue of research for both of these questions, since minimal elements in an interval order form an embedded semi-order in the interval order. See Fishburn (1985), Chapter 6 for a discussion of maximal embedded semi-orders. This embedded semi-order need not be of maximal cardinality, since minimality of elements is a stronger constraint than semi-orderedness. Determining $N_{\mathbf{max}}(m)$ and the expected value of N under a given generation mechanism is likely to require knowledge about the attainable cardinality of such an embedded semi-order, but also about the patterns of maximal antichain membership in these embedded semi-orders.

Chapter 3

Weak order partition of interval orders

One form of nonparametric inference on interval censored data involves rank-score statistics. In the simplest case of rank scoring, a selected group of individuals are assigned to a "control" category while the remainder are assigned to a "treatment" category prior to any measurement being made. In the presence of complete information about the ordering of all individuals according to some measured quantity, we can assign a rank to every one of them. A function of this rank, usually but not necessarily monotonic, is then summed over the treatment individuals, and the resulting rank score statistic is usually assessed on the basis of its exact or asymptotic distribution under the null hypothesis of no difference between treatment and control.

However, information about the measured quantity is often only available as interval censored data. Prentice (1978) introduced a closed-form rank score statistic for right-censored data which consists of the average rank score statistic over all linear extensions of the partial order induced by censoring, or, equivalently, over all rank vectors compatible with the data. The basic nonparametric assumption underlying this technique is that every possible ordering of the underlying data is of equal weight a priori. Attempting to apply this principle to interval censored data is complicated by the fact that in this case there is no closed form for average rank score statistics.

To produce such statistics, it is therefore necessary to enumerate all linear extensions. However, enumeration is rarely feasible for actual data sets. Consider Example 1.1 from Chapter 1 viewed as interval censored data. It consists of 6 observations and has 57 linear extensions, as ascertained through a backtracking enumeration algorithm (see § 1), while the breast cosmesis data analyzed in Finkelstein & Wolfe (1985), with 93 observations, have approximately 2.4×10^{92} linear extensions, as estimated using the techniques described in § 3. The available alternative to enumeration is to produce a sampling estimate of the rank score statistic, based on uniformly generated linear extensions of the underlying interval order.

Self & Grossman (1986) offer several algorithms to perform uniform generation of linear extensions; unfortunately, none of these algorithms actually achieves uniform generation. In this chapter, we propose a structuring view of the set of linear extensions of interval orders through what we call *weak order partitioning*. In essence, we will construct a scheme whereby this set can be partitioned into disjoint sets, each of which consists of the linear extensions of a weak order. This partition greatly simplifies the task of both uniform generation of linear extensions and sampling estimation of various rank-based statistics of interval orders.

In § 1, we provide some detail about backtracking algorithms as they can be used to generate linear extensions; our purpose there is to provide insight into the problem, as backtracking-based generation algorithms are not up to the task of generating the linear extensions of an interval order uniformly. We construct the weak order partition in § 2. This is accomplished in Theorem 3.4, the main result of this chapter. The construction involves a number of definitions, most important among which is that of marked configurations of the interval order. By partitioning the linear extensions into simple sets, Theorem 3.4 yields a straightforward technique not only to generate linear extensions uniformly, but also to estimate linear rank statistics without generating the linear extensions directly. We provide a brief description of a general sampling application based on this partition in § 3, as well as describe a few typical questions concerning interval orders which may be addressed with the aid of standard sampling techniques.

1 Backtracking algorithms

Backtracking is a general method to count or enumerate linear extensions of a poset in constant linear time. Kalvin & Varol (1983) provide a review of backtracking algorithms applied to the enumeration of linear extensions. Even for small posets, the number of compatible linear extensions is large enough so as to make the method unusable in practice. However, for the insight it provides on the nature and complexity of the problem, we will describe backtracking using the precedence matrix as a starting point.

1.1 Counting the linear extensions of a poset

We illustrate how backtracking methods can be used to perform the conceptually simple task of counting the number of linear extensions of a partially ordered set. To do so, we introduce the precedence matrix of a partially ordered set.

Definition 3.1 *Let $(X, \prec)$ be a poset, with $X = \{x_1, x_2, \ldots, x_n\}$. Then we can define the $n \times n$* precedence matrix $\mathbf{P}$ *of $(X, \prec)$ as follows:*

$$\mathbf{P}_{ij} = \mathbf{1}[x_i \prec x_j], i, j = 1, \ldots, n.$$

Our convention that $\prec$ be irreflexive causes $\mathbf{P}$ to be zero-diagonal. Moreover, for transitive relations such as partial orders, it is always possible to reorder the labels in X so as to make $\mathbf{P}$ strictly upper triangular.

For $\mathbf{P} \in \{0,1\}^{n \times n}$ a (square) precedence matrix of size n, denote by $\mathbf{P}^j$ the jth column of $\mathbf{P}$, and by $\mathbf{P}_{\cdot}^j$ the sum of the elements of $\mathbf{P}^j$. We denote by $\mathbf{P}_{(j_1, \ldots, j_k)}$ the matrix $\mathbf{P}$ with rows $j_1, \ldots, j_k$ and columns $j_1, \ldots, j_k$ deleted, but with remaining rows and columns retaining their original labeling. The latter convention simplifies notation.

The idea behind backtracking enumeration is to recursively apply the following principle: only elements of X which have no predecessor can head a linear extension of $(X, \prec)$. These are easily identifiable in terms of the precedence matrix as elements with index j such that $\mathbf{P}_{\cdot}^j = 0$. Once initial element x_j has been chosen to head the linear extension, the problem is reduced to finding a linear extension for the poset with precedence matrix $\mathbf{P}_{(j)}$, and to let x_j head it.

We introduce further notation. For $\mathbf{P}$ as above:

- with j in the index set of rows and columns of $\mathbf{P}$, let $\nu_j(\mathbf{P}) = \mathbf{1}\left[\mathbf{P}_{\cdot}^j = 0\right]$;
- let $\mathcal{N}(\mathbf{P})$ be the number of linear extensions consistent with precedence matrix $\mathbf{P}$, with $\mathcal{N}([0]) = 1$.

Then $\mathcal{N}(\mathbf{P})$ can be expressed as

$$\mathcal{N}(\mathbf{P}) = \sum_{j=1}^{n} \nu_j(\mathbf{P}) \mathcal{N}\left(\mathbf{P}_{(j)}\right). \tag{3.1}$$

Repeatedly expanding the right-hand side of 3.1 yields

$$\mathcal{N}(\mathbf{P}) = \sum_{j_1=1}^{n} \nu_{j_1}(\mathbf{P}) \sum_{\substack{j_2=1\\ j_2\neq j_1}}^{n} \nu_{j_2}\left(\mathbf{P}_{(j_1)}\right) \sum_{\substack{j_3=1\\ j_3\neq j_1\\ j_3\neq j_2}}^{n} \nu_{j_2}\left(\mathbf{P}_{(j_1,j_2)}\right) \cdots \sum_{\substack{j_n=1\\ j_n\neq j_1\\ \vdots\\ j_n\neq j_{n-1}}}^{n} \nu_{j_n}\left(\mathbf{P}_{(j_1,j_2,\ldots,j_n)}\right)$$

so that

$$\mathcal{N}(\mathbf{P}) = \sum_{\sigma} \prod_{j=1}^{n} \nu_{\sigma_j}\left(\mathbf{P}_{(\sigma_1,\ldots,\sigma_{j-1})}\right),$$

where the sum is taken over all permutations σ of $\{1,\ldots,n\}$. We note that the factors in the summed products can be expressed in terms of the original $\mathbf{P}_{ij}$, since

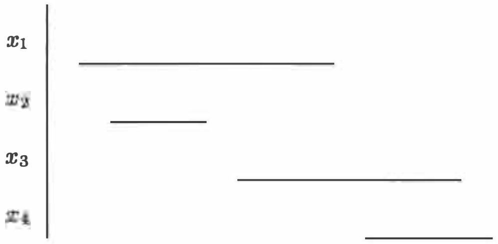
$$\nu_{\sigma_j}\left(\mathbf{P}_{(\sigma_1,\ldots,\sigma_{j-1})}\right) = \mathbf{1}\left[\sum_{i=j+1}^{n} \mathbf{P}_{(\sigma_i,\sigma_j)} = 0\right].$$

1.2 Enumeration and random generation

The recurrence formula (3.1) can be transformed into an algorithm which will enumerate the linear extensions themselves. Such an algorithm is equivalent to the algorithm proposed by Knuth & Szwarcfiter (1974) when indices are kept track of.

The enumeration algorithm, in turn, can easily be adapted as a pseudo-random generation algorithm: at every turn, choose the element to head the remaining sublist uniformly randomly among available candidates, until a linear extension is produced. Such a method does not produce linear extensions with uniform generation probability, as the following example demonstrates.

Example 3.1 *Consider the interval order $X = \{x_1, x_2, x_3, x_4\}$ endowed with the following real representation:*



Under the natural subscript ordering 1, 2, 3, 4 *of X, the precedence matrix of $\prec$ is*

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Representing linear extensions by their index sequence, the set of linear extensions $\mathcal{L}\left(\underline{X}\right)$ compatible with interval order $\underline{X} = (X, \prec)$ is

$$\mathcal{L}\left(\underline{X}\right) = \{(1,2,3,4),(1,2,4,3),(2,1,3,4),(2,1,4,3),(2,3,1,4)\}$$

The event that an element has been chosen to belong to the linear extension being generated is represented by the precedence matrix having the appropriate row and column struck out, and by leaving out the unchosen indices in the linear extension.

We can compute the probability of generation P_G for, say, extensions $(1,2,3,4)$ and $(2,1,3,4)$, as follows.

$$\begin{aligned}
P_G[(1,2,3,4)] &= P_G\left[(1,?,?,?)\right] P_G\left[(1,2,?,?)\left|\begin{array}{ccc} 2 & 3 & 4 \\ \hline 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right.\right] P_G\left[(1,2,3,?)\left|\begin{array}{cc} 3 & 4 \\ \hline 0 & 0 \\ 0 & 0 \end{array}\right.\right] P_G\left[(1,2,3,4)\left|\begin{array}{c} 4 \\ \hline 0 \end{array}\right.\right] \\
&= \frac{1}{2} \times 1 \times \frac{1}{2} \times 1 \\
&= \frac{1}{4}
\end{aligned}$$

$$\begin{aligned}
P_G[(2,1,3,4)] &= P_G\left[(2,?,?,?)\right] P_G\left[(2,1,?,?)\left|\begin{array}{ccc} 1 & 3 & 4 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right.\right] P_G\left[(2,1,3,?)\left|\begin{array}{cc} 3 & 4 \\ 0 & 0 \\ 0 & 0 \end{array}\right.\right] P_G\left[(2,1,3,4)\left|\begin{array}{c} 4 \\ 0 \end{array}\right.\right] \\
&= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times 1 \\
&= \frac{1}{8}
\end{aligned}$$

A generator with uniform generation probability across all linear extensions in $L(\prec)$ can be based on the backtracking random generator in the following manner:

Algorithm 3.1

BacktrackWithRejection($\underline{X}$)

Arguments: $\underline{X}$, a partially ordered set;

N, size of the sample to generate.

1. Find a quantity $a \leq \min_{\lambda \in \mathcal{L}(\underline{X})} P_G\left[\lambda\right]$.

2. Generate a linear extension λ through backtracking; include it in the sample with probability $a/P_G\,[\lambda$
3. Repeat step 2. until a sample of size N is generated.

Return sample.

Such an algorithm would generate any linear extensions in $\mathcal{L}\,(\underline{X})$ with equal probability. This algorithm stands in slight contrast with the closely related rejection scheme which generates a primary sample S, then resamples it by culling each of its members λ with probability $\min_{\lambda_0 \in S} P_G[\lambda_0]/P_G[\lambda]$, to create a secondary sample. Sample sizes are difficult to control with such an approach, since secondary samples issued from different primary samples cannot be merged without knowledge of the primary sample. On the other hand, finding an appropriate quantity a in Algorithm 3.1 may be costly.

In practice, we have found that neither approach is practicable because of the unacceptably high rejection rates. Attempts at uniform generation of linear extensions of the breast cosmesis data from Finkelstein & Wolfe (1985) ($n = 93$) using backtracking-based generation have led to acceptance rates of about 2.3×10^{-29}. This crippling acceptance rate has led us to consider alternative methods. Generation methods based on simple Markovian transitions, usually dubbed Monte Carlo Markov Chain methods, are attractive in many ways, not least of which is the simplicity with which some may be implemented. The difficulty in measuring the association between successively generated linear extensions using simple transitions, as well as the usual doubts concerning the convergence of simulated Markov chain to stationarity and the possibility of transitional bottlenecks in the set of linear extensions, have led us to consider a new approach.

2 Weak order partitioning of interval orders

The solution we settled on is expounded in the remainder of this Chapter. So-called *weak order partitioning* of interval orders holds the particular attraction of avoiding the generation of linear extensions altogether when linear statistics on the ranks of the data are desired. This property arises because a certain set of weak, rather than linear, extensions are produced by the method. Linear rank statistics on the linear extensions of a weak order can be computed directly by appealing to the average such rank statistic over each equivalence class of the weak order.

2.1 The marked configuration-induced weak order

We first formalize what we mean by partitioning a set of linear extensions using partial orders.

Definition 3.2 *For* $\underline{X} = (X, \prec)$ *a poset, an order partition of* $\mathcal{L}(\underline{X})$ *is a collection* $\{\mathcal{L}(X, R); R \in \mathcal{R}\}$ *such that*

1. $\mathcal{R} \subset (X \times X)$ *is a set of partial order relations on* X,
2. $\mathcal{L}(\underline{X}) = \bigcup_{R \in \mathcal{R}} \mathcal{L}(X, R)$, *and*
3. $\mathcal{L}(X, R_1) \bigcap \mathcal{L}(X, R_2) = \emptyset$ *for* $R_1 \neq R_2$, $R_1, R_2 \in \mathcal{R}$.

Because a partial order is the intersection of all the linear orders induced by its linear extensions (Theorem 1.4), we could also write dually that $\mathcal{R}$ is an order partition of $\prec$, in the above definition.

We will qualify the term "order" in "order partition" using the specific type of order contained in $\mathcal{R}$, viz. *weak order partition* in the current investigation.

By way of heuristics, we now construct a first weak order based on the maximal antichain structure of an interval ordered set. Recall that $\mathcal{M} = MA(\underline{X})$ denotes the set of maximal antichains of an interval order $\underline{X}$

Definition 3.3 *A configuration of an interval order* $\underline{X}$ *is a mapping* $C : X \to \mathcal{M}$ *where* $C(x) = M$ *only if* $x \in M$.

A configuration is simply an assignment of each element of X to a single maximal antichain of X. We will represent a configuration C by the matrix of indicators $\mathbf{C} = [c_{ij}; i = 1, \ldots, m; j = 1, \ldots, n]$, with $c_{ij} = \mathbf{1}[C(x_j) = M_i]$. Such a matrix will have a zero entry wherever the Petrie matrix of the interval order has a zero entry, and will contain a single 1 in every column. Configurations can also be represented by the sequence of integers $s_j = \arg_i [C_{ij} = 1]$, $j = 1, \ldots, n$, which we will call the *maximal antichain sequence* of C. We will use the same bolded letter to represent both the matrix and maximal antichain sequence of a configuration, relying on the obvious isomorphism between the representations.

We will also denote $\mathcal{C}(\underline{X}) = \{C;\ C \text{ is a configuration of } \underline{X}\}$.

Example 3.2 *Consider the interval order $\underline{X} = (X, \prec)$ corresponding to the data in Example 1.4, which has Petrie matrix* $\mathbf{A}$ *shown below. Then a configuration C of $\underline{X}$ could be represented by matrix* $\mathbf{C}$*:*

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

Equivalently, using maximal antichain sequence, we could put

$$\mathbf{C} = (1, 3, 1, 4, 4, 4).$$

Each configuration C of $\underline{X}$ induces a weak order on X in a simple way, as indicated in the next Lemma, the simple proof of which is again omitted.

Lemma 3.2 *Let $\underline{X} = (X, \prec)$ be an interval order. Let C be a configuration of $\underline{X}$ be represented by the maximal antichain sequence $\mathbf{C} = (s_1, \ldots, s_n)$, and define the relation $<_C$ by*

$$x_a <_C x_b \Leftrightarrow s_a < s_b \text{ for all } x_a, x_b \in X.$$

Then $<_C$ is a weak order on X, $\prec \subset <_C$ and $\approx_C = \sim_C$ is given by $x_a \approx_C x_b \Leftrightarrow s_a = s_b$.

We can conclude from Definition 1.4 and Lemma 3.2 that two configurations C and C' of an interval order given respectively by maximal antichain sequences $\mathbf{C} = (s_1, \ldots, s_n)$ and $\mathbf{C}' = (s'_1, \ldots, s'_n)$ will induce weak orderings with linear extensions in common, in the sense that $\mathcal{L}(X, <_C) \bigcap \mathcal{L}(X, <_{C'}) \neq \emptyset$) whenever $s_a \leq s_b \Leftrightarrow s'_a \leq s'_b$ for all $a, b \in 1, \ldots, n$.

Example 3.3 *Using the interval order from Example 3.2 above, we consider the two configurations C and C' represented by matrices*

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{C}' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Indicating equivalence classes by curly braces $\{\cdot\}$, the induced weak orders of Lemma 3.2 are

$$\{x_1, x_3\} <_C \{x_2\} <_C \{x_4, x_5, x_6\}$$

and

$$\{x_1, x_2, x_3\} <_{C'} \{x_4\} <_{C'} \{x_6\} <_{C'} \{x_5\}.$$

It is easy to see that

$$\mathcal{L}(X,<_C)\bigcap\mathcal{L}(X,<_{C'})=\{(x_1,x_3,x_2,x_4,x_6,x_5),(x_3,x_1,x_2,x_4,x_6,x_5)\}\neq\emptyset.$$

Thus $\{\mathcal{L}(X,<_C);C\in\mathcal{C}(\underline{X})\}$ is not in general a weak order partition for $\mathcal{L}(X,\prec)$. The fact that it does form a cover for $\mathcal{L}(X,\prec)$ will be shown in the course of proving Theorem 3.4. Configurations, however, form the first of two main building blocks which we will use to construct such a partition.

Before proceeding, we introduce further notation:

Definition 3.4 *For C a configuration of interval order $\underline{X}$, let $M_i(C)=\{x_j\in M_i;C(x_j)=M_i\}$, $i=1,\dots,m$, and let $M_{m+1}(C)=\emptyset$. We call $M_i(C)$ a configuration maximal antichain of C.*

We define now the concept of marker, the second building block in the construction of a weak order partition of $\mathcal{L}(X,\prec)$:

Definition 3.5 *For C a configuration of interval order $\underline{X}$ and $i=1,\dots,m$, we say that $M_i(C)$, can be marked or is markable if there exists $x\in M_i(C)$ with $x\notin M_{i+1}$.*

Such an x (including therefore any $x\in M_m$) will then be called a marker for maximal antichain M_i.

A configuration can be marked or is markable if all of its configuration maximal antichains are markable.

A marker, in a particular configuration, is thus simply an element which appears in the configuration maximal antichain corresponding to the maximal antichain in which it last occurs in the original interval order; "last occurs" is here, as usual, used in the sense of the maximal antichain linear ordering. Every element of X is a marker for exactly one maximal antichain in an interval ordering of X, in the sense of Chapter 1, § 1.3.

It must be noted that not every configuration can be marked; specifically, if, for some i, $M_i(C)$ is non-empty but does not contain a marker, then C cannot be marked (see Example 3.4 below).

Definition 3.6 *A marked configuration of an interval order is a mapping $C^*=(C_1^*,C_2^*):X\to\mathcal{M}\times\{0,1\}^m$ where C_1^* is a markable configuration of $\underline{X}$ and C_2^* is such that for every $M\in C_1^*(X)$ there is exactly one marker $x\in M$ with $C_2^*(x)=1$.*

One useful way to represent a marked configuration of $\underline{X}$ is as a matrix-vector pair (or maximal antichain sequence-vector pair) $(\mathbf{C}, \omega)$, where

1. $\mathbf{C}$ is the matrix or maximal antichain sequence of a markable configuration of $\underline{X}$;
2. ω is a vector $[\omega_i], i = 1, \ldots, m$, where ω_i is undefined if $M_i(C) = \emptyset$, and is a marker for $M_i(C)$ otherwise.

Again, we will use C^* to denote either the mapping (C_1^*, C_2^*) and the bold version $\mathbf{C}^*$ to denote the matrix-vector pair representation $(\mathbf{C}, \omega)$.

The example below should help to clarify our definitions.

Example 3.4 *Using the interval order and configuration from Example 3.3, and identifying ω only through element subscripts, we could have a marked configuration C^* with matrix-vector representation*

$$\mathbf{C}^* = (\mathbf{C}, \omega) = \left(\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 \\ - \\ 3 \\ 5 \end{bmatrix} \right),$$

where the $-$ indicates that the marker is undefined, an allowable state since $M_3(C) = \emptyset$.

We note that configuration C' in Example 3.3 cannot be marked, since $M_2(C') = \{x_2\}$, and x_2 cannot be a marker for M_2 since it last occurs in M_3.

Marked configurations may be used to induce weak order relations on X in the following manner.

Lemma 3.3 *Let $\underline{X}$ be an interval order. Let C^* be a marked configuration of $\underline{X}$ with maximal antichain sequence-vector representation $\mathbf{C}^* = [(s_1, \ldots, s_n), \omega]$, and define the relation $<_{C^*}$ as follows for all $x_a, x_b \in X$:*

$$x_a <_{C^*} x_b \Leftrightarrow [s_a < s_b \text{ or } x_b = \omega_{s_a}].$$

Then $<_{C^*}$ is a weak order on X, $\prec \subset <_{C^*}$, and $\approx_{C^*}$, for $x_a \neq x_b$ is given by $x_a \approx_{C^*} x_b \Leftrightarrow [s_a = s_b$ and $x_a \neq \omega_{s_b} \neq x_b]$.

Example 3.5 *The marked configuration of Example 3.4 above yields the following weak order, in the sense of Lemma 3.3:*

$$\{x_3\} <_{C^*} \{x_1\} <_{C^*} \{x_2\} <_{C^*} \{x_4, x_6\} <_{C^*} \{x_5\}.$$

Elements identified as markers occur immediately after all the elements of the configuration maximal antichain to which they are mapped.

The construction of $<_{C^*}$ shows that $(<_C) \subset (<_{C^*})$ for any markable configuration C and compatible vector of markers ω. The following Theorem shows that this tightening of configuration-induced weak orders is enough for us to meet our goal of partitioning the linear extensions of an interval order in simple subsets.

Theorem 3.4 *Let $\underline{X}$ be an interval order. Let $\mathcal{C}^*(\underline{X}) = \{C^*; C^* \text{ is a marked configuration of } \underline{X}\}$. Then the weak orderings induced by $\mathcal{C}^*(\underline{X})$ by the construction of Lemma 3.3 form a weak order partition of $\mathcal{L}(X, \prec)$.*

Proof. We first show that the weak orders $<_{C^*}$ cover the set of linear extensions, by constructing a marked configuration for each linear extension in $\mathcal{L}(X, \prec)$.

Let $\lambda = (x_{j_1}, x_{j_2}, \ldots, x_{j_n}) \in \mathcal{L}(\underline{X})$. For convenience, we let $y_k = x_{j_k}$, $k = 1, \ldots, n$. The linear ordering of the subscripts $k = 1, \ldots, n$ will then correspond exactly to the ordering of the y's in λ. We construct a configuration C by completely determining its dual sets $M_i(\cdot)$, $i = 1, \ldots, m$. As we construct the dual sets, we identify markers by entering them in a marker vector ω. The procedure is effected in such a way that λ belongs to the induced weak order. The induced ordering will be denoted $<_{C,\omega}$ during the course of this proof.

Place y_n in $M_m(C)$, which can always be done since the last element of a linear extension of X must belong to the last maximal antichain, by the linear ordering of the maximal antichains and the essentiality of M_m. Let $\omega_m = y_n$. Also assign $y_{n-1}, y_{n-2}, \ldots, y_{k_m}$ to $M_m(C)$, where k_m is chosen so that $y_{k_m} \in M_m$ but $y_{k_m - 1} \notin M_m$. Thereafter, assign $y_{k_r - 1}, y_{k_r - 2}, \ldots, y_{k_{r-1}}$ to $M_{i_{r-1}}(C)$, where $M_{i_{r-1}}$ is the largest maximal antichain to which $y_{k_r - 1}$ belongs, and where $k_{r-1} = 1$ or is chosen so that $y_{k_{r-1}} \in M_{i_{r-1}}$ but $y_{k_{r-1} - 1} \notin M_{i_{r-1}}$. Put $\omega_{i_{r-1}} = y_{k_r - 1}$, which is possible since $y_{k_r - 1} \notin M_{i_{r-1}+1}$ by construction. Since all y's belong to at least one maximal antichain, they will all be assigned to a configuration maximal antichain after a finite number of steps, say s steps, yielding $M_{i_r}(C) = \{y_{k_r}, \ldots, y_{k_{r+1}-1}\}$, $\omega_{i_r} = y_{k_{r+1}-1}$, $r = m - s, \ldots, m$, where $k_{r+1} = m + 1$.

If $y_a <_{C,\omega} y_b$, then $a < b$, for either $M_C(y_a) < M_C(y_b)$, in which case $a < b$ by construction, or $y_b \in M_C(y_a)$ and y_b is the selected marker for $M_C(y_a)$ in which case $a < b$ obtains again by construction. Hence the weak order induced by (C, ω) has precedence relations only where the subscripts $k = 1, \ldots, n$ also have precedence in the integer ordering; therefore $<_{C,\omega} \subset <_\lambda$. Extending the construction to every $\lambda \in \mathcal{L}(X, \prec)$ shows that

$$\mathcal{L}(X, \prec) \subset \bigcup_{(C,\omega) \in \mathcal{C}^*(\underline{X})} \mathcal{L}(X, <_{C,\omega}).$$

But since

$$(\prec) \subset (<_C) \subset (<_{C,\omega}) \Rightarrow \mathcal{L}(X, <_{C,\omega}) \subset \mathcal{L}(X, <_C) \subset \mathcal{L}(X, \prec) \text{ for all } (C, \omega) \in \mathcal{C}^*(\underline{X}),$$

this shows that

$$\mathcal{L}(X, \prec) = \bigcup_{(C,\omega) \in \mathcal{C}^*(\underline{X})} \mathcal{L}(X, <_{C,\omega}).$$

We must now show that $\mathcal{L}(X, <_{C,\omega}) \bigcap \mathcal{L}(X, <_{C',\omega'}) = \emptyset$ for all $(C, \omega) \neq (C', \omega')$, with $(C, \omega), (C', \omega') \in \mathcal{C}^*(\underline{X})$.

If $C = C'$, then there is an i such that $\omega_i \neq \omega_i'$. Thus $\omega_i' <_{C,\omega} \omega_i$, but $\omega_i <_{C',\omega'} \omega_i'$, so $\mathcal{L}(<_{C,\omega}) \bigcap \mathcal{L}(X, <_{C',\omega'}) = \emptyset$.

Ift $C \neq C'$ but $\omega = \omega'$, then there exists $x \in M_i(C)$ with $x \in M_j(C')$ for some $i \neq j$. Without loss of generality, assume that $j < i$; then $\omega_j' <_{C,\omega} x$ but $x <_{C',\omega'} \omega_j'$.

If $C \neq C'$ and $\omega \neq \omega'$, consider for some i the configuration maximal antichain $M_j(C')$ which contains ω_i. Then either $M_j = M_i$ or $M_j < M_i$, since M_i is the last maximal antichain where ω_i can occur, by definition of markers.

If $M_j = M_i$, then $\omega_i <_{C',\omega'} \omega_j'$, but since ω_j' last occurs in maximal antichain $M_j = M_i$, it must be that $\omega_j' <_{C,\omega} \omega_i$.

If $M_j < M_i$, then $\omega_i <_{C',\omega'} \omega_j'$ by construction. But since ω_j' occurs before $M_i(C)$, $\omega_j' <_{C,\omega} \omega_i$.

In all cases, $\mathcal{L}(X, <_{C,\omega}) \bigcap \mathcal{L}(X, <_{C',\omega'}) = \emptyset$. □

3 Sampling marked configurations

Marked configurations partition the space of linear extensions of an interval order, and it is therefore natural to use them to construct a sampling plan on that space. The most obvious applications of Theorem 3.4 will require pseudo-random generation of marked configurations according to a known probability measure. In 3.1, we provide a general scheme whereby marked configurations can be randomly generated. We describe some of the applications in 3.2.

3.1 Generating marked configurations

Our objective will be to describe a procedure to generate marked configurations with preset probabilities, and to sample appropriately from the marked configurations. A few more concepts related to covers of maximal antichains are necessary before we can make this procedure explicit.

Definition 3.7 *For $\underline{X}$ a partial order, we call $\mathcal{W} \subset \mathcal{M}$ a cover of X if $X = \bigcup_{M \in \mathcal{W}} M$.*

Covers will ultimately provide us with a practical method for generating marked configurations. The technique is described below, for interval order $\underline{X}$ with $|\mathcal{M}| = m$.

I. Generate a cover $\mathcal{W}$ for X with known probability $p(\mathcal{W})$.

II. Select a marker for each $M \in \mathcal{W}$.

III. Select a maximal antichain from $\mathcal{W}$ for each $x \in X$ which is not a marker.

Whereas II. and III. are straightforward and even suggest natural uniform probabilities of selection, viz. $|\{\text{markers of } M\}|^{-1}$ for marker selection and $|\{M \in \mathcal{W}; x \in M\}|^{-1}$ for maximal antichain selection, generating a set of maximal antichains $\mathcal{W}$ which covers X is a more difficult task. To resolve this difficulty, we appeal to minimal covers.

To generate a cover with known probability, the following procedure may be used:

Step 1. Generate a minimal cover $\mathcal{W}_0$ uniformly. Put $b = |\mathcal{W}_0|$.

Step 2. Generate a value a, $0 \leq a \leq m - b$, according to predetermined probability $p_b(a)$.

Step 3. Select a set $\mathcal{W}_1$ of a maximal antichains from $\mathcal{M} \setminus \mathcal{W}_0$ uniformly and set $\mathcal{W} = \mathcal{W}_0 \bigcup \mathcal{W}_1$ to be the generated cover.

Step 4. Determine the number and cardinality of all minimal covers which are subsets of $\mathcal{W}$, and thus determine the probability $p(\mathcal{W})$ of generating $\mathcal{W}$.

A few notes on each of the steps:

- **Step 1. Generate a minimal cover uniformly.**

The uniform generation techniques for minimal covers are based on Algorithm RandomMinCover, a non-uniform pseudo-random minimal cover generator based on Algorithm 2.4.

Algorithm 3.5

```
RandomMinCover($Y$)
Argument: $Y$, the set of minimal elements of an interval order;
$\mathcal{W} \leftarrow \emptyset$
$\mu_\rfloor \leftarrow \emptyset$
$p \leftarrow 1$
While $Y = \emptyset$
    Begin
    $y_0 := \operatorname*{argmin}_{y} \max_{y \in Y} y^*$
    $T \leftarrow \{M \in y_0^*; \mu_\rfloor < M\}$
    $\mu_0 \leftarrow$ ChooseRandom$(T)$
    $p \leftarrow \frac{p}{|T|}$
    $\mathcal{W} \leftarrow \mathcal{W} \cup \{\mu_0\}$
    $Y \leftarrow \{y \in Y; \max y_0^* < y^*\}$
    End
Return $(\mathcal{W}, p)$
```

The function **ChooseRandom** called with a set of maximal antichains of size k returns a uniformly randomly selected maximal antichain from the set, that is, every maximal antichain has probability $1/k$ of being returned. The values returned by Algorithm RandomMinCover are a pseudo-randomly generated minimal cover $\mathcal{W}$ along with its probability of generation p.

Generating a minimal cover uniformly can be done using one of the following two methods.

1. Finding the number of minimal covers is a less onerous task than enumerating them; if this number is small enough, minimal covers can be listed using Algorithm SimplifiedList-

MinCovers, and uniform generation of a minimal cover reduces to random selection from the list.

2. If the number of minimal covers is too large for enumeration, a Monte Carlo Markov Chain may be set using the Metropolis-Hastings algorithm described in Hastings (1970). Transitions are then based on proposals generated independently from the previous state, which provides the chain with a geometric rate of convergence to its stationary distribution (Smith & Tierney (1996)). A proposal is obtained by generating a minimal cover for the interval order using Algorithm RandomMinCover, with known probability of generation.

- **Step 2. Generate a number a of maximal antichains to add to the minimal cover.**

Known probabilities of generation do not guarantee efficient sampling. Since, at least conceptually, our ultimate sampling units are linear extensions, the most efficient generation scheme would have probability of generation proportional to the number of linear extensions belonging to the marked configuration-induced weak order, a set-up known as sampling with probability proportional to size Thompson (1992, Chapter 6).

It is dubious whether this ideal may be reached at all in our setting; however, a probability of generation which roughly increases with the weight of the marked configuration will ensure acceptable efficiency. We first note that weak orders containing large equivalences classes will have large sets of linear extensions. Larger equivalence classes occur in weak orders with fewer equivalence classes. A positive association between the probability of generating a weak order and the size of its linear extension set can therefore be achieved by setting $p_b(a)$ so as to place greater probability mass on smaller values of a. A discretized truncated exponential distribution probability function provided empirically good results (see Figure 3.1). Data-driven schemes are also possible, but have not been explored by us.

The cover generation procedure was designed specifically to include Step 2. as a device to achieve approximate positive relationship between probability of generation and size of the linear extension set. Other generation schemes were unsuccessful in that they were not efficient.

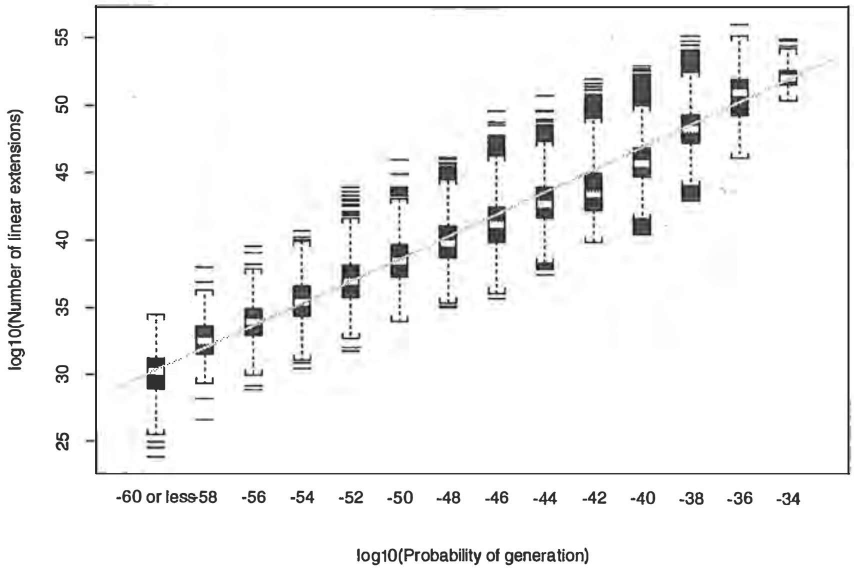


Figure 3.1: *Boxplots on log-log scale relating the number of linear extensions in a marked configuration-induced weak order to the probability of generating the marked configuration, for a sample of 100,000 marked configurations of the interval order underlying the breast cosmesis data (93 observations) analyzed in* [3]. *The straight line is a least-squares fit of the original data, and has slope* $\approx$ *0.83.*

- **Step 3. Select a subset of** a **maximal antichains uniformly to add to the minimal cover.**

  Uniformity is trivially achieved when the number of maximal antichains to add has been fixed beforehand.

- **Step 4. Determine the probability of generation of** $\mathcal{W}$**.**

  Because there may be more than one way of generating $\mathcal{W}$ using the method we are describing, the probability of its generation cannot be computed on the fly. In order to determine this probability, all possible pathways to the generation of $\mathcal{W}$ must be considered. A few features of interval orders and of our cover generation scheme conspire to make this task relatively simple.

  – Any cover of an interval order represents the full set of maximal antichains of another interval order; any minimal cover for this new interval order is also a minimal cover for the original interval order.

- Algorithm ListMinCover can be modified to list the number of minimal covers of an interval order according to the number of maximal antichains they contain.
- A non-zero probability of generating a cover from a given minimal cover depends only on the difference in cardinalities.
- Minimal covers are generated uniformly.

These observations can be brought together as follows:

Let $\mathcal{W}$ be a cover for interval order $\underline{X}$. Let $r = r(\mathcal{W})$ be the number of distinct cardinalities for minimal covers which are subsets of $\mathcal{W}$. Let $b_k = b_k(\mathcal{W})$, $k = 1, \ldots, r$, be the distinct cardinalities of minimal cover of $\mathcal{W}$, and let $c_k = c_k(\mathcal{W})$ be the number of minimal covers of $\mathcal{W}$ bearing the corresponding cardinality b_k. We can write down the probability of generating $\mathcal{W}$ as

$$p(\mathcal{W}) = \sum_{k=1}^{r} c_k p_{b_k} \left(|\mathcal{W}| - b_k \right)$$

3.2 Applications

The weak order partition of interval orders has strong potential for a variety of tasks involving interval orders, including uniform generation of linear extensions and estimation of various statistics of the linear extension set. We give four examples.

Linear extensions. To generate linear extensions uniformly, two-stage sampling (see Cochran (1977, Chapter 11), for instance) may be used. Marked configurations are sampled with replacement at the first stage, and their probability of generation is determined after they have been generated. Then an appropriate number of linear extensions are generated with replacement from the induced weak orders at the second stage. Generating linear extensions from a weak order is a very simple task, requiring only unconstrained shuffling of the weak order's equivalence classes while the order between equivalence classes is maintained. Basic sampling techniques allow the generation of linear extensions to be uniform.

Number of linear extensions. The number of linear extensions of an interval order may be estimated using marked configurations. A given weak order has a number of linear extensions which is simply computed: if E_i, $i = 1, \dots, r$ are the equivalence classes of weak order $\underline{Y}$, then $|\mathcal{L}(\underline{Y})| = \prod_{i=1}^{r} |E_i|!$. If C_s^*, $s = 1, \dots, S$ are a sample of marked configurations of interval order $\underline{X}$, each sampled with replacement with probability p_s, then the Hansen-Hurwitz estimator (see, e.g. Thompson, 1992, Chapter 3) is an unbiased estimator for $|\mathcal{L}(\underline{X})|$ and is given by

$$\widehat{|\mathcal{L}(\underline{X})|} = \frac{1}{S}\sum_{s=1}^{S} \frac{|\mathcal{L}(X, \prec_{C_s^*})|}{p_s}.$$

An unbiased estimate of its variance is given by

$$\widehat{\mathrm{Var}}\left(\widehat{|\mathcal{L}(\underline{X})|}\right) = \frac{1}{S(S-1)}\sum_{s=1}^{S}\left(\frac{|\mathcal{L}(X, \prec_{C_s^*})|}{p_s} - \widehat{|\mathcal{L}(\underline{X})|}\right)^2$$

Rank score statistics. Rank score statistics, which can be based on the average rank score of elements of X in the set of linear extensions, are simple to deal with at the level of weak orders. If, in a given weak order $\prec_0$, an element x belongs to an equivalence class E such that $\underline{k} = |\{x \in X; x \prec_0 E\}|$ and $\overline{k} = |\{x \in X; E \prec_0 x\}$, then x will take on ranks $\underline{k}+1, \underline{k}+2, \dots, \overline{k}-1$ an equal number of times amongst the linear extensions of $(X, \prec_0)$. If $c : X \rightarrow \mathbb{R}$ is a rank score function which depends only on the rank of x in a linear order $(x_{(1)}, \dots, x_{(n)})$ of X and on a covariate $Z(x)$ of the observation, then the value of the rank score of x over all linear extensions of $(X, \prec_0)$ will simply be $(\overline{s} - \underline{s} - 1)^{-1} \sum_{i=\underline{s}+1}^{\overline{s}-1} c\left(x_{(i)}, Z(x_{(i)})\right)$. Summing these average rank scores over all relevant rank and covariate values yields an average rank score statistic $\overline{\rho}_s$ for the weak order.

This simple property of weak orders along with the ease with which linear extensions are counted within them makes single-stage cluster sampling estimates of rank score statistics simple and readily available. A slightly biased but efficient estimate for the average rank score statistic over all linear extensions is given by

$$\widehat{\overline{\rho}} = \left[\sum_{s=1}^{S} \frac{|\mathcal{L}(X, \prec_{C_s^*})|\,\overline{\rho}_s}{p_s}\right] \Big/ \left[\sum_{s=1}^{S} \frac{|\mathcal{L}(X, \prec_{C_s^*})|}{p_s}\right]$$

See Särndal et al. (1992, Chapter 4).

Precedence proportions. Given two elements x, y of a partial order $\underline{X} = (X, \prec)$, we define the precedence proportion of x over y as $|\{\lambda \in \mathcal{L}(\underline{X}); x <_\lambda y\} / |\mathcal{L}(\underline{X})|$, the proportion of linear extensions of $\underline{X}$ where x precedes y. Precedence proportion has been studied in the related settings of correlation inequality (e.g. Winkler, 1986) and linear extension majority (e.g. Brightwell, Fishburn & Winkler, 1993). Precedence proportions in a weak order are always 0, 0.5 or 1, and sampling estimates of precedence proportions in an interval order using marked configuration-induced weak orders are thus simple to compute. The sampling estimate for precedence proportions takes on the same form as for rank score statistics.

Chapter 4

Nonparametric estimation of the CDF

In this chapter, we pursue the work initiated in Chapter 1 concerning nonparametric estimation of the CDF for interval censored data. In particular, we will endeavour to reconcile two perspectives on nonparametric maximum likelihood estimation for interval censored data which have gained wide currency during the past twenty-five years or so.

The first of these perspectives is the point of view represented in works such as Peto (1973), Turnbull (1976) and Gentleman & Geyer (1994), to name a few. It models the probability density function (PDF) observed under the censoring pattern as a discrete probability function on a set of disjoint intervals determined by the data. The CDF is undefined within these intervals and remains unvarying outside of them. The basic tool in this perspective is self-consistency and its natural implementation, the EM algorithm. Convergence of the EM algorithm to the nonparametric maximum likelihood estimate (NPMLE) of the PDF is guaranteed when the Kuhn-Tucker conditions are satisfied. Böhning, Schlattmann & Dietz (1996) propose, with the same view on the form of the NPMLE and again using the Kuhn-Tucker conditions, algorithms based on the Vertex Exchange Method and the Vertex Displacement Method which appear to be much more efficient than the EM algorithm.

The second point of view is based mostly on the work of Groeneboom (1991) and Groeneboom & Wellner (1992), as well as Jongbloed (1998) and Wellner & Zhan (1997). In this case, the NPMLE of the CDF is assumed to have jumps on the positive real line only at the endpoints of the intervals forming the data. Isotonic regression is used for estimation of the CDF in this setting, with the Iterative Convex Minorant algorithm and some variants thereof as its main implementations. The assumed form of the CDF and Fenchel duality conditions on the estimate (see § 3) guarantee the convergence of the modified Iterative Convex Minorant algorithm proposed by Jongbloed (1998) to the NPMLE.

Our aim within this Chapter is to show that apparently disparate perspectives on nonparametric likelihood estimation can actually share a single descriptive language rooted in order theory. This bridge is based on the discretization of Turnbull (1976), formalized in Theorem 1.10. An added value to the language of order theory is that some of its objects are easily expressed in matrix/vector notation, with a few extensions introduced in Appendix A. The simplicity of the notation and of the underlying computations may contribute to a more widespread use of correct methods to analyze interval censored data.

The main order theoretic concepts we will call upon are the linear ordering of maximal antichains in an interval order, and the structure of the data's duals with respect to their maximal antichains, embodied in the characteristic matrix. In § 2, we state the nonparametric likelihood of the event time CDF and revisit real-valued intervals in accordance with order theoretic concepts. One of the results which ensues naturally is the uniqueness of the CDF NPMLE on the set of maximal antichains of the interval order. In § 2, we discuss the non-uniqueness and the effective support of self-consistent estimates, and relate them to covers of maximal antichains. We also show how the simple distinction between essential and non-essential maximal antichains yields bounds on self-consistent estimates, and discuss recent work by Mykland & Ren (1996) and Wellner & Zhan (1997) in the light of our observations. In § 3, we apply the approach of Groeneboom & Wellner (1992) to our discrete setting. We show the similarity of the Fenchel duality conditions proposed by Groeneboom & Wellner (1992) and of self-consistency augmented by the Kuhn-Tucker conditions, as proposed by Gentleman & Geyer (1994). We then apply the Iterative Convex Minorant algorithm to the set of maximal antichains of the interval order represented by the data.

1 The likelihood and $\mathcal{M}$

The explicit mapping $H_{\mathcal{M}}$ of Theorem 1.10 determines the support of the NPMLE of F. We can carry out estimation strictly on the interval order and its maximal antichains, knowing that the results can be assigned to a real setting after we are done. An important aspect of the linear ordering of maximal antichains in interval orders is that we will remain justified in dealing with a cumulative distribution function on $\mathcal{M}$.

To see what is being estimated when we maximize the nonparametric likelihood, consider the image distribution of F restricted in its support to $H_{\mathcal{M}}(\mathcal{M})$ and unidentifiable within this support. To simplify notation, we extend the domain of the inverse mapping $H_{\mathcal{M}}^{-1}$ to real values in $H_{\mathcal{M}}(\mathcal{M})$, setting $H_{\mathcal{M}}^{-1}(x) = M$ if $x \in H_{\mathcal{M}}(M)$. We then obtain, with $t \notin H_{\mathcal{M}}(\mathcal{M})$ and $M(t) = \operatorname{argmax}_{M \in \mathcal{M}} [\sup H_{\mathcal{M}}(M) \leq t]$,

$$\begin{aligned} F(t \,|\, T \in H_{\mathcal{M}}(\mathcal{M}))) &= \text{Prob}\,[T \leq t \,|\, T \in H_{\mathcal{M}}(\mathcal{M})] \\ &= \text{Prob}\,[T \leq \sup M(t)) \,|\, T \in H_{\mathcal{M}}(\mathcal{M})] \\ &= \text{Prob}\left[H_{\mathcal{M}}^{-1}(T) \leq M(t) \,|\, T \in H_{\mathcal{M}}(\mathcal{M})\right] \\ &= F_{\mathcal{M}}(M(t)) \end{aligned}$$

where $F_{\mathcal{M}}$ is the distribution function of maximal antichains, uniquely identified for every equivalence class of distribution functions which are unidentifiable on $H_{\mathcal{M}}(\mathcal{M})$ and only on that set. The passage from $\sup M(t)$ to $M(t)$ in the second and third lines is justified by the right continuity of distribution functions.

Writing X_n for X and $\mathcal{M}_n$ for $\mathcal{M}$, the NPMLE of the event time CDF conditional on the censoring pattern with a sample size of n is then just

$$\hat{F}_n(t \,|\, q_{jk}, j = 1, \ldots, n; k = 1, \ldots, K_j) = \hat{F}_{\mathcal{M}_n}(M(t)) \text{ for } t \notin H_{\mathcal{M}_n}(\mathcal{M}_n).$$

Otherwise, we leave $\hat{F}_n$ undefined. In this setting, we will put $p_i = F_{\mathcal{M}}(M_i) - F_{\mathcal{M}}(M_{i-1}) = \text{Prob}_{\mathcal{M}}[M_i]$, with $M_0 \equiv \emptyset$ and $F_{\mathcal{M}}(\emptyset) = 0$.

The indicator notation introduced in § 1.1 can be defined in terms of maximal antichain membership

$$\alpha_{ij} = \mathbf{1}[(b_{2i-1}, b_{2i}) \in x_j] = \mathbf{1}[x_j \in M_i], \; i = 1, \ldots, m, j = 1, \ldots, n. \tag{4.1}$$

This notation is just the transposed version of the the one used by Turnbull (1976) and Gentleman & Geyer (1994).

The likelihood can be written as

$$L(F) = L(\mathbf{p}) = \prod_{j=1}^{n} \left[\sum_{i=1}^{m} \alpha_{ij} p_i \right] = (\mathbf{A}'\mathbf{p})^{\mathbf{e}_m} \tag{4.2}$$

subject to $\mathbf{0} \leq \mathbf{p} \leq \mathbf{e}$ and $\mathbf{e}'\mathbf{p} = 1$, where $\mathbf{A} = [\alpha_{ij}]$ is the Petrie matrix of the data. See A for the notational conventions, an explanation of matrix/vector exponentiation and the definition of $\mathbf{e}_m$.

Equation (4.2) can be reparametrized in terms of a discrete CDF vector on $\mathcal{M}$. Letting $\underline{\sigma} = [\sigma_1, \ldots, \sigma_m]' = \mathbf{K}\mathbf{p}$, where $\mathbf{K} = \mathbf{1}[i \geq j] \in \{0,1\}^{m \times m}$ is the cumulative sum operator matrix previously described in Lemma 1.8, we can define $\underline{\mathbf{W}} = \mathbf{K}'^{-1}\mathbf{A} \in \{-1,0,1\}^{m \times n}$ and redefine the likelihood function by

$$L(\underline{\sigma}) = (\underline{\mathbf{W}}'\,\underline{\sigma})^{\mathbf{e}_m}$$

subject to the constraint

$$0 < \sigma_1 \leq \cdots \leq \sigma_{m-1} < \sigma_m = 1. \tag{4.3}$$

The strict inequalities in this constraint are justified later by the application of Lemma 4.2 to the essential maximal antichains M_1 and M_m. The reparametrization from $\mathbf{p}$ to σ is not merely cosmetic: as we will see in § 3, the constraint that σ lie in a bounded region of isotonicity allows the application of isotonic regression techniques to determine the MLE of σ.

There is another useful form for the likelihood which is justified by the interpretation of the characteristic matrix seen in § 1.7:

$$L(\sigma) = \prod_{0 \leq i < k \leq m} (\sigma_k - \sigma_i)^{\chi_{i+1,k}} \tag{4.4}$$

subject to (4.3), where we define $\sigma_0 = 0$ and $\sigma_m = 1$.

As one would expect, universal elements are superfluous to the likelihood as they merely contribute a factor of one to it. We will assume in the sequel that no element of X is universal. In terms of the characteristic matrix, this will mean that $\chi_{1,m} = 0$.

In Turnbull (1976) and Gentleman & Geyer (1994), the authors caution against cases where (in our transposed notation) there are maximal antichains M_i and M_k such that $\alpha_{ij} = \alpha_{kj}$ for all j, as such a situation can lead to non-uniqueness of the NPMLE. However, Lemma 1.5 and Theorem 1.10 show together that this situation can never occur, since $\mathbf{A}$ must have full row-rank, or since, equivalently, no two maximal antichains can have the same membership without violating the starting/ending property described in § 1.4.

We can extend this observation to the following Theorem.

Theorem 4.1 *With* $\mathcal{M} = \{M_i; i = 1, \ldots, m\}$ *and* $p_i = \text{Prob}_{\mathcal{M}}[M_i]$, *the NPMLE of* $\mathbf{p}$ *is unique.*

Proof. As noted in Gentleman & Geyer (1994), the NPMLE $\hat{\mathbf{p}}$ is unique if $\mathbf{H} = \mathbf{A}\mathbf{D}^{-2}_{\mathbf{A}'\hat{\mathbf{p}}}\mathbf{A}'$ is positive definite. But $\mathbf{H}$ is certainly non-negative definite since $\mathbf{H} = \left(\mathbf{A}\mathbf{D}^{-1}_{\mathbf{A}'\hat{\mathbf{p}}}\right)\left(\mathbf{A}\mathbf{D}^{-1}_{\mathbf{A}'\hat{\mathbf{p}}}\right)'$. Also, $\mathbf{A}'\hat{\mathbf{p}} \neq 0$ elementwise, since otherwise the likelihood is 0. Therefore $\mathbf{D}^{-1}_{\mathbf{A}'\hat{\mathbf{p}}}$ has full rank n. Since by Lemma 1.5, $\mathbf{A}$ has full row rank $m \leq n$, $\mathbf{H}$ must have rank m and so is positive definite. □

Theorem 4.1 applies to all cases of interval censored data (for example all those listed on p. 11). This can be compared with the proof of Wellner & Zhan (1997) which is specific to doubly censored data.

Theorem 1.10 opens a number of avenues concerning the interpretation of consistency for estimators of $\mathbf{p}$ viewed from an order theoretic point of view. The dimensionality of the estimand and the mapping $H_{\mathcal{M}}$ both depend on the censoring pattern. Depending on this pattern, the number of maximal antichains may or may not increase as n increases. If it does increase, $H_{\mathcal{M}}(\mathcal{M})$ may or may not converge to a dense set in $\mathbb{R}$. Whatever conditions apply, the results of Redner (1981) can be used to show consistency of the NPMLE up to the identifiability of the underlying distribution. Asymptotic distributional results, in such a case, may depend on the mechanism which gives rise to the censoring pattern. Alternatively, by selecting a particular member of the equivalence class as the estimate and assuming so-called strong separation of interval endpoints, Groeneboom (1996) obtains explicit asymptotic results.

2 Self-consistency

Our interpretation of interval censored data in terms of their order theoretic structure shows that the estimation problem conditioned on the censoring pattern is essentially discrete.

We will first apply the language of order theory to investigate self-consistent estimates of the CDF. The product limit estimator for right-censored data, first proposed by Kaplan & Meier (1958) [see also (4.19)] was later shown by Efron (1967) to be self-consistent. Turnbull (1974,1976) then used self-consistency as the basis for an estimation algorithm, later shown in Dempster, Laird & Rubin (1977) to be a particular application of the EM algorithm.

Self-consistent estimates of the CDF exhibit two closely related properties of particular interest, on which order theory casts a new light. The first is non-uniqueness, and the second is the fact that such estimates can assign a mass of zero to some maximal antichains. We will call *effective support* of an estimate the set of maximal antichains on which it puts positive mass. The quest for a simple a priori characterization of the effective support of the NPMLE will motivate discussions in this section and the next. To date, a simple set of necessary and sufficient conditions for a maximal antichain to receive no mass at the NPMLE is still eluding us, and may in fact not exist.

We discuss the relationship of self-consistency with covers of maximal antichains and propose some bounds on self-consistent estimates originating from the maximal antichain structure of the data. We then discuss two examples involving self-consistency from the recent literature on interval censored data, and show how apparently complex problems can be greatly simplified by the use of order theoretic language.

2.1 Self-consistent estimates and maximal antichain covers

The self-consistency condition of Turnbull (1976) can be expressed compactly as follows: $\tilde{\mathbf{p}}$ is a self-consistent of $\mathbf{p}$ if and only if

$$n\tilde{\mathbf{p}} = \mathbf{D}_{\tilde{\mathbf{p}}}\mathbf{A}\left(\mathbf{A}'\tilde{\mathbf{p}}\right)^{-\mathbf{1}}, \qquad (4.5)$$

Turnbull showed that the NPMLE is self-consistent in that it satisfies (4.5); however, it is now a well recognized fact (see Groeneboom & Wellner, 1992; Gentleman & Geyer, 1994; Wellner & Zhan, 1997) that in general there exist several distinct values of $\tilde{\mathbf{p}}$ which are self-consistent but do not maximize the likelihood. In order to be the NPMLE, a self-consistent estimate must satisfy the Kuhn-Tucker conditions listed in Gentleman & Geyer (1994). We return to these conditions in § 3.1.

Starting from some initial estimate $\mathbf{p}^{(0)}$, (4.5) can be transformed into an expression of the EM algorithm where $\mathbf{p}^{(r)}$, $r = 1, 2, \ldots$, converges to a self-consistent estimate $\tilde{\mathbf{p}}$, viz.

$$n\mathbf{p}^{(r)} = \mathbf{D}_{\mathbf{p}^{(r-1)}}\mathbf{A}\left(\mathbf{A}'\mathbf{p}^{(r-1)}\right)^{-\mathbf{1}} \qquad (4.6)$$

The convergence properties of the EM algorithm are formally demonstrated in Dempster, Laird & Rubin (1977) and Wu (1983). In particular, the conditions expressed in Wu (1983) show that the EM algorithm will converge to the NPMLE if the initial seed puts mass on every maximal antichain, that is, if $\mathbf{p}^{(0)} > \mathbf{0}$. It is easy to generate self-consistent estimates with negative entries in $\tilde{\mathbf{p}}$. We will tacitly assume that only non-negative entries in any self-consistent estimate $\tilde{\mathbf{p}}$ are allowed. This condition will always be satisfied for self-consistent estimates generated by the EM algorithm seeded with $\mathbf{p}^{(0)} > \mathbf{0}$.

Though the effective support of a self-consistent estimate may possibly not include all maximal antichains, the collection of maximal antichains on which a self-consistent estimate does put mass must form a cover for X. We formalize this observation in the following Lemma.

Lemma 4.2 *If* $\tilde{\mathbf{p}}$ *is any self-consistent estimate of* $\mathbf{p}$*, then*

$$\bigcup_{i:\tilde{p}_i>0} M_i = X.$$

In particular, if $M_i \in \mathcal{M}$ *is essential, then* $\tilde{p}_i > 0$.

Proof. The elementwise version of self-consistency reads

$$\tilde{p}_i = \frac{1}{n}\sum_{j=1}^{n} \frac{\alpha_{ij}\tilde{p}_i}{\tilde{\eta}_j} \tag{4.7}$$

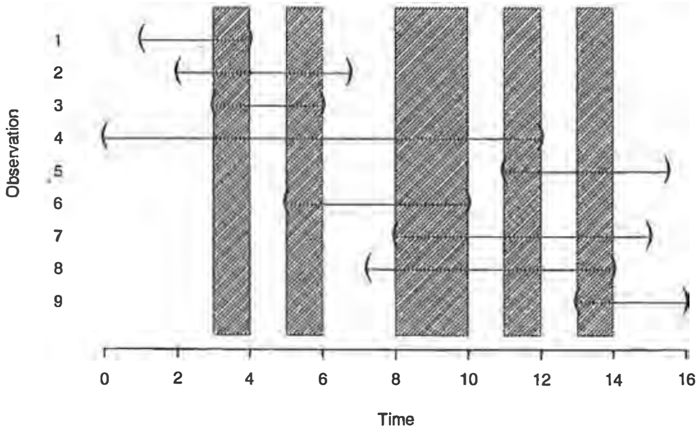
where $\tilde{\eta}_j = \sum_{i=1}^{m} \alpha_{ij}\tilde{p}_i$, for $i = 1, \ldots, m$. If $\{M_i \in \mathcal{M}; \tilde{p}_i > 0\}$ does not form a cover for X, then there is a j_0 such that $\eta_{j_0} = \sum_{i=1}^{m} \alpha_{ij_0} p_i = 0$. For all i such that $\alpha_{ij_0} = 1$, the self-consistency equation above must then fail as the left-hand side p_i is equal to zero while the right-hand side is undefined. □

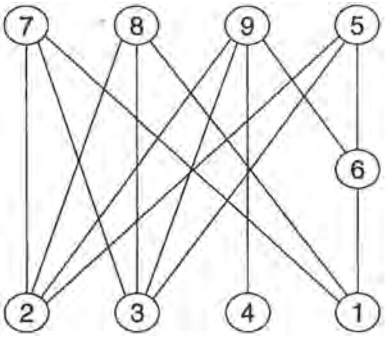
To appreciate the consequence of assigning a value of 0 to one or more of the p_i, we appeal to Lemma 1.6. Letting $p_i = 0$ for some i is equivalent to removing a non-essential maximal antichain. In a real interval representation, this could be achieved by shortening the intervals represented in R_{i-1} and L_i (the endpoint equivalence classes defined in Theorem 1.7) until they cease to overlap.

Observation 4.3 *Let* $\tilde{\mathbf{p}}$ *be a self-consistent estimate of* $\mathbf{p}$ *for interval order* $(X, \prec)$*. Let* $\mathcal{I}(\tilde{\mathbf{p}}) = \{i; 1 \le i \le m, \tilde{p}_i > 0\}$*, and let* $\prec_\circ$ *be the interval order induced by* $\mathcal{M}_\circ = \{M_i \in \mathcal{M}; i \in \mathcal{I}(\tilde{\mathbf{p}})\}$*. Then* $\tilde{\mathbf{p}}_\circ = [\tilde{p}_i]_{i\in\mathcal{I}(\tilde{\mathbf{p}})}$ *is the NPMLE of the probability function vector* $\mathbf{p}_\circ$ *on* $\mathcal{M}_\circ = MA\,(X, \prec_\circ)$.

We can thus view any self-consistent estimate consistent with the data as the NPMLE of a similar data set with added precedence relations.

Example 4.1 *Consider the interval censored data of Figure 4.1, on an arbitrary time scale.*

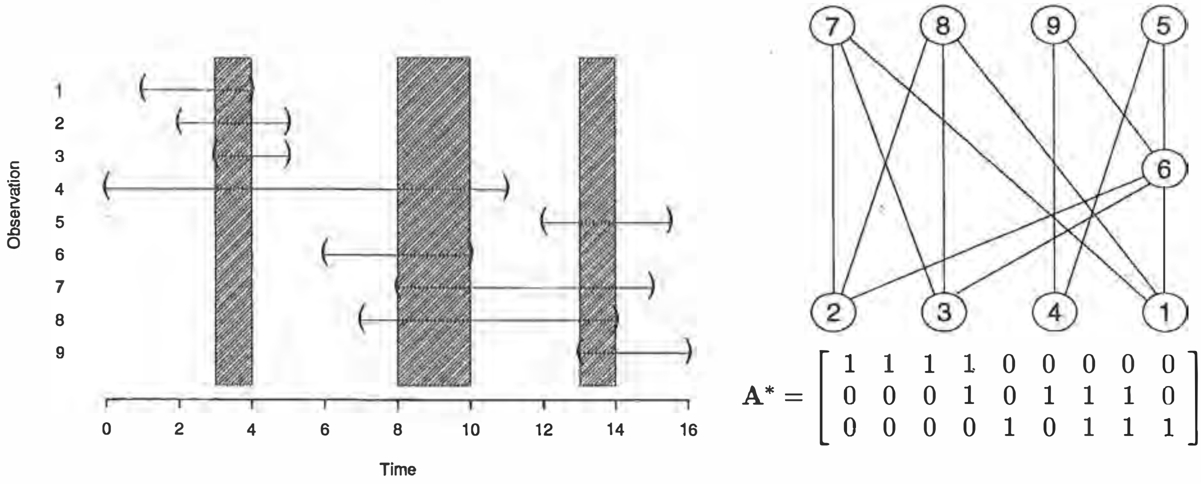




$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Figure 4.1: *Example data, interval representation of their maximal antichains, Hasse diagram and Petrie matrix.*

The NPMLE of $\mathbf{p}$ *is* $\hat{\mathbf{p}} = (3/10, 1/5, 1/10, 0, 2/5)'$, *which corresponds to maximal antichain cover* $\{M_1, M_2, M_3, M_5\}$. *Another self-consistent estimate for the data is* $\tilde{\mathbf{p}} \approx (0.4026, 0, 0.2436, 0, 0.3538)'$, *with cover* $\{M_1, M_3, M_5\}$. *We can conclude immediately that* $\tilde{\mathbf{p}}_\circ \approx (0.4026, 0.2436, 0.3538)'$ *maximizes the likelihood of the interval censored data displayed in Figure 4.2. The Hasse diagrams shows that the interval order of Figure 4.2 corresponds to that of Figure 4.1 augmented by precedence relations* $(2,6)$, $(3,6)$ *and* $(4,5)$.



$$\mathbf{A}^* = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Figure 4.2: *A similar representation for a data set with NPMLE* $\tilde{\mathbf{p}}_\circ \approx (0.4026, 0.2436, 0.3538)'$.

The other self-consistent estimate for the data of Figure 4.1 is $(5/18, 5/18, 0, 0, 4/9)'$, *corresponding to the interval order induced by the cover* $\{M_1, M_2, M_5\}$.

The case of current status data provides another illustration. For this type of data, there always exists a self-consistent estimate $\tilde{\mathbf{p}}$ with $\tilde{p}_1 = (1/n)\sum_{k=1}^{m-1} \chi_{1,k}$, $\tilde{p}_m = (1/n)\sum_{i=2}^{m} \chi_{i,m} = 1 - \tilde{p}_1$, and $\tilde{p}_i = 0$ for $i = 2, \ldots, m-1$. This self-consistent estimate is the NPMLE for $(X, \prec_\circ)$ where $x \prec_\circ y$ if and only if x is left-censored and y is right-censored. All overlaps are ignored, and mass is assigned only to the minimal cover $\{M_1, M_m\}$.

In the interval order on X induced by a minimal cover, every maximal antichain must be essential. Otherwise some maximal antichain may be removed from the set without affecting its covering property, and thus the cover could not have been minimal in the first place. Lemma 4.2 shows that every minimal cover will be the effective support of some self-consistent estimate. In other words, if $\mathcal{W}$ is a minimal cover of $\underline{X}$, then there exists a self-consistent estimate $\tilde{\mathbf{p}}(\mathcal{W})$ such that $\mathcal{W} = \{M_i \in \mathcal{M}; \tilde{p}_i(\mathcal{W}) > 0\}$. Of course, some self-consistent estimates will correspond to non-minimal covers.

A consequence of this observation is that it provides us in principle with lower and upper bounds on the number of self-consistent estimates for a particular data set, corresponding respectively to the number of minimal covers and the number of covers for the data. Though we are not aware of a practical method for determining the number of covers for an interval order, the number of minimal covers can be determined using a slight modification of Algorithms 2.4 and 2.5 presented in Chapter 2, thereby providing us with a lower bound on the number of self-consistent estimates for a particular data set. For example, the breast cosmesis data presented in Finkelstein & Wolfe (1985), with $n = 93$ and $m = 30$, correspond to an interval order with 2646 minimal covers, a lower bound on the number of self-consistent estimates for these data.

2.2 Bounds on self-consistent estimates

Although the isotonic regression approach of § 3 may be better suited to explain the tendency of the NPMLE and other self-consistent estimates to assign no mass to some maximal antichains, this phenomenon can be addressed partly from the point of view of self-consistency.

Theorem 4.5 and Corollary 4.6 below provide some simple bounds on the entries of self-consistent estimates of $\mathbf{p}$. These bounds are derived from Turnbull's self-consistency condition and from the maximal antichain structure of interval orders. The proof and constructions used to derive these results are instructive in their own way, in spite of being slightly technical.

Let $\underset{\sim}{X} = (X, \prec)$ with $X = \{x_1, \ldots, x_n\}$ be an interval order. Recall the definition of the characteristic matrix χ (Chapter 1, § 1.7). We let $\mathcal{J}_i = \{j; x_j \in M_i\}$ be the index set of $M_i \in \mathcal{M}$ and write $n_i = |M_i| = \sum_{k=1}^{i} \chi_{k,i} + \sum_{k=i+1}^{m} \chi_{i,k}$, for $i = 1, \ldots, m$. For every maximal antichain M_i, recall that χ_{ii} is the number of simplicial elements it contains. With $\mathbf{A}$ the Petrie matrix of $\underset{\sim}{X}$, we put $\tilde{\eta} = \mathbf{A}'\tilde{\mathbf{p}}$.

Denote $\beta_k^i = \mathbf{1}[M_k \in M_i^{**}]$, (see Definition 1.7) and $\beta^i = (\beta_1^i, \ldots, \beta_m^i)'$, the double dual indicator vector of M_i, $i = 1, \ldots, m$. The following are elementary facts:

- $\tilde{\eta}_j > 0$ for all $j = 1, \ldots, n$; and
- $\tilde{\eta}_j \geq \tilde{p}_i$ for all $j \in \mathcal{J}_i$.

If $\tilde{\mathbf{p}}$ is any self-consistent estimate of $\mathbf{p}$, then $\beta^{i\prime}\tilde{\mathbf{p}} \geq \tilde{\eta}_j$ for all $j \in \mathcal{J}_i$. We also assume that there are no universal elements, since they have no bearing on the estimation of $\mathbf{p}$. This corresponds to the assumption that $\chi_{1,m} = 0$. These observations along with Lemma 4.2 establish the following Lemma.

Lemma 4.4 *Let $\underline{X}$ be an interval order with $m = |\mathcal{M}|$. If $\tilde{\mathbf{p}} \in [0,1]^m$ is such that $\mathbf{e}'\tilde{\mathbf{p}} = 1$ and if $M_i \in \mathcal{M}$, then*

$$0 \leq \hat{p}_i \leq \hat{\eta}_j \leq \beta^{i\prime}\hat{\mathbf{p}} \leq 1$$

and

$$0 < \eta_j < 1$$

for all $j \in \mathcal{J}_i$, $i = 1, \ldots, m$.

Theorem 4.5 *Let $\underline{X}$ be an interval order, $M_i \in \mathcal{M}$, $i = 1, \ldots, m$, and $\tilde{\mathbf{p}}$ be a self-consistent estimate of $\mathbf{p}$. Let $n_i = |M_i|$ and let χ_{ii} denote the number of simplicial elements in M_i.*

Then

$$\frac{\chi_{ii}}{n - n_i + \chi_{ii}} \leq \tilde{p}_i \leq \frac{n_i}{n}.$$

The left inequality is strict if and only if $0 < \chi_{ii} < n_i$. The left and right hand sides are equal if and only if $n_i = \chi_{ii}$, that is, if and only if all members of M_i are simplicial.

Proof. For clarity, we omit the ˜ sign within the proof, which applies to all $i = 1, \ldots, m$.

Self-consistency equation (4.7) is equivalent to the statement that

$$n = \sum_{j \in \mathcal{J}_i} \frac{1}{\tilde{\eta}_j} \text{ or } \tilde{p}_i = 0 \tag{4.8}$$

for $i = 1, \ldots, m$.

We first show that $p_i \leq n_i/n$. Since $p_i \leq \eta_j$ for all $j \in \mathcal{J}_i$, we obtain

$$\frac{n_i}{p_i} \geq \sum_{j \in \mathcal{J}_i} \frac{1}{\eta_j} = n$$

from which the result is immediate. If $\chi_{ii} = n_i$, equality occurs since then $\eta_j = p_i$ for all $j \in \mathcal{J}_i$.

Assume now that $0 < \chi_{ii} < n_i$, so that $p_i > 0$ by Lemma 4.2.

To show that $\frac{\chi_{ii}}{n - n_i + \chi_{ii}} < p_i$, we use Lemma 4.4. Put $\mathcal{S}_i = \{j \in \mathcal{J}_i; x_j \text{ is simplicial}\} \neq \emptyset$. Then

$$\begin{aligned} n &= \sum_{j \in \mathcal{J}_i} \frac{1}{\eta_j} \\ &= \frac{\chi_{ii}}{p_i} + \sum_{j \in \mathcal{J}_i \setminus \mathcal{S}_i} \frac{1}{\eta_j} \\ &> \frac{\chi_{ii}}{p_i} + n_i - \chi_{ii} \end{aligned} \tag{4.9}$$

whence the result follows.

If $p_i = 0$, then $\chi_{ii} = 0$ by Lemma 4.2, and the result is immediate. If $\chi_{ii} = 0$, then there exists a self-consistent estimate which assigns a mass of zero to M_i, by the discussion in § 2.1. Constraining self-consistent estimates of $\mathbf{p}$ to have non-negative entries makes zero the lowest achievable bound for p_i. Lastly, we note that $n_i = \chi_{ii}$ if and only if $\chi_{ii}/(n - n_i + \chi_{ii}) = n_i/n$.

□

The lower bound on $\tilde{p}_i$ for non-essential maximal antichains (for which $\chi_{ii} = 0$) is zero and cannot be improved. By contrast, the bounds on essential maximal antichains may be tightened somewhat.

Corollary 4.6 *Under the setting of Theorem 4.5, let*

$$\mathcal{E}_i^{**} = \{k;\, M_k \text{ is essential and } M_k \notin M_i^{**}\}.$$

Suppose M_i is an essential maximal antichain. Let

$$f_i^{(0)} = \frac{\chi_{ii}}{n - n_i + \chi_{ii}}$$

and define for $r \geq 1$

$$f_i^{(r)} = \frac{\chi_{ii}\left(1 - \sum_{k \in \mathcal{E}_i^{**}} f_k^{(r-1)}\right)}{n\left(1 - \sum_{k \in \mathcal{E}_i^{**}} f_k^{(r-1)}\right) - n_i + \chi_{ii}} \tag{4.10}$$

Then $f_i^{(r)}$ converges to some value $f_i \in [0,1]$ which is a fixed point of (4.10) as $r \to +\infty$, for $i = 1, \ldots, m$; if $\tilde{\mathbf{p}}$ is any self-consistent estimate of $\mathbf{p}$, then

$$f_i \leq p_i \leq \min\left(\frac{n_i}{n}, 1 - \sum_{k \in \mathcal{E}_i^{**}} f_k\right)$$

Proof. As with the proof of Theorem 4.5, the ˜ sign is omitted in the following, which applies to $i = 1, \ldots, m$.

First note that we can tighten the bound shown in (4.9) using Lemma 4.4 at the cost of the strictness of the inequality:

$$n \geq \frac{\chi_{ii}}{p_i} + \frac{n_i - \chi_{ii}}{\beta^{i\prime}\mathbf{p}}. \tag{4.11}$$

For any $c_i \geq \beta_i'\mathbf{p}$, we will have, by (4.11),

$$n \geq \frac{\chi_{ii}}{p_i} + \frac{n_i - \chi_{ii}}{c_i}$$

whence

$$\frac{\chi_{ii}}{n - \frac{n_i - \chi_{ii}}{c_i}} \leq p_i \tag{4.12}$$

Equation (4.12) must apply with $c_i = 1 - \sum_{j \in \mathcal{E}_i^{**}} f_j^{(0)} \geq \beta_i' \mathbf{p}$.

Simple induction arguments show both that the $f_i^{(r)}$ are all bounded above by the corresponding p_i, $i = 1, \ldots, m$, and are non-decreasing in r; hence it must be that $f_i^{(r)} \dashrightarrow f_i$ for some f_i which still bounds p_i below. □

In the case of current status data, Corollary 4.6 provides an interesting analysis of self-consistent estimates. Simple calculations show that if $\chi_{11}\chi_{mm} \geq (n_1 - \chi_{11})(n_m - \chi_{mm})$, then $f_1 = n_1/n$ and $f_m = n_m/n$. In this case, the only self-consistent estimate, and therefore the NPMLE, is given by $\tilde{\mathbf{p}} = (n_1/n, 0, \ldots, 0, n_m/n)$. If $\chi_{11}\chi_{mm} < (n_1 - \chi_{11})(n_m - \chi_{mm})$, then $f_1 = \chi_{11}/(n_m - \chi_{mm} + \chi_{11})$ and $f_m = \chi_{mm}/(n_1 - \chi_{11} + \chi_{mm})$.

In the case of right-censored data, long-winded but straightforward calculations show that upper and lower bounds are equal for every maximal antichain, and correspond to the expression given in (4.19). This is to be expected, as right-censored data only supports one self-consistent estimate, but shows that the bounds may be tight in some cases. Whether they are tight in general, i.e. whether there always exists a self-consistent estimate with some entry equal to its lower or upper bound in Corollary 4.6, is an open problem.

Theorems 4.5 and Corollary 4.6 are interesting practically in that they can provide good initial estimates in iterative estimation procedures to determine the NPMLE, and theoretically as they flesh out the basic result of Lemma 4.2 by appealing strictly to the maximal antichain structure of the underlying interval order. Their application is limited first in that they provide bounds on all self-consistent estimates, and secondly in that they take an unweighted view of the maximal antichain double duals. For these reasons, no refinement of the lower and upper bounds of Corollary 4.6 can be expected to converge to a single value.

2.3 Two examples

We consider two examples taken from the recent literature on interval censored data to motivate the use of our language.

Our first example concerns the results of Mykland & Ren (1996) on doubly censored data. In Theorem 2, they characterize the NPMLE of the survival function (equivalently, of the CDF) as a self-consistent estimator satisfying Kuhn-Tucker-type conditions. In Corollary 3, they provide an equality which must be satisfied at a particular self-consistent estimate of the survival function. The EM algorithm is used to determine the estimate satisfying this equation. This estimate has jumps precisely at the exact observations, apart possibly from the right endpoint of the first occurrence of left censoring and the left endpoint of the last occurrence of right censoring. They propose a condition under which this self-consistent estimator is also the NPMLE in Corollary 4 and indicate where jumps might occur if it is not in Corollary 5.

In the language of order theory, their Theorem 2 is equivalent to the Kuhn-Tucker conditions satisfied at a self-consistent estimate proposed by Gentleman & Geyer (1994), to which we will return in § 3.1. The version they propose is the particular form these conditions take in the cases of doubly censored and current status data. Corollary 3 states that there exists a self-consistent estimate which puts mass only at essential maximal antichains, which we know to be true since the essential maximal antichains form a cover for the data in the doubly censored case. Corollary 4 indicates that this self-consistent estimator is the NPMLE whenever all maximal antichains in the data are essential, which must be true since every maximal antichain then has positive mass by Theorem 4.5; Corollary 5 identifies non-essential maximal antichains as being the only other locations on which the NPMLE can put mass if their estimate does not maximize the likelihood. These results are more easily expressed and interpreted within an order theoretic framework than in the rather elaborate notation usually employed to distinguish between various types of censoring.

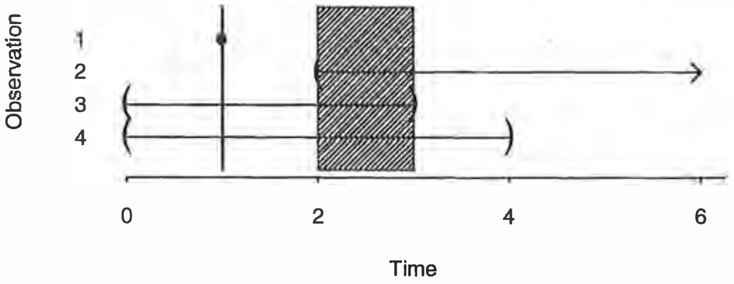
A statement by Mykland & Ren (1996) to the effect that several authors have shown "the strong consistency and weak convergence of any self-consistent estimator of [the survival function]", is not quite correct. Consistency as shown in Gu & Zhang (1993), for instance, requires a regularity condition to be satisfied. This condition is that if q is a random interval endpoint or exact observation, then

$$\text{Prob}[q \text{ is an exact observation}|q = t] > 0 \text{ for all } t > 0. \tag{4.13}$$

Current status data, a special case of doubly censored data, are patently excluded by this condition.

Given that there always exists a self-consistent estimate for doubly censored data with $p_1 = 1 - p_m$, we should not expect consistency to be provable in this case. Other obvious cases of doubly censored data may be excluded by this condition, such as those with mixed continuous and point inspection processes along the time line. By contrast, consistency of the NPMLE is well-documented for the general case of interval censored data. Since determining the NPMLE requires no more work than evaluating the self-consistent estimator of Mykland and Ren, the NPMLE should remain the estimator of choice unless or until a better estimator is proposed.

Secondly, we discuss a simple example used by Wellner & Zhan (1997) to show that the EM algorithm does not necessarily converge to the NPMLE, which leads them to deem it "ambiguous". In fact the EM algorithm will always converge to the NPMLE if its initial seed puts mass on all maximal antichains. Their Example 5.1 of doubly censored data is composed of the exact data and intervals 1, $(2, +\infty)$, $(0, 3)$ and $(0, 4)$.



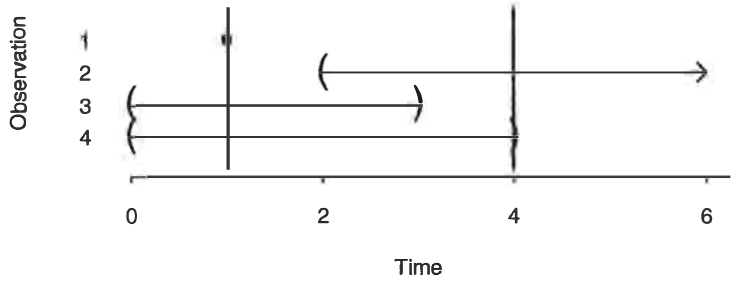
$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

Figure 4.3: *Data and Petrie matrix for Wellner and Zhan's Example 5.1*

The data contain two universal points, which can essentially be discarded, except inasmuch as they determine the mapping $H_{\mathcal{M}}$ of Theorem 1.10. We can immediately deduce from the Petrie matrix that the NPMLE vector is $(1/2, 1/2)$. Any $\mathbf{p}^{(0)} > 0$ used to seed the EM algorithm (4.6) will converge to the NPMLE.

Wellner & Zhan (1997), however, use the estimator of Groeneboom & Wellner (1992), which models the CDF as having a possible jump at every interval endpoint, namely at $t = 1, 2, 3$ and 4. They suggest two different seeds to estimate the CDF using the EM algorithm: $\sigma_1^{(0)} = (0.1, 0.1, 0.1, 0.2)$ and $\sigma_2^{(0)} = (0.1, 0.1, 0.15, 0.2)$. While the latter leads the EM algorithm to the NPMLE $(1/2, 1/2, 1, 1)$, the former causes it to converge to $(2/3, 2/3, 2/3, 1)$.

The wrong convergence value is easy to understand when considered from the point of view of maximal antichains. By placing mass only at $t = 1$ and $t = 4$, the seed reinvents, as it were, the combinatorial properties of the data. Figure 4.3 shows clearly that intervals $(0, 3)$ and $(0, 4)$ are in fact equivalent in terms of their duals, and should therefore be assigned the same mass. Since the EM algorithm cannot create mass where none is assigned, $\sigma_1^{(0)}$ imposes on the data the structure depicted in Figure 4.4.

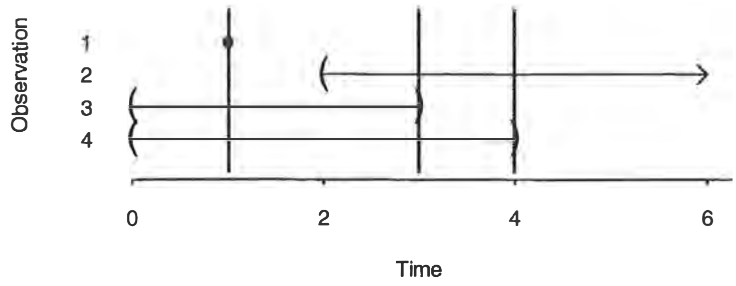


$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Figure 4.4: *Wrong maximal antichain structure imposed by* $\sigma_1^{(0)}$ *on the data and corresponding Petrie matrix.*

Again the nonparametric likelihood estimate $(2/3, 1/3)$ can be read off directly from the Petrie matrix. The estimate is correct; the data are wrong. In fact $(2/3, 1/3)$ is not even a self-consistent estimate for the original data set.

The second seed $\sigma_2^{(0)}$ is also instructive, in that it places mass at $t = 1$, $t = 3$ and $t = 4$. Both maximal antichains of the original data are here given mass (at their right endpoint), but extra mass is assigned at $t = 4$.



$$\mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Figure 4.5: *Wrong maximal antichain structure imposed by* $\sigma_2^{(0)}$ *on the data and corresponding Petrie matrix.*

In this case, the mass at $t = 4$ corresponds to an antichain which is not maximal, a fact made obvious by the row rank deficiency of $\mathbf{A}_2$. The EM algorithm ignores the extra antichain, since mass must be assigned in such a way that both $(0, 3)$ and $(0, 4)$ receive the same probability.

This example illustrates the importance of assigning mass to endpoint equivalence classes, as defined in Theorem 1.7, rather than to the endpoints themselves.

3 Isotonic regression

In this section, we discuss isotonic regression methods for estimating CDF under interval censoring proposed by Groeneboom (1991) and Groeneboom & Wellner (1992). We will first consider the Fenchel duality conditions of Groeneboom & Wellner (1992). These conditions characterize the NPMLE and guarantee its uniqueness, but the artificial requirement of allowing the estimate to have jumps only at interval endpoints may have masked their generality. We will show these conditions to be equivalent on $\mathcal{M}$ to self-consistency augmented by the Kuhn-Tucker conditions as proposed by Gentleman & Geyer (1994). Our second concern is to reformulate isotonic regression methods to make them applicable on the set of maximal antichains.

3.1 Fenchel duality and Kuhn-Tucker conditions

Recalling that $m = |\mathcal{M}|$, we define the matrix $\mathbf{K}_m$ as in Lemma 1.8 by $\mathbf{K}_m = [\mathbf{1}(i \geq j)]$, $i, j = 1, \ldots, m$. We recall that $\mathbf{K}_m$ is the cumulative sum transformation, $\mathbf{K}_m'$ is the reverse cumulative sum transformation, and $\mathbf{K}_m^{-1} = \mathbf{1}[i = j] - \mathbf{1}[i - 1 = j]$ is the backward difference transformation. We can then denote the CDF on $\mathcal{M}$ by $\underline{\sigma} = \left[\begin{array}{c} \sigma \\ \hline 1 \end{array}\right] = \mathbf{K}_m \mathbf{p}$ under the constraint $\mathbf{e}'\mathbf{p} = 1$. It follows that $\sigma \in (0, 1)^{m-1}$. We also set $\underline{\mathbf{W}} = \mathbf{K}_m'^{-1} \mathbf{A}$. Using the notational convention of (4.5), we can describe a log-likelihood partial derivative vector evaluated at σ by

$$\underline{\mathbf{w}} = \underline{\mathbf{w}}(\sigma) = \underline{\mathbf{W}} \left(\underline{\mathbf{W}}'\underline{\sigma}\right)^{-1} = \begin{bmatrix} \mathbf{w}_\sigma \\ a \end{bmatrix},$$

where the value of a at a self-consistent estimate will be made explicit below. Groeneboom and Wellner's Fenchel duality conditions on the existence and uniqueness of the CDF NPMLE can be translated on the discrete space $\mathcal{M}$ as follows.

Theorem 4.7 (Groeneboom & Wellner, 1992, Proposition II.1.3) *If* $\underline{\sigma} = (\sigma_1, \ldots, \sigma_{m-1}, 1)$ *is the CDF on* $\mathcal{M}$ *for some interval order* $\underline{X} = (X, \prec)$, *then* $\underline{\hat{\sigma}}$ *maximizes the likelihood* $\left(\underline{\mathbf{W}}'\underline{\sigma}\right)^{\mathbf{e}}$ *over* $0 < \sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_{m-1} < 1$ *if and only if*

$$\mathbf{K}'_{m-1}\mathbf{w}_{\hat{\sigma}} \leq \mathbf{0} \textit{ elementwise} \tag{4.14}$$

and

$$\hat{\sigma}'\mathbf{w}_{\hat{\sigma}} = 0 \tag{4.15}$$

One of the Kuhn-Tucker conditions (Gentleman & Geyer, 1994) corresponds directly to (4.15) the other to the requirement that the Lagrange multipliers $n\mathbf{e} - \mathbf{A}\left(\mathbf{A}'\tilde{\mathbf{p}}\right)^{-\mathbf{I}}$ be non-negative, that is

$$\begin{aligned}
\mathbf{A}\left(\mathbf{A}'\tilde{\mathbf{p}}\right)^{-\mathbf{I}} &\leq n\mathbf{e} \text{ elementwise, so that} \\
\mathbf{K}'_m\underline{\mathbf{W}}\left(\underline{\mathbf{W}}'\underline{\tilde{\sigma}}\right)^{-\mathbf{I}} &\leq n\mathbf{e}, \text{ i.e.} \\
\left[\begin{array}{c|c} \mathbf{K}'_{m-1} & \mathbf{e}_{m-1} \\ \hline \mathbf{0}'_{m-1} & 1 \end{array}\right]\left[\begin{array}{c} \mathbf{w}_{\hat{\sigma}} \\ \hline n \end{array}\right] &\leq n\mathbf{e}.
\end{aligned}$$

This expression is satisfied at an estimate if and only if the first of the Fenchel duality conditions (4.14) is satisfied.

Since it is well-known that there are self-consistent estimates that are not NPMLEs, we next examine why self-consistency alone is not sufficient to guarantee maximization of the likelihood. Premultiplying self-consistency equation (4.5) by $\mathbf{K}_m$ and expressing it in terms of $\underline{\sigma}$ yields

$$n\underline{\sigma} = \mathbf{K}_m\mathbf{D}_{\mathbf{K}_m^{-1}\underline{\sigma}}\mathbf{K}'_m\underline{\mathbf{W}}\left(\underline{\mathbf{W}}'\underline{\sigma}\right)^{-\mathbf{I}} = \mathbf{K}_m\mathbf{D}_{\mathbf{K}_m^{-1}\underline{\sigma}}\mathbf{K}'_m\underline{\mathbf{w}}. \tag{4.16}$$

The structure of the matrix premultiplying $\underline{\mathbf{w}}$ is as follows:

$$\mathbf{K}_m\mathbf{D}_{\mathbf{K}_m^{-1}\sigma}\mathbf{K}'_m = \begin{bmatrix}
\sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 & \cdots & \sigma_1 & \sigma_1 \\
\sigma_1 & \sigma_2 & \sigma_2 & \sigma_2 & \cdots & \sigma_2 & \sigma_2 \\
\sigma_1 & \sigma_2 & \sigma_3 & \sigma_3 & \cdots & \sigma_3 & \sigma_3 \\
\sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 & \cdots & \sigma_4 & \sigma_4 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 & \cdots & \sigma_{m-1} & \sigma_{m-1} \\
\sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 & \cdots & \sigma_{m-1} & 1
\end{bmatrix}.$$

Equating the last two elements of the vectors in (4.16), we obtain

$$\begin{aligned} n &= \left[\begin{array}{c|c} \tilde{\sigma}' & 1 \end{array}\right] \left[\begin{array}{c} \mathbf{w}_{\tilde{\sigma}} \\ \hline a \end{array}\right] \text{ and} \\ n\tilde{\sigma}_{m-1} &= \left[\begin{array}{c|c} \tilde{\sigma}' & \tilde{\sigma}_{m-1} \end{array}\right] \left[\begin{array}{c} \mathbf{w}_{\tilde{\sigma}} \\ \hline a \end{array}\right] \end{aligned}$$

for $\tilde{\sigma}$ a self-consistent estimate of σ. Hence $a = n$ at a self-consistent estimate of σ, whence (4.16) reduces to

$$\mathbf{K}_{m-1}\mathbf{D}_{\mathbf{K}_{m-1}^{-1}\tilde{\sigma}}\mathbf{K}'_{m-1}\mathbf{w}_{\tilde{\sigma}} = \mathbf{0}. \tag{4.17}$$

We are now in a position to compare self-consistency to the Fenchel duality conditions. Considering the first and last rows of the matrix premultiplying $\mathbf{w}_{\tilde{\sigma}}$ in (4.17), we get that, at self-consistent estimate $\tilde{\sigma}$,

$$\mathbf{e}'\mathbf{w}_{\tilde{\sigma}} = 0 \text{ and } \tilde{\sigma}'\mathbf{w}_{\tilde{\sigma}} = 0.$$

Although (4.17) forces a complex constraint to be respected by an estimate for it to be self-consistent, the crucial missing condition for a self-consistent estimate to uniquely maximize the likelihood is therefore (4.14).

3.2 Isotonic regression on $\mathcal{M}$

Given that we can apply Fenchel duality conditions on an estimate of the CDF on $\mathcal{M}$, the use of isotonic regression methods can be justified in a manner which parallels the discussion in Groeneboom & Wellner (1992). Our goal here will be to reexpress the convex minorant approach introduced by them in terms of a maximum likelihood estimation technique on $\mathcal{M}$. This reinterpretation, in our view, grounds isotonic regression techniques on more intuitive and more easily applicable bases. It also applies a preliminary data reduction which improves the performance of the algorithms involved.

We use likelihood equation (4.4) subject to the isotonicity constraint (4.3). The form of the likelihood will guarantee that $\hat{\sigma}_b > \hat{\sigma}_{a-1}$ whenever $\chi_{ab} > 0$. In particular, it will cause all estimates to lie strictly between 0 and 1, since the factors $(\sigma_i - \sigma_{k_1})$ and $(\sigma_{k_2} - \sigma_i)$ must appear in the likelihood for all i and some k_1, k_2, by the starting/ending property mentioned in § 1.4. In general, the ordering thus induced is only partial and the isotonic constraint (4.3) remains necessary unless all maximal antichains are essential.

In this section, we deal with the more convenient log-likelihood function

$$\log L(\sigma) = \sum_{k=1}^{m-1} \left[\chi_{1k} \log \sigma_j + \chi_{k+1,m} \log(1-\sigma_k) + \sum_{i=1}^{k-1} \chi_{i+1,k} \log(\sigma_k - \sigma_i) \right], \tag{4.18}$$

which we must maximize subject to (4.3).

The main idea now lies in isotonic regression, by which the estimate is made to satisfy the constraint (4.3). References on isotonic regression are Barlow et al. (1972) and, more recently, Robertson et al. (1988).

The case of right-censored data

Our first case of interest is right-censored data. Every maximal antichain is essential in such data; Lemma 4.2 indicates that the unconstrained NPMLE of the CDF will be strictly increasing over $\mathcal{M}$ since all $\tilde{p}_i > 0$, thus satisfying (4.3). We can express the nonparametric maximum likelihood estimate in terms of the entries of the characteristic matrix.

$$\hat{\sigma}_i = \frac{\chi_{ii} + \left(\sum_{j=i+1}^{m} \chi_{jj} + \sum_{j=i+1}^{m-1} \chi_{jm} \right) \hat{\sigma}_{i-1}}{\sum_{j=i}^{m} \chi_{jj} + \sum_{j=i+1}^{m-1} \chi_{jm}} \quad \text{for } i = 1, \ldots, m-1 \tag{4.19}$$

putting $\sigma_0 = 0$.

We derive explicitly (4.19) from first principles, but starting with the interval order structure of the data. Right-censored data consists only of simplicial elements ($\chi_{ii} > 0$ for $i = 1, \ldots, m$) and right-censored elements ($\chi_{ij} > 0$ for $i < j$ only if $j = m$). Bearing in mind that the unconstrained NPMLE of σ will already be isotonic, we obtain a reduction of the log-likelihood (4.18) to

$$\log L(\sigma) = \chi_{11} \log(\sigma_1) + \chi_{mm} \log(1-\sigma_{m-1}) + \sum_{i=2}^{m-1} \chi_{ii} \log(\sigma_i - \sigma_{i-1}) + \sum_{i=1}^{m-2} \chi_{i+1,m} \log(1-\sigma_i) \tag{4.20}$$

Taking the derivative of each σ_k, $k = 1, \ldots, m-1$ in turn yields estimating equations

$$\chi_{11}(1-\hat{\sigma}_1)(\hat{\sigma}_2 - \hat{\sigma}_1) + \chi_{22}\hat{\sigma}_1(1-\hat{\sigma}_1) + \chi_{2m}\hat{\sigma}_1(\hat{\sigma}_2 - \hat{\sigma}_1) = 0 \tag{4.21}$$

$$\begin{aligned} \chi_{kk}(1-\hat{\sigma}_k)(\hat{\sigma}_{k+1} - \hat{\sigma}_k) + \chi_{k+1,k+1}(\hat{\sigma}_k - \hat{\sigma}_{k-1})(1-\hat{\sigma}_k) & \\ + \chi_{k+1,m}(\hat{\sigma}_k - \hat{\sigma}_{k-1})(\hat{\sigma}_{k+1} - \hat{\sigma}_k) &= 0 \end{aligned} \tag{4.22}$$

$$\chi_{m-1,m-1} - (\chi_{m-1,m-1} + \chi_{m,m})\hat{\sigma}_{m-1} + \chi_{m,m}\hat{\sigma}_{m-2} = 0 \tag{4.23}$$

Claim 1 *For $k = 1, \ldots, m-1$, with $\hat{\sigma}_0 \overset{def}{=} 0$, (4.19) provides the NPMLE of σ.*

Proof. We first proceed by induction on $k = m-1, m-2, \ldots, 2$. First note that (4.19) is a reexpression of (4.23), showing the expression is true for $k = m-1$.

Assume then that the expression is true with k replaced by $k+1$, $2 \leq k \leq m-2$. Use (4.22) and replace $\hat{\sigma}_{k+1}$ by the appropriately modified expression given by induction hypothesis (4.19). Multiply through by $\sum_{j=k+1}^{m} \chi_{jj} + \sum_{j=k+2}^{m-1} \chi_{jm}$, and factor out $(1-\hat{\sigma}_k)\,\chi_{k+1,k+1}$; then solve the remaining linear equation for $\hat{\sigma}_k$ in terms of $\hat{\sigma}_{k-1}$ to obtain the induction hypothesis for k.

To show the induction hypothesis holds for $k = 1$, use (4.19) to substitute for $\hat{\sigma}_2$ in (4.21), multiply through by $\sum_{j=2}^{m} \chi_{jj} + \sum_{j=3}^{m-1} \chi_{jm}$, simplify and factor out $(1-\hat{\sigma}_1)$. The remaining linear equation is (4.19) with $k = 1$. □

These estimators for σ_k correspond to the Kaplan-Meier estimators, as becomes clear if we rewrite (4.19) as

$$\hat{\sigma}_k = \frac{\chi_{kk} - \chi_{kk}\hat{\sigma}_{k-1} + \left(\sum_{j=k}^{m} \chi_{jj} + \sum_{j=k+1}^{m-1} \chi_{jm}\right)\hat{\sigma}_{k-1}}{\sum_{j=k}^{m} \chi_{jj} + \sum_{j=k+1}^{m-1} \chi_{jm}} \tag{4.24}$$

$$= \hat{\sigma}_{k-1} + \frac{\chi_{kk}\,(1-\hat{\sigma}_{k-1})}{\sum_{j=k}^{m} \chi_{jj} + \sum_{j=k+1}^{m-1} \chi_{jm}} \tag{4.25}$$

$$= 1 - (1-\hat{\sigma}_{k-1})\left(1 - \frac{\chi_{kk}}{\sum_{j=k}^{m} \chi_{jj} + \sum_{j=k+1}^{m-1} \chi_{jm}}\right), \tag{4.26}$$

or one less the product limit NPMLE for survival function $S(t) = 1 - F(t)$ introduced by Kaplan & Meier (1958).

The right-censored data case has been somewhat of a misguiding tyrant since the product-limit estimator (4.24) was first derived by Kaplan & Meier (1958). The first cause of confusion is the fact that all maximal antichains in right-censored data are essential. Consequently, $\mathcal{M}$ is the only minimal cover and induces the only self-consistent estimator, which must of course also be the NPMLE. Since Efron (1967) showed the equivalence of self-consistency and nonparametric likelihood maximization for right-censored data, self-consistency has been considered an attractive feature in an estimator, stifling research in non-self-consistent or asymptotically self-consistent estimators. We have seen in § 2.3 that self-consistent estimators can be badly behaved. In particular, self-consistent estimators have been shown [see (4.13)] to be consistent only by positing a positive probability for an observed endpoint to be an exact observation everywhere on the support of F, a stringent requirement which cannot be weakened. A second cause of confusion arises because $H_{\mathcal{M}}$ maps $\mathcal{M}$ to points, rather than intervals on the real line. In order to mimic the consequent jump behaviour of the CDF estimates, statisticians have placed strong requirements on the placement of mass by their estimates and have made the tacit assumption that the only acceptable CDF estimates are uniquely defined step functions. We hope that by separating estimation on $\mathcal{M}$ from the mapping to the real line $H_{\mathcal{M}}$, we have made it clear that right-censored data have an exceptional structure which cannot be expected to apply to the general case of interval censored data.

The case of current status data

Our second example is current status data. In this case, only terms of the form σ_i or $1-\sigma_i$ occur in the log-likelihood, so that (4.18) can be rewritten as

$$\log L(\sigma) = \sum_{i=1}^{m-1} [\chi_{1,j} \log \sigma_i + \chi_{i+1,m} \log(1-\sigma_i)] \tag{4.27}$$

Because each term of the log-likelihood only involves one entry of σ, we can solve for unconstrained $\hat{\sigma}_i$ and then isotonize the estimates, in a manner similar to the one illustrated in Theorem 1.10 of Barlow et al. (1972) or Theorem 1.5.1 of Robertson et al. (1988).

Setting the partial derivative of the log-likelihood with respect to σ_k equal to zero for each $k = 1, \ldots, m-1$, yields the following (non-isotonized) estimates:

$$\hat{\sigma}_k^{\text{ni}} = \frac{\chi_{1,k}}{\chi_{1,k} + \chi_{k+1,m}} \tag{4.28}$$

each associated with a weight of $\chi_{1,k} + \chi_{k+1,m}$ representing the number of elements of X in maximal antichain k.

Before isotonizing this estimate, it is instructive to couch the derivation of the NPMLE for current status data found in Groeneboom & Wellner (1992) in terms of the maximal antichain structure of the data, or, in effect, to consider the translation process of the problem from the real line to $\mathcal{M}$. Their data model is of the form $(x_j, \delta_j) \in \mathbb{R}^+ \times \{0,1\}$, $j = 1, \ldots, n$, each observed pair being associated with an unobserved event time t_j, with $\delta_j = \mathbf{1}[t_j < x_j]$. Let $(x_{(j)}, \delta_{(j)})$ denote the data set ordered according to the linear ordering on the x_j's. Then Proposition 1.2 of Groeneboom & Wellner (1992) characterizes the NPMLE of the event time CDF at $x_{(k)}$ as being the left derivative of the greatest convex minorant of the cumulative sum diagram (CSD),

$$\left(k, \sum_{j\leq k} \delta_{(j)}\right) \text{ at } k. \tag{4.29}$$

The greatest convex minorant of a CSD is the graph of the supremum of all convex functions which do not exceed the CSD (Barlow et al., 1972, Chapter 1). Example 4.2 illustrates the concept.

The nature of current status data is that the dual of each element of X over $\mathcal{M}$ extends either to M_1 or to M_m, as previously noted. The endpoint equivalence classes of Theorem 1.7 are thus structured not only according to the maximal antichains of X, but also according to the censoring pattern of the elements. Specifically, all elements belong either to L_1 or R_m. In the CSD given by (4.29), right-endpoint equivalence class R_k corresponds to a vertical rise of χ_{1k} over an equal number of horizontal units for $k = 1, \ldots, m-1$; left-endpoint equivalence class L_k corresponds to no rise over $\chi_{k,m}$ units for $k = 2, \ldots, m$. Thus (4.29) can serve as a representation of the conjoint weak order of Theorem 1.7, where the precedence relation between any pair of successive classes is indicated by a bend or elbow in the curve.

An equivalent representation of this cumulative sum diagram is

$$\left(\sum_{i=1}^{k} [\chi_{1,k} + \chi_{k+1,m}], \sum_{i=1}^{k} \chi_{1,k}\right). \tag{4.30}$$

By "equivalent", we mean that the greatest convex minorant of (4.30) must be equal to that of (4.29).

A basic result of isotonic regression theory is that the left derivative of the greatest convex minorant of cumulative sum diagram $\left(\sum_{i=1}^{k} w_i, \sum_{i=1}^{k} w_i g_i\right)$ is the regression of g_i on i with weights w_i, $i = 1, \ldots, m$, subject to $g_1 \leq g_2 \leq \cdots \leq g_m$, otherwise known as the isotonic regression of g_i. The cumulative sum diagram given by (4.30) corresponds to $g_i = \chi_{1,i} / (\chi_{1,i} + \chi_{i+1,m})$ with weights $w_i = \chi_{1,i} + \chi_{i+1,m}$. Thus the NPMLE of σ is just the isotonic regression of the unconstrained likelihood estimator given by (4.28).

Example 4.2 *Consider the example of current status data shown in Figure 4.6.*

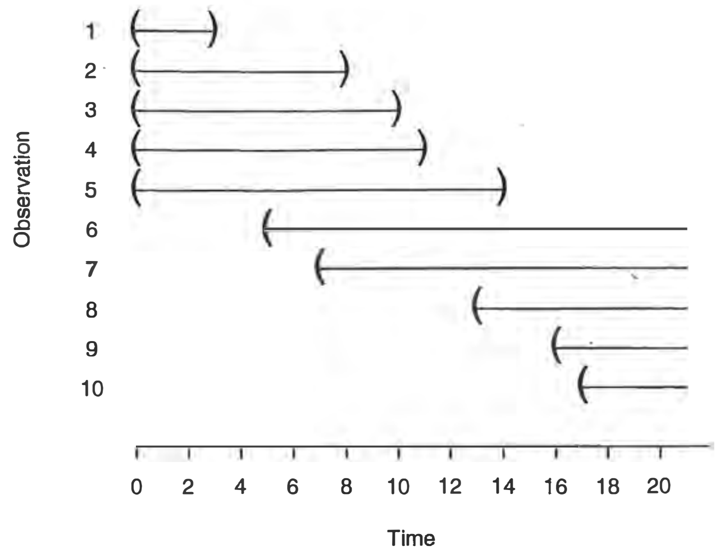


Figure 4.6: *Real representation of current status data.*

The corresponding interval order $\underline{X} = (X, \prec)$ *is given by Petrie and characteristic matrices*

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} ; \quad \chi = \begin{bmatrix} 1 & 3 & 1 & 0 \\ & 0 & 0 & 2 \\ & & 0 & 1 \\ & & & 1 \end{bmatrix} .$$

The NPMLE of σ*, constrained to lie between 0 and 1 but not constrained to be isotonic is given by*

$$\begin{aligned} \hat{\sigma}_1^{\text{ni}} &= \chi_{1,1}/(\chi_{1,1} + \chi_{2,4}) &= 1/(1+2) &= 1/3 \\ \hat{\sigma}_2^{\text{ni}} &= \chi_{1,2}/(\chi_{1,2} + \chi_{3,4}) &= 3/(3+1) &= 3/4 \\ \hat{\sigma}_3^{\text{ni}} &= \chi_{1,3}/(\chi_{1,3} + \chi_{4,4}) &= 1/(1+2) &= 1/3 \end{aligned}$$

The cumulative sum diagrams (4.29) and (4.30), in this case, have representations given in Figure 4.7.

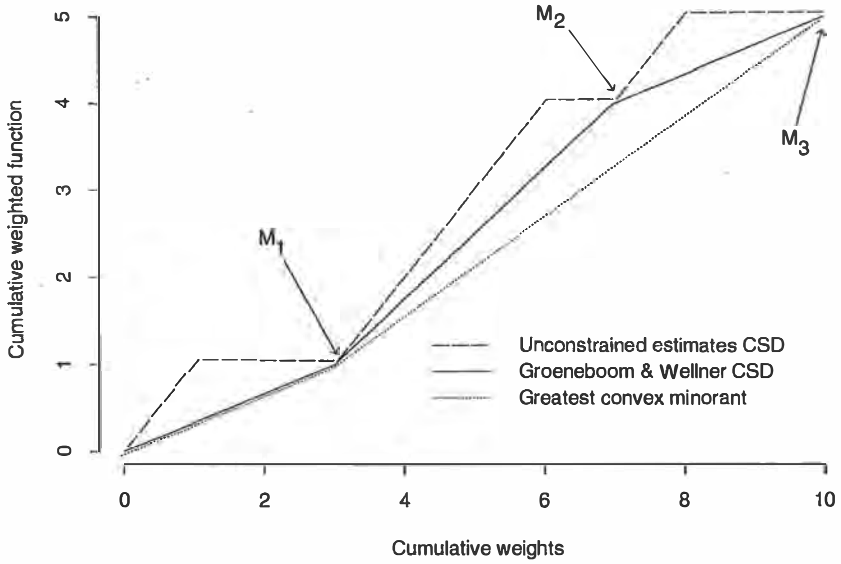


Figure 4.7: *Cumulative sum diagram and greatest convex minorant of example.*

The isotonic regression of $\hat{\sigma}_k^{\text{ni}}$ *on* k *with weights* $w_1 = 3$, $w_2 = 4$ *and* $w_3 = 3$ *can be read off as the left derivative of the greatest convex minorant at the appropriate cumulative weight values. Alternatively, the Pool-Adjacent Violator Algorithm or any other algorithm used to perform isotonic regression (see, e.g., Barlow et al., 1972, Chapter 2) can be used to find the isotonic regression of* $\hat{\sigma}_k^{\text{ni}}$. *In this example, isotonization will be achieved by letting*

$$\begin{aligned} \hat{\sigma}_1 &= \hat{\sigma}_1^{\text{ni}} = && 1/3 \\ \hat{\sigma}_2 &= \hat{\sigma}_3 = && \frac{\chi_{1,2} + \chi_{1,3}}{\chi_{1,2} + \chi_{3,4} + \chi_{1,3} + \chi_{4,4}} = 4/7 \end{aligned}$$

which yields probability mass estimates on the maximal antichains of $\hat{p}_1 = 1/3$, $\hat{p}_2 = 5/21$, $\hat{p}_3 = 0$ *and* $\hat{p}_4 = 3/7$.

The general case: the iterative convex minorant algorithm

For the general case of interval censored data, isotonic regression is used in the same spirit as for current status data. A complication arises, however, in that the terms of the log-likelihood generally involve more than one entry from the vector σ. In contrast to the case of current status data, the second derivative of $\log L(\sigma)$ is not a diagonal matrix. However, consider the general expression for the partial derivative of (4.18) with respect to σ_i, $i = 1, \ldots, m-1$:

$$\frac{\partial}{\partial \sigma_i} \log L(\sigma) = \frac{\chi_{1,i}}{\sigma_i} - \frac{\chi_{i+1,m}}{1-\sigma_i} + \sum_{k=1}^{i-1} \frac{\chi_{k+1,i}}{\sigma_i - \sigma_k} - \sum_{k=i+1}^{m} \frac{\chi_{i+1,k}}{\sigma_k - \sigma_i} = w_i(\sigma). \tag{4.31}$$

Now assume, within an iterative process, that all σ_k are fixed for $k \neq i$, and that we wish to solve for σ_i in $w_i(\sigma) = 0$, without regards to the isotonic constraint. Positing some starting value σ_i^{old} for σ_i, we could use Fisher scoring to obtain an approximate solution $\sigma_i^{(\text{new}),\text{ni}}$ via the equation

$$\sigma_i^{(\text{new}),\text{ni}} = \sigma_i^{(\text{old})} - \left(\frac{\partial^2}{\partial \sigma_i^2} \log L(\sigma)\right)^{-1} \Bigg|_{\sigma_i = \sigma_i^{(\text{old})}} \left(\frac{\partial}{\partial \sigma_i} \log L(\sigma)\right) \Bigg|_{\sigma_i = \sigma_i^{(\text{old})}}$$

The "ni" superscript above emphasizes the fact that this second-order approximation based on a fixed $\sigma^{(\text{old})}$ is not isotonized.

We can isotonize this approximation, realizing that an approximate weight for $\sigma_i^{(\text{new}),\text{ni}}$ will be the information value at $\sigma_i^{(\text{old})}$, still assuming all other $\sigma_k^{(\text{old})}$ to be fixed.

Putting

$$\underline{\mathbf{g}}(\sigma) = \underline{\mathbf{W}}^{2\mathbf{I}} \left(\underline{\mathbf{W}'}\, \underline{\sigma}\right)^{-2\mathbf{I}} = \left[\begin{array}{c} \mathbf{g}(\sigma) \\ \hline a \end{array}\right]$$

for some a, we can reexpress these weights as a vector, replacing the "old" and "new" superscripts by an iteration superscript (r):

$$\frac{\partial^2}{\partial \sigma_i^2} \log L(\sigma) \Bigg|_{\sigma = \sigma^{(r)}} = \mathbf{g}(\sigma^{(\mathbf{r})})$$

Determining an isotonic $\sigma^{(r+1)}$ from the approximation $\sigma^{(r+1),\text{ni}}$ can be done by finding the slope of the greatest convex minorant of the CSD

$$\left(\mathbf{K}\mathbf{g}(\sigma^{(r)}), \mathbf{K}\left[\sigma^{(r+1),\text{ni}} \circ \mathbf{g}(\sigma^{(r)})\right]\right) = \left(\mathbf{K}\mathbf{g}(\sigma^{(r)}), \mathbf{K}\left[\sigma^{(r)} \circ \mathbf{g}(\sigma^{(r)}) + \mathbf{w}(\sigma^{(r)})\right]\right)$$

which is precisely the process of the Iterative Convex Minorant (ICM) algorithm of Groeneboom and Wellner.

The "upwards" and "downwards" jumps described by Groeneboom & Wellner (1992) in § 3, Part 2] as determining blocks of constant values for the CDF correspond in fact to the left- and right-endpoint equivalence classes described in § 1.6, which themselves identify the boundaries between maximal antichains.

While the ICM algorithm does not always converge, a modified algorithm which maximizes the likelihood along the line $\lambda\sigma^{(r)} + (1-\lambda)\sigma^{(r+1)}$ for $0 \leq \lambda \leq 1$ was shown by Jongbloed (1998) to converge globally. A constant value of $\lambda = 0$ was observed to take the ICM algorithm to convergence in all the examples we surveyed. The hybrid EM/ICM algorithm introduced by Wellner & Zhan (1997) was shown by way of simulation to converge much more rapidly than either the EM or ICM algorithm. In this case, an extra step of the nature of (4.6) is taken after the isotonization step. Preliminary trials suggest that the rate of convergence is improved by concentrating on the maximal antichains and that the hybrid EM/ICM algorithm provides substantial improvement in this case as well.

Perspectives

The work on minimal covers and weak order partitioning, though combinatorial in nature, was motivated by the need to provide a basic methodology to perform nonparametric analysis on interval censored data under the assumption of underlying order (page 4). A comparative assessment between this assumption and the assumption of pure incomparability still needs to be done, at least empirically. Theorem 3.4 established a strong connection between the maximal antichain structure and the linear extension set of an interval order, by way of marked configurations. This connection may be a promising starting point to compare the two assumptions for interval censored data.

In the case of CDF estimation, the language of order theory has identified new equivalent invariants which are central to the estimation process, namely the Petrie matrix and the characteristic matrix. The importance of dealing with endpoint equivalence classes rather than real-valued endpoints has led to a simple proof of uniqueness of the NPMLE of $\mathbf{p}$, considered as an estimate on $\mathcal{M}$. The language has also removed some confusion about the structure of self-consistent estimates, and provided a simple expression for some important results concerning them. The simple algebraic language of order theory has made the identity between the Kuhn-Tucker and Fenchel duality conditions obvious. It has also enabled a painless simplification of the isotonic regression methods of Groeneboom & Wellner (1992). Order theory has demonstrated itself to be an efficient data reduction tool and a help rather than a hindrance in applications.

Self-consistency and nonparametric maximum likelihood lie under the umbrella of the assumption of pure incomparability (page 4), relying as they do on the maximal antichain structure and disregarding the linear extension set. Estimates obtained through these methods routinely remove maximal antichains by assigning zero mass to them, thereby creating precedence between observations where none existed in the data, as illustrated in Example 4.1. When new precedence relations are created by a CDF estimate, the estimate becomes incompatible with some linear extensions of the interval order. We can consider these maximal antichain removals as the end result of the pooling of maximal antichains necessary to force the CDF estimate to be isotonic. This pooling can potentially create many new and spurious ordering relations, especially for small sample sizes. An important and useful research task could be to determine a CDF estimate based on the assumption of underlying order. It may be possible, for instance, to construct a hybrid estimate designed both to put mass on the maximal antichains (to be mapped onto the real line by $H_{\mathcal{M}}(\mathcal{M})$) and to approximate as closely as possible the pairwise precedence proportions found in the set of linear extensions. The optimal properties of alternative estimates based on the set of linear extensions would likely differ from those of the classical nonparametric estimates. For instance, such estimates would not need to be self-consistent, a requirement which may have intuitive appeal but often forces estimates into unnatural contortions. Consistency and asymptotic self-consistency may be sufficient targets for which to aim, in the context of a precedence proportion-optimal estimator.

In light of the above comments, a closer investigation of the properties of rank tests based on the NPMLE of the CDF, such as those described in Petroni & Wolfe (1994) and Fay (1996), should be undertaken. The robustness of such rank tests to incomparability misspecification by CDF estimates must be determined.

Another application of our approach concerns bivariate interval censored data. Such data can be visualized in the plane as a set of points, line segments, half-lines and variously bounded and unbounded squares, which we will all group under the name of "boxes". By extending the arguments of Peto (1973) and Turnbull (1976), we can easily identify the only areas in the plane where the bivariate NPMLE of the probability function can put mass. We must first determine the maximal cliques of the intersection graph of the boxes (see Golumbic, 1980, Chapters 1 and 2), that is, of the graph $(X, \sim)$, where X stands for the set of boxes (i.e. the bivariate data) and $a \sim b$ for $a, b \in X$ only if the boxes represented by a and b have non-null intersection. Box intersection graphs, however, are not triangulated, and as a consequence the fast algorithms for determining the maximal cliques of perfect graphs (Golumbic, 1980, Chapter 2) are not available. However, algorithms based on the product of the univariate maximal antichain sets can alleviate this problem. A maximal clique can be identified with the intersection of the boxes it contains on the plane. It is then a simple matter to show that the NPMLE of the probability function can only put mass within these intersections.

Self-consistency via a properly seeded EM algorithm can then be used to determine the probability function estimate. This approach has been taken before for the case of right-censored data, but identifiability problems have caused a flurry of competing estimators to be devised, some very closely related to the NPMLE. The CDF estimate for bivariate will be subject to a partial ordering called a *trapezoid order* (see Dagan, Golumbic & Pinter, 1988, and Felsner, Müller & Wernisch, 1997, among others). It may be that isotonic regression techniques can be applied to this estimation problem using this partial ordering rather than the linear ordering which obtains in the univariate case.

Closely related to the question of characterizing *a priori* the effective support of the probability function NPMLE is that of determining whether $H_{\mathcal{M}}(\mathcal{M})$ is always the best we can do in terms of predicting the support of this NPMLE. To couch the problem more formally, consider the $m \times m$ characteristic matrix χ of an interval order. The question is then: is there always an interval order with $m \times m$ characteristic matrix Γ such that $\Gamma_{ij} = 0$ if and only if $\chi_{ij} = 0$, and such that its probability function NPMLE has no entry equal to zero?

Another problem of interest is to refine the relationship between covers and self-consistent estimates. Why do some covers not correspond to a self-consistent estimate? The question is in fact a generalization of the NPMLE effective support characterization problem mentioned above. The advantage of dealing with the potentially large array of all self-consistent estimates is that we may find a clue in the lattice-like structure of covers as to why self-consistency requires the removal of some maximal antichains and how it can do so while maximizing the likelihood.

One of the main limitations of our approach lies in the necessity for permutation arguments to apply under null hypotheses. This necessity is best illustrated with the simple case of rank tests, where bias can be introduced if the inspection processes Q_j described in Chapter 1, § 2.1 are not permutable between different samples (see Mantel, 1967, for a description of the source of this bias for doubly censored data). It may be that the need for permutation arguments can be obviated with more research.

Another limitation, likely to be correctable, is that no proper toolbox of order theoretic methods seems to exist for consistency analysis. It seems likely, however, that the theory of infinite interval orders and of random graphs can be brought to bear on this question. Consistency of the NPMLE has been well-established, as we mentioned in Chapter 4, § 1, using approaches different from order theory. Alternative estimators may eventually require some asymptotic firepower from our approach in order to be shown consistent.

The looming absence in our study is that of nonparametric hazards models. We have not considered them at all for reasons of focus, and a full order theoretic study of them is still needed.

In spite of these few shortcomings, this work has introduced concepts and a language based on order theory that have helped to clarify several outstanding issues in the analysis of interval censored data. The mathematical nature of such data is profoundly combinatorial; combinatorial mathematics, specifically order theory, has much to say about its structure and analysis. This language has the potential to draw together the methodologies associated with different types of censored data.

Appendix A

Matrix notation and Hadamard exponentiation

We first briefly expose a shorthand notation and a few conventions used to simplify the expression of several constructs and equations occuring in this work. Boldface capital letters represent complex matrices or vectors, boldface lowercase letters will represent complex vectors, and normal lowercase letters will represent complex numbers. C refers to the set of complex numbers. A matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ will have i, jth element a_{ij}, $i = 1, \ldots, m, j = 1, \ldots, n$, and a vector $\mathbf{v} \in \mathbb{C}^{m \times 1}$ will have ith element v_i.

- For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ define $\mathbf{v}_1 < \mathbf{v}_2$ to mean $v_i < w_i$ for $i = 1, \ldots, n$, and similarly for $>$, $\leq$ and $\geq$.
- Define the identity matrix $\mathbf{I}_m = [\delta_{ij}]_{i,j=1,\ldots,m}$, where $\delta_{ii} = 1$ and $\delta_{ij} = 0$ whenever $i \neq j$.
- Define the column vector $\mathbf{e}_m^k = [\delta_{ik}]$, the kth column of $\mathbf{I}_m$.
- Define the column vector $\mathbf{e}_m = [1]^{m \times 1}$, having all entries equal to 1.
- For a vector $\mathbf{v} \in \mathbb{C}^{m \times 1}$, define $\mathbf{D}_\mathbf{v} \in \mathbb{C}^{m \times m}$ by $(\mathbf{D}_\mathbf{v})_{ij} = \delta_{ij} v_i$, $i, j = 1, \ldots, m$, the diagonal matrix with diagonal $\mathbf{v}$.
- Denote transposition by $'$, that is, if $\mathbf{B} = \mathbf{A}'$, then $b_{ij} = a_{ji}$.
- For $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m \times n}$, define the Hadamard (or elementwise) product $\circ$ by $(\mathbf{A} \circ \mathbf{B})_{ij} = a_{ij} b_{ij}$.
- Define $0^0 = 1$, and any empty product $\prod_{a \in \emptyset} a = 1$.

Definition A.1 *Let* $\mathbf{A} = [k_{sj}] \in \mathrm{C}^{m\times n}$ *and* $\mathbf{B} = [m_{si}] \in \mathrm{C}^{m\times r}$ *Then define* $\mathbf{A}^{\mathbf{B}} = \left[(\mathbf{A}^{\mathbf{B}})_{ij}\right] \in \mathrm{C}^{r\times n}$ *by*

$$\left(\mathbf{A}^{\mathbf{B}}\right)_{ij} = \begin{cases} \prod_{s=1}^{m} a_{sj}^{b_{si}} & \textit{if } a_{sj} \neq 0 \textit{ or } b_{sj} \in \mathrm{I\!R}^{+} \bigcup 0 \textit{ for all } s = 1,\ldots,m \\ \text{undefined} & \textit{otherwise.} \end{cases}$$

We call the operator $h(\mathbf{A},\mathbf{B}) = \mathbf{A}^{\mathbf{B}}$ Hadamard exponentiation.

Example A.1 *Let* $\mathbf{v}, \in \mathrm{I\!R}^{m\times 1}$ *Then* $\mathbf{v}^{\mathbf{e}^{k}} = v_k$ *and* $\mathbf{v}^{\mathbf{e}} = \prod_{s=1}^{m} v_s$.

Example A.2 *Let* $\mathbf{q} \in \mathrm{I\!R}^{m\times 1}$, *and let* $\mathbf{v} \in \mathrm{I\!R}^{m\times 1}$ *be such that* $v_k > 0$ *whenever* $q_k < 0$. *Then* $\mathbf{v}^{\mathbf{q}} = \prod_{s=1}^{m} v_s^{q_s}$.

Example A.3 *Let* $\mathbf{v} \in \mathrm{I\!R}^{m\times 1}$ *be such that* $v_k \neq 0$ *for* $k = 1,\ldots,m$. *Then* $\mathbf{v}^{-\mathbf{I}_m} = \left[v_1^{-1},\ldots,v_m^{-1}\right]'$.

Example A.4 *Let* $\alpha \in \mathrm{C} \setminus \{0\}$ *and* $\mathbf{v} \in \mathrm{C}^{m\times 1}$. *Then* $\alpha^{\mathbf{v}'} = [\alpha^{v_1},\ldots,\alpha^{v_m}]'$.

Hadamard exponentiation is a generalization of matrix-matrix exponentiation as defined in Barradas & Cohen (1994).

Theorem A.1 is not used in this work, but is included for completeness.

Theorem A.1 *Let* $\mathbf{A}, \mathbf{A}_1, \mathbf{A}_2 \in \mathrm{C}^{m\times n}$, $\mathbf{B}, \mathbf{B}_1, \mathbf{B}_2 \in \mathrm{C}^{m\times r}$, $\mathbf{C} \in \mathrm{C}^{r\times q}$. *Then*

1. $\mathbf{A}^{\mathbf{I}_m} = \mathbf{A}$.
2. *If* $\alpha \in \mathrm{C}$, *then* $(\alpha\mathbf{A})^{\mathbf{B}} = \mathbf{D}_{\alpha^{\mathbf{e}'\mathbf{B}}}\mathbf{A}^{\mathbf{B}}$.
3. $\mathbf{A}^{\mathbf{B}_1} \circ \mathbf{A}^{\mathbf{B}_2} = \mathbf{A}^{\mathbf{B}_1+\mathbf{B}_2}$.
4. $\mathbf{A}_1^{\mathbf{B}} \circ \mathbf{A}_2^{\mathbf{B}} = (\mathbf{A}_1 \circ \mathbf{A}_2)^{\mathbf{B}}$.
5. $(\mathbf{A}^{\mathbf{B}})^{\mathbf{C}} = \mathbf{A}^{\mathbf{BC}}$.

Proof of 5.

$$
\begin{aligned}
[(\mathbf{A}^{\mathbf{B}})^{\mathbf{C}}] &= \left[\prod_{s=1}^{m} a_{sj}^{b_{si}}\right]^{L}_{i=1,\ldots,r;j=1,\ldots,n} \\
&= \left[\prod_{i=1}^{r}\left(\prod_{s=1}^{m} a_{sj}^{b_{si}}\right)^{c_{ik}}\right]_{k=1,\ldots,q;j=1,\ldots,n} \\
&= \left[\prod_{s=1}^{m}\left(\prod_{i=1}^{r} a_{sj}^{b_{si}c_{ik}}\right)\right]_{k=1,\ldots,q;j=1,\ldots,n} \\
&= \left[\prod_{s=1}^{m} a_{sj}^{\sum_{i=1}^{r} b_{si}c_{ik}}\right]_{k=1,\ldots,q;j=1,\ldots,n} \\
&= \left[\prod_{s=1}^{m} a_{sj}^{(BC)_{sk}}\right]_{k=1,\ldots,q;j=1,\ldots,n} \\
&= \mathbf{A}^{\mathbf{BC}}
\end{aligned}
$$

□

Bibliography

ABEL, U. (1986). A nonparametric test against ordered alternatives for data defined by intervals. *Statistica Neerlandica* **40**, 87–91.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. & BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions: the Theory and Application of Isotonic Regression.* Wiley, New York.

BARRADAS, I. & COHEN, J. E. (1994). Iterated exponentiation, matrix-matrix exponentiation, and entropy. *J. Math. Anal. Appl.* **183**, 76–88.

BECKER, R. A., CHAMBERS, J. M. & WILKS, A. R. (1988). *The New S Language.* Wadsworth, Pacific Grove.

BEHRENDT, G. (1988). Maximal antichains in partially ordered sets. *Ars Combin.* **25C**, 149–151.

BERTOSSI, A. A. & BONUCELLI, M. A. (1987). Some parallel algorithms on interval graphs. *Discrete Appl. Math.* **16**, 101–111.

BÖHNING, D., SCHLATTMANN, P. & DIETZ, E. (1996). Interval censored data: A note on the nonparametric maximum likelihood estimator of the distribution function. *Biometrika* **83**, 462–466.

BRIGHTWELL, G., FISHBURN, P. C. & WINKLER, P. (1993). Interval orders and linear extension cycles. *Ars Combin.* **36**, 283–288.

COCHRAN, W. G. (1977). *Sampling Techniques.* Wiley, New York, 3rd edition.

DAGAN, I., GOLUMBIC, M. C. & PINTER, R. Y. (1988). Trapezoid graphs and their coloring. *Discrete Appl. Math.* **21**, 35–46.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1–38.

DILWORTH, R. P. (1950). A decomposition theorem for partially ordered sets. *Ann. of Math.* **51**, 161–166.

DUSHNIK, B. & MILLER, E. W. (1941). Partially ordered sets. *Amer. J. Math.* **63**, 600–610.

EFRON, B. (1967). The two-sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4**, 831–853.

FAY, M. P. (1996). Rank invariant tests for interval censored data under the grouped continuous model. *Biometrics* **52**, 811–822.

FEINBERG, M. (1963). Fibonacci-Tribonacci. *Fibonacci Quart.* **1**, 71–74.

FELSNER, S. (1992). *Interval Orders: Combinatorial Structure and Algorithms.* PhD thesis, Technischen Universität Berlin.

FELSNER, S., MÜLLER, R. & WERNISCH, L. (1997). Trapezoid graphs and generalizations, geometry and algorithms. *Discrete Appl. Math.* **74**, 13–32.

FINKELSTEIN, D. M. & WOLFE, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41**, 933–945.

FISHBURN, P. C. (1973). Interval representations for interval orders and semiorders. *J. Math. Psych.* **10**, 91–105.

FISHBURN, P. C. (1985). *Interval Orders and Interval Graphs.* Wiley, New York.

FULKERSON, D. R. & GROSS, O. A. (1965). Incidence matrices and interval graphs. *Pac. J. Math.* **15**, 835–855.

GAVRIL, F. (1972). Algorithms for minimum coloring, maximum clique, minimum covering by cliques, and maximum independent set of a chordal graph. *SIAM J. Comput.* **1**, 180–187.

GEHAN, E. A. (1965a). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52**, 650–653.

GEHAN, E. A. (1965b). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52**, 203–223.

GENTLEMAN, R. & GEYER, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika* **81**, 618–623.

GOLUMBIC, M. C. (1980). *Algorithmic Graph Theory and Perfect Graphs.* Academic Press, New York.

GROENEBOOM, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. Technical Report 91-53, Faculty of Technical Mathematics and Informatics, Delft University of Technology.

Groeneboom, P. (1996). Lectures on inverse problems. In Bernard, P., editor, *Lectures on Probability Theory and Statistics, École d'été de Probabilités de Saint-Flour XXIV-1994*, volume 1648 of *Lecture Notes in Mathematics*, pages 67–164. Springer.

Groeneboom, P. & Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, volume 19 of *DMV Seminar*. Birkhäuser Verlag, Basel.

Gu, M. G. & Zhang, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21**, 611–624.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Horn, R. A. & Johnson, C. R. (1985). *Matrix Analysis.* Cambridge University Press, Cambridge.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.* **7**, 310–321.

Kalvin, A. D. & Varol, Y. L. (1983). On the generation of all topological sortings. *J. Algorithms* **4**, 150–162.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457–481.

Knuth, D. E. & Szwarcfiter, J. L. (1974). A structured program to generate all topological sorting arrangements. *Inform. Proc. Lett.* **2**, 153–157.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day.

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observation. *Biometrics* **23**, 65–78.

Matthews, P. (1991). Generating a random linear extension of a partial order. *Ann. Probab.* **19**, 1367–1392.

Mykland, P. A. & Ren, J.-J. (1996). Algorithms for coomputing self-consistent and maximum likelihood estimators with doubly censored data. *Ann. Statist.* **24**, 1740–1764.

Peto, R. (1973). Experimental survival curves for interval censored data. *Appl. Statist.* **22**, 86–91.

Petroni, G. R. & Wolfe, R. A. (1994). A two-sample test for stochastic ordering with interval-censored data. *Biometrics* **50**, 77–87.

PRENTICE, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–180.

REDNER, R. (1981). Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions. *Ann. Statist.* **9**, 225–228.

RHEE, C. & LIANG, Y. D. (1996). An NC algorithm for the clique cover problem in cocomparability graphs and its application. *Inform. Process. Lett.* **57**, 287–290.

ROBERTSON, T., WRIGHT, F. T. & DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference.* Wiley, Chichester.

SAATY, T. L. (1970). *Optimization in Integers and Related Extremal Problems.* McGraw-Hill, New York.

SÄRNDAL, C.-E., SWENSSON, B. & WRETMAN, J. (1992). *Model Assisted Survey Sampling.* Springer, New York.

SCHEMPER, M. (1984). A generalized Friedman test for data defined by intervals. *Biom. J.* **3**, 305–308.

SCHEMPER, M. (1991). Generalized rank transformations for tests of survival. *Biom. J.* **33**, 73–79.

SELF, S. G. & GROSSMAN, E. A. (1986). Linear rank tests for interval-censored data with application to PCB levels in adipose tissue of transformer repair workers. *Biometrics* **42**, 521–530.

SMITH, R. L. & TIERNEY, L. (1996). Exact transition probabilities for the independence Metropolis sampler. From the MCMC preprint server.

SPICKERMAN, W. R. (1982). Binet's formula for the Tribonacci sequence. *Fibonacci Quart.* **20**, 118–120.

SZPILRAJN, E. (1930). Sur l'extension de l'ordre partiel. *Fundam. Math.* **16**, 386–389.

THOMPSON, S. K. (1992). *Sampling.* Wiley, New York.

TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69**, 169–173.

TURNBULL, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Statist. Soc. B* **38**, 290–295.

WELLNER, J. A. & ZHAN, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J. Amer. Statist. Assoc.* **92**, 945–959.

WILLE, R. (1981). Restructuring lattice theory. In Rival, I., editor, *Ordered Sets*, volume 83 of *Series C: Mathematical and Physical Sciences*, pages 445–470, Dordrecht-Boston, MA. NATO Advanced Study Institute, D. Reidel Publishing Company.

WINKLER, P. M. (1986). Correlation and order. *Contemp. Math.* **57**, 151–174.

WU, C. F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11**, 95–103.