# Statistical Models for Rank Data using Partial Orders: an Introduction

**Geoff K. Nicholls**
*Department of Statistics*
*University of Oxford, UK*

**Jeong Eun Lee**
*Department of Statistics*
*University of Auckland, NZ*

**Dongqing Li**
*Department of Statistics*
*University of Oxford, UK*

## Abstract

In rank-order data, assessors give preference orders over items in choice sets. Each order lists items in order from best to worst. Data of this sort include results from multi-player competitions, voting preferences, records of queues and data recording the orders in which a set of events are seen to occur. Well known parametric models for list-data include the Mallows model and the Plackett-Luce model. These models seek a total order which is "central" to the lists provided by the assessors. Extensions model the list-data as realisations of a mixture of distributions each centred on a total order. Recent work has relaxed the requirement that the centering order be a total order and instead summarise the ordered lists using partial orders. Lists are random linear extensions of a partial order or linear extensions observed with noise. Evaluation of the likelihood costs #P so applications are restricted to choice sets of up to around 20 elements. For longer lists, Vertex Series Parallel partial orders admit likelihood evaluation in linear time. They are not covered here.

**Keywords:** Bayesian Inference, Partial Orders, Linear Extensions, Plckett-Luce models.

## 1. Introduction

This document is intended as a introduction to statistical modelling of rank-order data using partial orders. We sometimes call the orders in the data "lists", because the first data of this sort we looked at were "witness lists" recording the lists of witnesses to legal documents. Most work to date analysing and summarising rank-order data works with total orders. A total order on a set of items is what you are used to - the list data are total orders - each item has a place and we can always say how any pair of items are ordered. Partial orders generalise this setup by allowing some pairs of items to be "incomparable". That doesn't mean the two items are tied, or equal in rank, it means that there is no order relation between the pair. Instead of trying to summarise a collection of lists with a total order we try to summarise it with something weaker: a partial order. We don't enforce an order in our summary if there really isn't one.

How is it possible for a pair of items to be "incomparable"? Sometimes we have a rule for determining preference, and the rule simply doesn't determine an order. For example, suppose we prefer fruit that are sweeter and crisper. If we are presented with a green apple and ripe banana, so one is crisper and the other sweeter, our rule doesn't determine a preference. As another example, in Oxford, the master of a college can direct the college tutors and the administrative staff (for example, the registrar) to do some task. However, the tutors and the administrators don't really have any sort of power relation. The hierarchy in the college would be something like master $\succ$ tutor and master $\succ$ registrar but there is no order relation between tutor and registrar.

We therefore seek a mathematical language for order relations that are incomplete. We will assume that the order relations are transitive. For example, if master $\succ$ tutor and tutor $\succ$ student then master $\succ$ student. Transitivity is quite a basic property of preference relations that even pre-speech babies and animals understand. Although it wont always hold, we gain alot of information and loose little real applicability by imposing it in our model. A mathematical object specifying a set of order relations which are transitive and may be incomplete is a partial order. We define this a bit more formally below.

The partial order models we describe below could be used in most settings where Mallows or Plackett-Luce models are currently used. Consideration should be given to whether the models fit the data well and express prior knowledge naturally, but that's statistical modelling. When we do statistical modelling it sometimes holds that the observation model is derived from a stochastic process that is written down as a serious attempt to model events in the real world. In that case "the elements of the model are in one to one correspondence with elements of reality". This is the setting where partial order models arise most naturally and seem unavoidable. To date, all published statistical modelling work in which the "parameter" is a partial order and the "data" are ordered lists was motivated in this way. The key examples are Mannila (2008), Sakoparnig and Beerenwinkel (2012) and Nicholls et al. (2025) and citations therein.

Statistical models are not usually motivated or constructed as models of reality; they set out to model data distributions in a heuristic way. Mallows models are good at capturing the distribution of rank order lists, but are often applied in settings where the preference orders they recover are simply summaries of patterns in the data. Put another way, no unknown true ranking, existing in the real world, is defined.

In Nicholls et al. (2025), the partial order represents a social hierarchy in which some individuals have higher status than others. If we think about the social hierarchies we live in, they are not always total orders. Since hierarchies express power relations, they are almost always transitive. That pretty much leads us to partial order representations. The main application of this parameterisation was to reconstruct a social status hierarchy for bishops who lived in the eleventh and twelfth centuries in England, Wales and Normandy. The data were the witness lists mentioned above. Historians know that witnesses are listed in order of status so lists respect the hierarchy, with higher status bishops ahead of those of lower status: in our terms the data are "linear extensions" (AKA "topological orders") that respect the partial order. Below we give a stochastic process (a random queue) that models the random lists given the partial order. Since the partial order expresses relations that really do exist between actors, and the data are realised from the partial order by a stochastic process intended to model reality, the elements of our model correspond to elements of reality, or at least, that's the intention.

In this setting there exists an unknown true social hierarchy and we would like to recover it. The statistical problem is to recover a partial order from random orders that respect the partial order. This is the type of problem where Bayesian inference is natural: there is an unknown true value for the "parameter" (the true social hierarchy); because the parameter is something that exists in the real world, we can have prior knowledge about its value (some social structures are a priori more probable than others).

We are a bit loose with terminology. There isn't always an assessor who forms a preference order and delivers an ordered list. For example, Jiang and Nicholls (2021) analyse race outcomes in Formula One races. Also, when discussing applications to social hierarchies, the items in the choice set are sometimes called *actors*, following the terminology of social network analysis. Below we define the terms *poset* and *partial order* in a way that distinguishes them, but then use the terms exchangeably. Finally, as we explain below, posets are one to one with transitively closed DAGs: in the notation for a poset, $(S, \succ)$ and a graph $(V, E)$ the choice set $S$ plays the role of the vertex set $V$, so an item $i \in S$ is a vertex or node on $V$, and the relations $\succ$ match the edge set $E$. Posets are easily visualised using DAGs (a directed edge $\langle a, b\rangle$ conveys $a \succ b$) so we switch between terminology, referring to DAGs as if they were posets, without any particular care.

In the following we give some mathematical definitions for partial orders and then the observation model (giving the likelihood) and the prior. The observation model is a process in which the partial order is a parameter and the data are random lists. The prior and posterior are probability distributions over partial orders.

Some of the applications we discuss (for example, to social hierarchies and queues) will seem pretty niche. This is just to be concrete. The models we describe can be applied in almost any setting where models like Mallows and Placket-Luce are currently applied.

## 2. Partial orders and suborders

In preparing this material, especially the later parts discussing models for random partial orders, we found Brightwell (1993) helpful. Let $\mathcal{M} = [M]$, $[M] = \{1, \dots, M\}$ be the universe of objects over which preferences can be given and suppose there are $M = |\mathcal{M}|$ in all. Let $\mathcal{B}_{\mathcal{M}}$ be the set of all subsets of $\mathcal{M}$ excluding the empty set and let $S \in \mathcal{B}_{\mathcal{M}}$ be a given *choice set* with $m = |S|$ elements.

### 2.1 Partial orders

A *poset* $h$ is a choice set equipped with a *partial order* $\succ_h$. We use strict partial orders to describe preferences over objects in a choice set. Poset $h = (\mathcal{M}, \succ_h)$ on $\mathcal{M}$ is a set with order relations $j_1 \succ_h j_2$ for $j_1, j_2 \in \mathcal{M}$. An example is shown in Figure 1. The relations are anti-symmetric (if $j_1 \succ_h j_2$ then not $j_2 \succ_h j_1$) and transitive (if $j_1 \succ_h j_2$ and $j_2 \succ_h j_3$ then $j_1 \succ_h j_3$) but need not be complete, so there may be pairs where neither $j_1 \succ_h j_2$ nor $j_2 \succ_h j_1$. We work with strict partial orders, so relations are irreflexive (no relations $j \succ_h j$, and no ties). Elements $j \in \mathcal{M}$ such that $\{j' \in \mathcal{M}; j' \succ j\}$ is empty are called maximal elements. Denote by $\max(h)$ the set of maximal elements of $h$. As $\mathcal{M}$ is finite, $\max(h)$ is non-empty by Zorn's lemma. If the choice set is restricted from $\mathcal{M}$ to $S$, then preferences

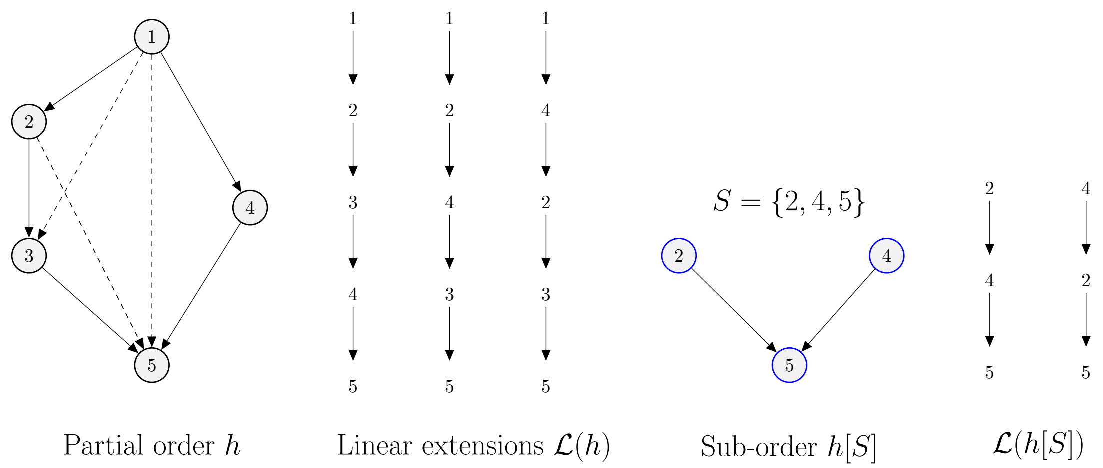



Figure 1: The example partial order $h$ (left) on $\mathcal{M} = \{1, 2, 3, 4, 5\}$ has three linear extensions (left-centre). Dashed lines in $h$ are relations implied by transitivity. Its suborder on $S = \{2, 4, 5\}$ (right-centre) has two linear extensions (right).

are given by the *suborder* $h[S] = (S, \succ_h)$ containing just the subset of relations $j_1 \succ_h j_2$ in $h$ for pairs $j_1, j_2 \in S$. An example is given in Figure 1. Let $\mathcal{H}_S$ be the set of all posets on $S$. We often refer to models for random posets $(S, \succ_h)$ as models for random partial orders, as the choice sets are always fixed, only the relations in the order component are random.

Partial orders $h \in \mathcal{H}_S$ are in one to one correspondence with transitively-closed, directed acyclic graphs with vertex set $S$; the DAG representing $h$ has a directed edge $\langle j_1, j_2 \rangle$ for each order relation $j_1 \succ_h j_2$ in $h$. When we illustrate partial orders we display the DAG. Let $\mathcal{D}_\mathcal{M}$ be the set of all labelled DAGs with $M$ vertices. For $g \in \mathcal{D}_\mathcal{M}$ let $C(g)$ denote the transitive closure of $g$. We simply add to $g$ all the edges implied by transitivity: if $h = C(g)$ then $h \in \mathcal{H}_\mathcal{M}$ is a partial order. Let $R(g)$ be the transitive reduction. This is the unique DAG with the least edges such that, when we make the transitive closure, we get back $C(g)$. When we draw posets we tend to show the reduction as it is a simpler graph conveying the same information. In Figure 1 the transitive reduction of $h$ is shown with solid lines. Relations/edges which are in the closure but not the reduction are dashed.

If for all pairs $j_1, j_2 \in S$ either $j_1 \succ_h j_2$ or $j_2 \succ_h j_1$ then $h$ is a *complete order*. Let $\mathcal{C}_S$ be the set of all complete orders of $S$. Complete orders are one to one with $\mathcal{P}_S$, the set of all permutations of the elements of $S$. An empty order is a partial order with no order relations at all.

In the following we make use of the *depth* of a partial order which we write $d(h)$. The depth is the length of the longest chain or "path" on the DAG representing the partial order (counting vertices visited) so $d : \mathcal{H}_S \to \{1, ..., m\}$. For example the depth of the partial order $h$ at left in Figure 1 is four and the depth of its suborder $h[S]$ at right is two. If $h \in \mathcal{C}_S$ is a complete order then $d(h) = m$ and if $h$ is the empty order then $d(h) = 1$.

### 2.2 Linear extensions of a poset

A *linear extension* of $h$ is any complete order $\ell \in \mathcal{C}_S$ satisfying $j_1 \succ_h j_2 \Rightarrow j_1 \succ_\ell j_2$, so the linear extension "completes" the partial order $\succ_h$. We sometimes abbreviate linear extension as "LE". Figure 1 gives the linear extensions of the example partial order $h$ and its suborder $h[S]$. Denote by $\mathcal{L}[h]$ the set of all linear extensions of $h = (S, \succ_h)$,

$$\mathcal{L}[h] = \{\ell \in \mathcal{C}_S : j_1 \succ_h j_2 \Rightarrow j_1 \succ_\ell j_2 \ \forall\ j_1, j_2 \in S\}$$

For $j \in S$ let

$$\mathcal{L}_j[h] = \{\ell \in \mathcal{L}[h] : \max(\ell) = j\}$$

be the set of all linear extensions of $h$ with maximal element $j$.

### 2.3 The dimension of a poset

It is possible to define the *dimension* of a poset. This is a useful concept when we come to building prior models for random partial orders.

Let $\ell$ and $\ell'$ be two linear extensions of $h = (S, \succ_h)$. The intersection order $\ell \cap \ell' = (\mathcal{M}, \succ_{\ell\cap\ell'})$ has all the order relations shared by $\ell$ and $\ell'$ that is

$$\succ_{\ell\cap\ell'} = \bigcup_{j_1, j_2 \in \mathcal{M}} \{j_1 \succ j_2 : j_1 \succ_\ell j_2 \text{ and } j_1 \succ_{\ell'} j_2\}$$

For example, if $\succ_\ell$ has the relations $1 \succ_\ell 2 \succ_\ell 3$ and $\succ_{\ell'}$ has $1 \succ_{\ell'} 3 \succ_{\ell'} 2$ then $\ell \cap \ell'$ has the relations $1 \succ_{\ell\cap\ell'} 2$ and $1 \succ_{\ell\cap\ell'} 3$.

If we intersect all the linear extensions in $\mathcal{L}[h]$ we get back $h$,

$$\bigcap_{\ell \in \mathcal{L}[h]} \ell = h.$$

This is because every order relation in $h$ must be present in every LE in $\mathcal{L}[h]$ (as the linear extension "completes" the orders in the poset) so when we intersect all these relations are preserved as they are present in every LE. On the other hand if a relation isn't in $h$, so for example $j_1$ and $j_2$ are unordered in $h$, then the pair will appear as $j_1 \succ_\ell j_2$ in some LEs and as $j_2 \succ_{\ell'} j_1$ in others. When we intersect LEs, these relations will be removed.

Let $L = \{\ell_1, \ldots, \ell_N\}$ be a set of complete orders, $\ell_i \in \mathcal{C}_S$. If it holds that

$$h = \bigcap_{i=1}^{N} \ell_i$$

then we say the set of complete orders $L$ realises $h$. Each of these complete orders must be a linear extension of $h$ so we must have $L \subseteq \mathcal{L}[h]$. The *dimension* of a partial order $D(h)$ is the number of elements in the smallest set of complete orders realising $h$.

For example, the dimension of a complete order $h$ is always one, because we can take $L = \{h\}$. The dimension of the empty order (with $|S| > 1$) is always two as it cant be one and we can always realise it with two LEs by taking any complete order of $S$ and then reversing all the relations: the intersection of these relations will be empty.

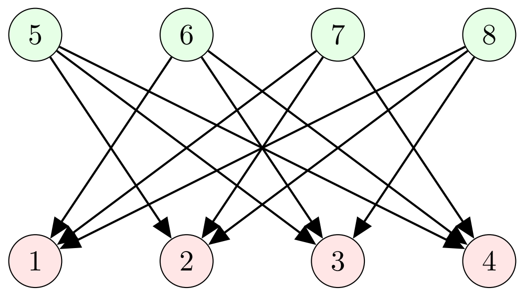



Figure 2: The crown poset on $m = 8$ items. For each even $m \geq 4$, this form of poset has maximum dimension ($D(h) = \lfloor m/2 \rfloor = 4$ here) among all posets in $\mathcal{H}_{[m]}$.

A theorem due to Hiraguchi (1951) says that the dimension of a poset on a set $S$ with $m$ elements is at most $\lfloor m/2 \rfloor$. This is useful when we come to parameterise partial orders in a prior: if we parameterise partial orders using a set of $K$ total orders then we know we can represent any partial order satisfying $\lfloor m/2 \rfloor \leq K$.

As a curiosity it is interesting to see what a poset of dimension $\lfloor m/2 \rfloor$ looks like. They have a special form called a crown. The crown poset on $m = 8$ items is shown in Figure 2. As a (rather pointless) exercise you might like to find four complete orders that realise it.

## 3. Observation model

Data are preference orders or "lists". An observed preference order $y = (S, \succ_y),\ y \in \mathcal{C}_S$ is a complete order on its choice set $S$, or equivalently the ordered permutation $y_{1:m} = (y_1, \dots y_m),\ y_{1:m} \in \mathcal{P}_S$ or "list" we get by indexing $y_i,\ i = 1, \dots m$ so that $y_1 \succ_y y_2 \succ_y \dots \succ_y y_m$. When referring to data, we use the terms list and complete order interchangeably.

There are two ways to realise an order: an assessor may be given a choice set $S$ and return a preference order considering only the elements of $S$, in which case $y$ is *realised on the choice set* $S$; alternatively they may make a preference order $y' \in \mathcal{C}_{\mathcal{M}}$ on the universe of choices and return the suborder $y = y'[S]$, thinning out any choices not in $S$ while retaining the order of those that remain. In this case, $y$ is *realised as a suborder* on $S$. If preferences are *context-independent*, then these two rules give the same observation models and otherwise, preferences are context-dependent. See Appendix A.1 for further discussion of this point.

We focus on the case where the data are realised on a choice set. We model the observed orders $y \in \mathcal{C}_S$ as random linear extensions respecting partially ordered preferences $h \in \mathcal{H}_S$.

### 3.1 Noise-free observation model for the rank-order data

In our first and simplest observation model, a list $y \in \mathcal{C}_S$ observed on a poset $h \in \mathcal{H}_S$ is a linear extension sampled uniformly at random from $\mathcal{L}[h]$ (we modify this later to allow "errors" in lists). Let $p_S(y|h)$ give the probability to realise list $y \in \mathcal{L}[h]$ when the poset is $h = (S, \succ_h)$. If $|\mathcal{L}[h]|$ is the number of elements in $\mathcal{L}[h]$ then the likelihood for $h$ is

$$p_S(y|h) = |\mathcal{L}[h]|^{-1} \mathbb{I}_{y \in \mathcal{L}[h]}. \tag{1}$$

Let $h = (S, \succ_h)$ and $X_0 = \ell_0$ be given with $m = |S|$ and $\ell_0 \in \mathcal{L}[h]$.

For $t \geq 1$, suppose $X_t = \ell$. The next queue state $X_{t+1}$ is determined as follows.

select $i \sim U\{1, \dots, m-1\}$ and propose to swap $\ell_i$ and $\ell_{i+1}$:

set $\ell'_k = \ell_k$ for each $k \neq i, i+1$

set $\ell'_i = \ell_{i+1}$ and $\ell'_{i+1} = \ell_i$

if $\ell' \in \mathcal{L}[h]$ set $X_{t+1} = \ell'$ and otherwise set $X_{t+1} = \ell$.

Figure 3: Stochastic process evolving the queue. The queue state $X_t$ evolves according to a process where neighboring pairs of items in the queue are randomly swapped.

This model simply expresses the idea that the objects constrained by $h$ are otherwise exchangeable, so if $y$ and $y'$ are two lists in $\mathcal{L}[h]$, then we require $p_S(y|h) = p_S(y'|h)$.

The model in (1) is derived in Nicholls et al. (2025) from a stochastic queue process in which the random linear extension is a draw from the equilibrium of a stochastic process on $\mathcal{L}[h]$. We referred to this in the introduction. Suppose the items we are ordering are "actors" (in the language of social network models). The set $S$ is the set of actor labels. Suppose the actors are in a social hierarchy expressed as a partial order $h = (S, \succ_h)$: for $j_1, j_2 \in S$ the relation $j_1 \succ_h j_2$ means that actor $j_1$ has higher status than actor $j_2$. Suppose the actors queue for a resource with higher status individuals coming ahead of lower status individuals in the queue. As the actors stand in the queue they randomly swap positions, being careful not to swap ahead of a higher status actor. Let $X_t,\ t \geq 0$ give the state of the queue at time step $t$. Here $X_t$ is a Markov Chain on $\mathcal{L}[h]$ defined by the update given in Figure 3.

The Markov chain $X_t$ defining the evolving queue is irreducible on $\mathcal{L}[h]$ (Jiang and Nicholls (2021) discuss this well known result). It is also "doubly stochastic": if

$$P_{\ell,\ell'} = \Pr(X_{t+1} = \ell' | X_t = \ell)$$

then

$$P_{\ell,\ell'} = \begin{cases} (m-1)^{-1} & \text{if } \ell, \ell' \text{ differ by swapping a pair of neighbors and} \\ 0 & \text{otherwise.} \end{cases}$$

Since $P = P^T$ the constant vector $\pi \propto 1$ with $|\mathcal{L}[h]|$ elements satisfies $\pi P = \pi$ and so the equilibrium of the chain is uniform on $\mathcal{L}[h]$, that is $X_t \to p_S(\cdot|h)$ in distribution.

### 3.2 Computing the number of linear extensions of a partial order

Computing $|\mathcal{L}[h]|$ is #P-complete (Brightwell and Winkler, 1991), so we cannot evaluate $p_S(y|h)$ for general $h \in \mathcal{H}_S$ and large $m$. Counting is fast up to about $m = 20$ and feasible up to about $m = 40$ using `lecount`(), which implements methods in Kangas et al. (2016). For greater $m$-values, we follow Jiang et al. (2023) and restrict $h$ to a class of partial orders where $|\mathcal{L}[h]|$ can be computed in linear time.

```
nle←function(h = (S, ≻_h))
    If |S| = 1 return total=1
    total=0
    for j in max(h)
        set h_{-j} = (S \ {j}, ≻_h)
        total←total+nle(h_{-j})
    return total
```

Figure 4: Function nle($h$) counting the linear extensions of $h$ recursively using (2).

Knuth and Szwarcfiter (1974) give a recursive algorithm for counting linear extensions. The set of linear extensions of a poset $h$ can be partitioned on the first item in each list. Each of these sets contains a collection of lists all headed by the same item. We have

$$|\mathcal{L}[h]| = \sum_{j \in \max(h)} |\mathcal{L}_j[h]|. \tag{2}$$

If we take LEs in $\mathcal{L}_j(h)$, the set of all LEs headed by $j$, and remove the top element $j$ from each list, that's the same as removing $j$ from $h$ at the start, to get the suborder $h_{-j} = (S \setminus \{j\}, \succ_h)$, and then forming the set $\mathcal{L}[h_{-j}]$ of LEs of the suborder. It follows that

$$|\mathcal{L}_j[h]| = |\mathcal{L}[h_{-j}]| \tag{3}$$

and so to count $|\mathcal{L}_j[h]|$ we apply (2) to $\mathcal{L}[h_{-j}]$ and so on recursively removing one node each time. The recursion stops when we reach a partial order with one element, as that has one linear extension. The counting algorithm is given in Figure 4.

### 3.3 Random linear extensions are repeated selection models

Besides counting LEs, we can also use (3) to realise LEs $y \sim p_S(\cdot|h)$ in a sequential way, building up the list one element at a time. This corresponds to an observation model in which the assessor first selects their top preference, then their second and so on. Models of this kind are called *repeated selection* models (Ragain and Ugander, 2018; Seshadri et al., 2020). Many models for preference orders are sequential choice models: our partial order models, the Plackett-Luce model, the CRS model in Seshadri et al. (2019) and some but not all Mallows models (depending on the choice of distance measures taken over $\mathcal{C}_{\mathcal{M}}$). Let $y_{i:m} = (y_i, \ldots, y_m)$ so that $h[y_{i:m}] = (y_{i:m}, \succ_h)$ is the suborder of $h$ restricted to $y_{i:m}$. We can write $p_S(y|h)$ as a telescoping product over suborders

$$p_S(y|h) = \prod_{i=1}^{m-1} q_{y_{i:m}}(y_i|h[y_{i:m}]), \tag{4}$$

where

$$q_{y_{i:m}}(y_i|h[y_{i:m}]) = \frac{|\mathcal{L}_{y_i}[h[y_{i:m}]]|}{|\mathcal{L}[h[y_{i:m}]]|} \tag{5}$$

is the probability $y_i$ is selected next from the remaining choices, and that probability is just the proportion of linear extensions headed by $y_i$. The product in (4) is equal to $p_S(y|h)$ in (1) because

$$|\mathcal{L}_{y_i}[h[y_{i:m}]]| = |\mathcal{L}[h[y_{i+1:m}]]|$$

using (3) again, so the product in (4) is a telescoping product of cancelling factors. It follows that our random linear extension model $p_S(\cdot|h)$ is a repeated selection model.

This further motivates our story of the observation model as a queue. Consider the following process.

1. Set $S^{(1)} = S$. This set records the actors remaining in the queue.
2. For $i = 1, \ldots, m-1$:
   (a) the actors in $S^{(i)}$ form a queue which reaches equilibrium according to the stochastic process set out in Figure 3;
   (b) $y_i$ is taken to be the first actor in the queue;
   (c) set $S^{(i+1)} = S^{(i)} \setminus \{y_i\}$ (update the queue, removing $y_i$).
3. Set $y_m = S^{(m)}$ (the last actor remaining).

The probability to select $y_i$ from the actors $S^{(i)} = \{y_i, \ldots, y_m\}$ that remain the queue is the proportion of linear extensions of $h[y_{i:m}]$ in which $y_i$ comes first and this is $q_{y_{i:m}}(y_i|h[y_{i:m}])$ in (5). Events are independent so the probability to realise $y$ is given by (4).

### 3.4 The full noise free observation model

In many rank-order datasets, the choice set varies from one list to another. An assessor with preferences expressed by a poset $h = (\mathcal{M}, \succ_h)$ gives $N$ preference orders $Y = (Y_1, \ldots, Y_N)$, where $Y_i \in \mathcal{C}_{S_i}$ orders the elements of choice set $S_i \in \mathcal{B}_\mathcal{M}$. For $i \in [N]$ let $m_i = |S_i|$. As lists, $Y_i = (Y_{i,1}, \ldots, Y_{i,m_i})$. Notice we use $y$ for a generic list, $Y_i$ for the $i$'th list in the data and $Y$ for the full data set of lists. The assessor's preferences over $S_i$ are determined by the suborder $h[S_i] = (S_i, \succ_h)$. If the lists are independent given $h$ then the likelihood is

$$p_{S_{1:N}}(Y|h) = \prod_{i=1}^{N} p_{S_i}(Y_i|h[S_i]), \tag{6}$$

where $p_{S_i}(Y_i|h[S_i]) = |\mathcal{L}[h[S_i]]|^{-1}$ per Equation (1).

The observation model we defined here is context dependent in the sense of Appendix A.1. This is discussed with an example in Appendix A.2. Also, we have discussed order data which are "incomplete" in the sense that we only get the order $Y_i$ on $S_i$ and not the full order on $\mathcal{M}$. Top-$k$ data are also incomplete, but qualitatively different as the set $S = \{y_1, \ldots, y_k\}$ is a random outcome of the observation process. However, it is straightforward to write down the likelihood for top-$k$ data in a sequential choice model by truncating the product in (4) at $k$.

### 3.5 The MLE for the partial order

Our focus will be on Bayesian inference for the unknown true poset. However, there is a simple estimator that works well in the noise free case. Consider list data realised with the noise free observation model in Section 3.4 and suppose the choice set is always $\mathcal{M}$ so $S_i = \mathcal{M}$ for $i = 1, \ldots, N$ and all the lists are random linear extensions of all $M$ elements in the poset $h = (\mathcal{M}, \succ_h)$.

We know from Section 2.3 that if we intersect the linear extensions of a poset then we get back the poset. In fact we can get back the poset with as few as $D(h) \le \lfloor M/2 \rfloor$ linear extensions, if we are lucky and our data contain a realiser for $h$. The intersection order,

$$\hat{h} = \bigcap_{i=1}^{N} Y_i,$$

is therefore a consistent estimator for $h$: any order order relation $j_1 \succ_h j_2$ that is present in $h$ will be present in every list (since they respect the poset) and if $j_1$ and $j_2$ are unordered in $h$ then (if $N$ is large enough) we will sometimes see $j_1$ before $j_2$ in the lists and sometimes after, so these relations will be removed in taking the intersection.

In fact $\hat{h}$ is also the MLE for the likelihood in (6),

$$\arg\max_{h \in \mathcal{H}_{\mathcal{M}}} p_{S_{1:N}}(Y|h) = \hat{h},$$

in the case where $S_i = \mathcal{M},\ i = 1, \ldots, N$ so we have full length lists. This observation (and alot more) can be found in Beerenwinkel et al. (2007). It follows because

$$p_{S_{1:N}}(Y|h) = \frac{1}{|\mathcal{L}[h]|^N} \prod_{i=1}^{N} \mathbb{I}_{Y_i \in \mathcal{L}[h]},$$

so we seek the poset $h$ with the least linear extensions that doesn't conflict any of the lists. That is the intersection order: if we add any relation to $\hat{h}$ we must conflict a relation in the data and we we remove one then we only increase the number of linear extensions.

This reasoning works because the data are full length lists. If we have choice sets $S_i \subset \mathcal{M},\ i = 1, \ldots, N$ then it may hold that some pairs $j_1, j_2$ of items are not present together in any choice set. Now it is possible that the relation between $j_1$ and $j_2$ isn't determined. Also it isn't clear how to intersect LEs of suborders on different choice sets. However, we can make an estimator by keeping all relations which are attested in the data and not contradicted. Reasoning as for the MLE shows this will be consistent in the limit of an infinite number of choice sets with the property that each pair $j_1, j_2$ appears in an infinite number of choice sets. However, it is no longer the MLE.

### 3.6 Extending the observation model to allow for noise

We allow for noise in the observation model in Section 3 by allowing individuals to "jump the queue", exploiting the sequential selection structure of the poset likelihood in (4). This is a simple adjustment of the sequential choice interpretaiton of the noise free model set out in Section 3.3.

1. Set $S^{(1)} = S$. This set records the actors remaining in the queue.
2. For $i = 1, \dots, m-1$:
   (a) the actors in $S^{(i)}$ form a queue which reaches equilibrium according to the stochastic process set out in Figure 3.
   (b) with probability $p$ we choose actor $y_i$ at random from $S^{(i)}$;
   (c) otherwise $y_i$ is taken to be the first actor in the queue;
   (d) set $S^{(i+1)} = S^{(i)} \setminus \{y_i\}$ (update the queue, removing $y_i$).
3. Set $y_m = S^{(m)}$ (the last actor remaining).

Figure 5: Sequential-choice description of noisy queue-jumping observation model realising preference ordered lists.

We build up the list sequentially. At each step, with probability $p$ we choose the next actor randomly from those remaining in the queue and otherwise we take the next person in the queue. The process realising a list $y = (y_1, \dots, y_m)$ is given in Figure 5. Since $S^{(i)} = \{y_i, \dots, y_m\}$ has $m - i + 1$ elements, the probability to choose $y_i$ at step $i$ is

$$q(y_i|h[y_{i:m}], p) = \frac{p}{m-i+1} + (1-p)\, q_{y_{i:m}}(y_i|h[y_{i:m}]), \tag{7}$$

with $q_{y_{i:m}}(y_i|h[y_{i:m}])$ given in (5). Since the selection events are independent,

$$p_S(y_{1:m}|h, p) = \prod_{i=1}^{m-1} q(y_i|h[y_{i:m}], p). \tag{8}$$

If we have $N$ conditionally independent lists $Y = (Y_1, \dots, Y_N)$ ordering the elements of choice sets $S_1, \dots, S_N$ then the likelihood for $h$ and $p$ is $p_{S_{1:N}}(Y|h, p)$, a product, as in (6). When we use this model in the posterior, we add a parameter $p$ with prior $\mathrm{Beta}(\alpha_p, \beta_p)$. If we are fitting a poset model then presumably $p$ is fairly small, so we take $\alpha_p = 1$ and $\beta_p = 9$ for a prior mean $p$ equal 0.1, though that depends on the application.

## 4. Bayesian Inference for Partial Orders

If we wish to carry out Bayesian inference for $h$, we need a prior for $h$. When we construct a prior we should pay attention to the key questions we wish to answer with the inference. If we are uncertain about some property of the parameter (the poset $h$ here) then our prior should be non-informative with respect to that property: the prior should be non-informative with respect to the key scientific hypotheses. One function of $h$ that is often of interest is the depth $d(h)$ of the unknown true partial order. For example, this is often of interest in statistical inference for social hierarchy: if the true poset is the empty order

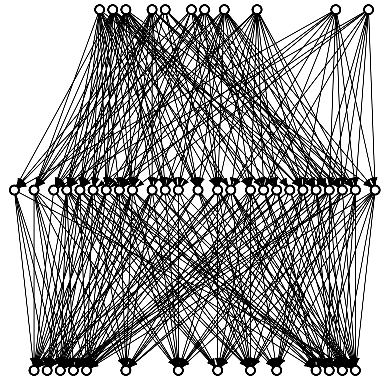


Figure 6: A poset on 50 nodes drawn approximately uniformly at random from $\mathcal{H}_{\mathcal{M}}$ with $M = 50$, illustrating the concentration on orders of depth three at large $M$.

then there are no order relations among the actors. If it is a total order then that is a qualitatively different sort of hierarchy. Both these hierarchies arise as extremes in real communities but are as different as a feudal monarchy and an egalitarian socialist society.

Since the depth of the reconstructed poset is often of interest in the inference, we will ask for the prior distribution over partial orders to be uninformative of the depth of the poset. We seek a prior with the property that if $h \sim \pi_{\mathcal{M}}(\cdot)$ then $d(h) \sim U\{1, \ldots, M\}$, at least approximately.

Another desirable property for a prior it that it be *marignally consistent*. Marginal consistency is the property that the prior distribution $\pi_S(\cdot)$ we write down for posets on a subset $S \in \mathcal{B}_{\mathcal{M}}$ of items is the marginal of our prior for posets on the universe of choices $\mathcal{M}$. Intuitively, our prior for relations on a subset of items is the same as the distribution of those relations in a superset. If we sample a poset from $\pi_{\mathcal{M}}(\cdot)$ and then keep only order relations between items in $S$ that should be the same as sampling a poset from $\pi_S(\cdot)$.

**Definition 1** *A family of probability distributions $\pi_S,\ S \in \mathcal{B}_{\mathcal{M}}$ over posets is marginally consistent if it holds that $h \sim \pi_{\mathcal{M}}(\cdot)$ implies $h[S] \sim \pi_S(\cdot)$ for every $S \in \mathcal{B}_{\mathcal{M}}$.*

This is a property we impose on a *family* of prior probability distributions $\pi_S(\cdot),\ S \in \mathcal{B}_{\mathcal{M}}$. If we simply write down $\pi_S(h),\ h \in \mathcal{H}_S$ for each $S \in \mathcal{B}_{\mathcal{M}}$ then it need not hold.

In Section 4.1 we give an example of a prior family which is not marginally consistent and is strongly informative of depth. We then give a prior on posets which is marginally consistent and gives us control of the prior probability distribution over poset depth.

### 4.1 The uniform prior over partial orders

We might be tempted to take as our prior over partial orders the uniform prior

$$\tilde{\pi}_S(h) = |\mathcal{H}_S|^{-1}, \qquad h \in \mathcal{H}_S.$$

It has two undesirable properties: it is not marginally consistent; it is informative of depth.

Marginal consistency for $\tilde{\pi}_S(\cdot),\ S \in \mathcal{B}_\mathcal{M}$ would require that if $h \sim \tilde{\pi}_\mathcal{M}(\cdot)$ then $h[S] \sim \tilde{\pi}_S(\cdot)$, that is, if we sample a poset uniformly at random from all posets in $\mathcal{H}_\mathcal{M}$ and then for some $S \in \mathcal{B}_\mathcal{M}$ we take the suborder $h[S]$ then this suborder is uniformly distributed over posets in $\mathcal{H}_S$. It is easy to see that that is not going to hold as there are 19 posets with three items and 3 posets with two items and 19 isn't divisible by 3. Essentially $h[S]$ is a many to one map from $\mathcal{H}_\mathcal{M}$ to $\mathcal{H}_S$ and so the probability to get some poset $g \in \mathcal{H}_S$ is the probability to pick one of the posets $h \in \mathcal{H}_\mathcal{M}$ that satisfy $h[S] = g$. Since the preimage of the posets in $\mathcal{H}_S$ cant always have the same number of posets, $h[S]$ cant be uniform.

The other property of the uniform distribution is that if $h \sim \tilde{\pi}_\mathcal{M}(\cdot)$ then the depth $d(h)$ is a random variable which converges to three in the limit that $M \to \infty$: most big posets have depth three (Kleitman and Rothschild, 1975). A poset sampled uniformly at random from $\mathcal{H}_\mathcal{M}$ with $M = 50$ using the MCMC algorithm given in Muir Watt (2015) and Appendix C is shown in Figure 6. This illustrates the rather dramatic depth weighting present in this prior. This prior probability distribution over posets doesn't express the prior knowledge we actually have so we wouldn't want to choose this prior to represent our state of knowledge before seeing the data.

### 4.2 Latent variable parameterisation of a poset

In order to parameterise a prior which is marginally consistent with a parameter controling depth, we start by representing $h$ in terms of a set of continuous latent variables. We adapt models for random partial orders given by Winkler (1985) and Nicholls and Muir Watt (2011). An example of the parameterisation is shown in Figure 7.

Let $U \in \mathbb{R}^{M \times K}$ be a matrix of preference weights with one row $U_{j,:} \in \mathbb{R}^K$ for each object $j \in \mathcal{M}$ in the universe of choices and one column $U_{:,k}$ for each "feature" $k = 1, \ldots, K$. For a pair of objects $j_1, j_2 \in \mathcal{M}$, our rule $h(U) = (\mathcal{M}, \succ_U)$ mapping $U$ to a poset will set $j_1 \succ_U j_2$ when the two rows of weights satisfy $U_{j_1,k} > U_{j_2,k}$ for each $k = 1, \ldots, K$. Looking at Figure 7, we see each row of $U$ is plotted as a path. The rule says that if the paths for items $j_1$ and $j_2$ cross then $j_1$ and $j_2$ are unordered. If they dont cross then the order will be $j_1 \succ_U j_2$ when the path for $j_1$ is higher than that for $j_2$.

We can write this map as an intersection over the orders of the columns of $U$. Let $h(U_{:,k}) = (\mathcal{M}, \succ_k)$ be the complete order on column $k$ with

$$j_1 \succ_k j_2 \Leftrightarrow U_{j_1,k} > U_{j_2,k}.$$

This is always a complete order (well, almost surely) because $>$ is a complete order on the real entries in the $k$'th column of $U$. We define $\succ_U$ to be the set of order relations

$$j_1 \succ_U j_2 \Leftrightarrow j_1 \succ_k j_2,\ \forall k = 1, \ldots, K.$$

This is an intersection order, which we defined in Section 2.3, so we write

$$h(U) = \bigcap_{k=1}^{K} h(U_{:,k}) \tag{9}$$

as $h(U)$ contains just the order relations shared by all $h(U_{:,k}),\ k = 1, \ldots, K$.

For any poset $h \in \mathcal{H}_\mathcal{M}$ there is a set of $U$ matrices that represent it, so if we parameterise $h$ using $U$ we can represent any poset.

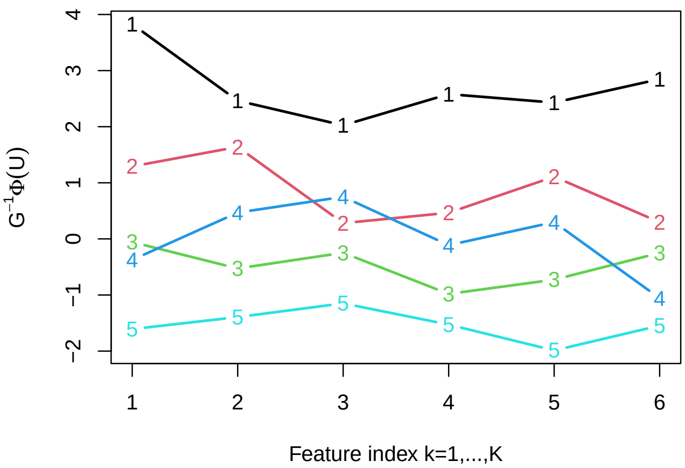



Figure 7: A $U$-matrix with $M = 5$ rows and $K = 6$ columns representing the partial order at left in Figure 1. Each path plots the sequence of values $G^{-1}(\Phi(U_{j,k})), k = 1, \ldots, K$ in a row of feature values for choice $j \in \mathcal{M}$. The path for choice 1 lies entirely above that for choice 2, so $1 \succ_h 2$ in Figure 1. However, the path for choice 4 intersects the path for choice 2, so these choices are unordered in Figure 1

.

**Lemma 2** *If $K \geq \lfloor M/2 \rfloor$ then there is $U \in R^{M \times K}$ such that $h(U) = h$.*

**Proof** In fact the set of $U$ matrices that map to $h$ contains an open set in $R^{M \times K}$, so it has non-zero measure as we now show. For any poset $h$ on $M$ elements and any $K \geq \lfloor M/2 \rfloor$ there exists a set $\{\ell^{(1)}, \ldots, \ell^{(K)}\}$ of complete orders $\ell^{(k)} = (\mathcal{M}, \succ_k),\ k \in [K]$, which intersect to give $h$ (Hiraguchi, 1951). The set $\mathcal{W}_{M,k} = \{U_{:,k} \in \mathbb{R}^M : h(U_{:,k}) = \ell^{(k)}\}$ has infinite volume measure (the constraint is just $U_{1^{(k)},k} > U_{\ell_2^{(k)},k} > \cdots > U_{\ell_M^{(k)},k}$, where $\ell_{1:M}^{(k)}$ is the list representation of $\ell^{(k)}$), so the set $\mathcal{W}_{M,[K]} = \mathcal{W}_{M,1} \times \mathcal{W}_{M,2} \times \cdots \times \mathcal{W}_{M,K}$ also has infinite volume measure in $R^{M \times K}$ and finally $h(U) = h$ for all $U \in \mathcal{W}_{M,[K]}$. ■

### 4.3 Adding Covariates to the poset parameterisation

This latent variable setup makes it straightforward to add covariates. These are covariates on the items. For example if the items being ranked are people, and the partial order is a social hierarchy, then we might have a covariate for the wealth of each person in $\mathcal{M}$. For $j \in \mathcal{M}$, let $x_j = (x_{j,1}, \ldots, x_{j,r})$ be a vector of $r$ covariates associated with object $j$, and let $X$ be the $M \times r$ design matrix of covariate values. Let $\beta \in \mathbb{R}^r$ be a vector of effects and set $\alpha = X\beta$ so $\alpha$ is an $M \times 1$ vector. Take

$$\eta = U + \alpha 1_K^T, \tag{10}$$

where $\alpha 1_K^T$ is an outer product of $\alpha$ with a vector of $K$ ones. The partial order is computed from the shifted weights, $h = h(\eta)$. The rule is now $(\mathcal{M}, \succ_\eta) = h(\eta)$ with $h(\eta) = \cap_k h(\eta_{:,k})$ or equivalently

$$\begin{aligned} j_1 \succ_k j_2 &\Leftrightarrow \eta_{j_1,k} > \eta_{j_2,k} \\ j_1 \succ_\eta j_2 &\Leftrightarrow j_1 \succ_k j_2,\ \forall k = 1, \dots, K. \end{aligned} \tag{11}$$

The $\eta$ matrix replaces the $U$ matrix. Each row is now $\eta_{j,:} = U_{j,:} + \alpha_j 1_K^T$, so a large positive effect $\alpha_j = x_j^T \beta$ means the covariates $x_j$ of item $j \in \mathcal{M}$ are associated with a higher position for $j$ in the partial order. Why does this work? Looking at $\eta_{j,:}$ we see that all elements in the $j$'th row of $U$ are shifted up by $\alpha_j$; that tends to make $\eta_{j,k} > \eta_{j',k}$ more likely. Looking at the rule defining $h(\eta)$ in (11), this tends to move $j$ up in the partial order. We don't include an intercept among the covariates as that duplicates the degree of freedom corresponding to the mean of $U_{j,:}$.

When we have covariates we take as a prior for $\beta$ the $r$-component standard normal distribution $\beta \sim N(0, I_r)$. This has the right sort of scale relative to the variation in $U$ we specify in the next section.

### 4.4 A prior for partial orders from a prior on the latent variables

In Appendix B we review the Plackett-Luce model for rankings, which will be a special case of the partial order models we write down. Readers may find it useful to review that material before continuing.

We get a prior over $h \in \mathcal{H}_\mathcal{M}$ by taking a prior over $\eta$. In order to control $d(h(\eta))$, Nicholls et al. (2025) take the rows of the $\eta$-matrix to be correlated multivariate normal variables, $\eta_{j,:} \sim N(\alpha_j 1_K, \Sigma_\rho)$ independent for $j \in \mathcal{M}$. The covariance matrix $\Sigma_\rho$ has a constant unit diagonal $(\Sigma_\rho)_{k,k} = 1$ and constant off diagonal $(\Sigma_\rho)_{k,k'} = \rho$ for $\rho \in [0, 1)$ and $k \neq k'$, so

$$\Sigma_\rho = \begin{bmatrix} 1 & \rho & . & . & . & \rho \\ \rho & 1 & \rho & . & . & . \\ . & \rho & \ddots & \ddots & & . \\ . & & \ddots & \ddots & \rho & . \\ . & & & \rho & 1 & \rho \\ \rho & . & . & . & \rho & 1 \end{bmatrix}.$$

The correlation parameter $\rho$ is positive and controls the typical depth of $h(\eta)$. When $\rho$ is close to one, the values $\eta_{j,k}$ don't vary much with $k$. The rows of $\eta$, are the different "paths" in Figure 7 and these tend to be quite flat when $\rho$ is large. Flat paths are less likely to intersect and that means there are alot of order relations, so the depth is quite large. Going back to the map we defined in (11), the orders $\succ_k$ are all the same and $d(h(\eta))$ is close to $M$. When $\rho$ is small the preference weights $\eta_{j,k}$ and $\eta_{j,k'}$ are nearly independent so the orders $\succ_k$ and $\succ_{k'}$ share fewer order relations. In this case the paths vary up and down alot so they are likely to cross. This gives fewer order relations and lower depth. Nicholls et al. (2025) show using simulation that taking $\rho \sim \text{Beta}(1, 1/6)$ gives a marginal prior distribution for $h$ which is reasonably uninformative of depth, though this depends on $M$ and the application.

In order to get a model for random partial orders reducing to Plackett-Luce as a special case, we modify this prior. Referring to Appendix B, the Plackett-Luce distribution $PL(\alpha, \mathcal{M})$ is a probability distribution over total orders $\ell \in \mathcal{C}_{\mathcal{M}}$ with "skill scores" $\alpha$.

We use a copula construction in order to retain control over the depth distribution. Let $G^{-1}(g) = -\log(-\log(g))$ be the inverse CDF of a standard Gumbel random variable, and let $\Phi$ be the CDF of a standard normal.

**Theorem 3** *(Partial Order Model) For $\alpha$ and $\Sigma_\rho$ defined above, if we take*

$$U_{j,:} \sim N(0, \Sigma_\rho), \quad \text{independent for each } j \in \mathcal{M}, \tag{12}$$
$$\eta_{j,:} = G^{-1}(\Phi(U_{j,:})) + \alpha_j 1_K^T, \quad \text{and} \tag{13}$$
$$h = h(\eta(U, \beta)), \tag{14}$$

*then $h(\eta_{:,k}) \sim PL(\alpha; \mathcal{M})$ for each $k = 1, \ldots, K$. If $K = 1$ then $h(\eta) \sim PL(\alpha; \mathcal{M})$.*

**Proof** The CDF of a standard normal $\Phi$ is applied to each element of $U_{j,:}$, and $U_{j,k} \sim N(0,1)$ marginally so $\Phi(U_{j,:})$ is a vector of correlated uniform random variables. Applying the inverse CDF of the Gumbel distribution to each element of this vector gives a vector of correlated standard Gumbel random variables, which we shift by $\alpha$ to get $\eta_{j,k} \sim \text{Gumbel}(\alpha_j),\ k = 1, \ldots, K$. They are are independent for each $j \in \mathcal{M}$, so $h(\eta_{:,k})$ is a complete order distributed as $h(\eta_{:,k}) \sim PL(\alpha; \mathcal{M})$ by Theorem 7. If $K = 1$ then the whole partial order $h(\eta(U, \beta)) = h(\eta_{:,1})$ is a complete order with the same distribution. ■

Equations (12) and (13) determine a prior distribution for $h \in \mathcal{H}_S$ for each choice set $S \in \mathcal{B}_{\mathcal{M}}$ with $m = |S|$. Let

$$\eta(U, \beta) = G^{-1}(\Phi(U)) + X\beta\, 1_K^T$$

be the $m \times K$ matrix with rows $\eta_{j,:},\ j \in S$. All operators are applied element by element to matrix arguments. We write $U \sim N(0_{Km}, I_m \otimes \Sigma_\rho)$ for the joint distribution of all elements of $U$ (taken as a vector $(U_{1,1}, \ldots, U_{1,K}, U_{2,1}, \ldots, U_{m,K})$ and using the Kroneker product, so that $I_m \otimes \Sigma_\rho$ is block diagonal). The random partial order $h(\eta(U, \beta))$ has prior distribution

$$\pi_{\mathcal{M}}(h|\rho, \beta) = E_U(\mathbb{I}_{h(\eta(U,\beta))=h}), \quad h \in \mathcal{H}_S. \tag{15}$$

We just write this down to be explicit. We would never estimate this or try to use $h$ itself as the parameter. It is easier to just work with $U$ and $\beta$ and their priors. We can easily compute $h = h(\eta(U, \beta))$ when we come to evaluate the likelihood.

### 4.5 Properties of the latent variable prior for posets

Every partial order in $\mathcal{H}_{\mathcal{M}}$ has non-zero prior probability in the model in (15). The prior doesnt "rule out" any poset.

**Corollary 4** *If $K \geq \lfloor M/2 \rfloor$, then $\pi_S(h|\rho, \beta) > 0$ for any $\rho, \beta$, $h \in \mathcal{H}_S$ and $S \in \mathcal{B}_{\mathcal{M}}$.*

**Proof** If $m = |S|$ and $S \in \mathcal{B}_{\mathcal{M}}$, then $m \le M$, so $K \ge \lfloor m/2 \rfloor$. By Lemma 2 there exists a open set $\mathcal{W}^{(\eta)}_{m,[K]}$ with infinite volume measure in $R^{m\times K}$ such that $h(\eta) = h$ for all $\eta \in \mathcal{W}^{(\eta)}_{m,[K]}$. Given $\eta$ and $\beta$, we invert $\eta(U,\beta)$ to get $U(\eta,\beta) = \Phi^{-1}(G(\eta - X\beta 1_K^T))$. At fixed $\beta$ the bijection $\eta(U,\beta)$ is a continuous function mapping $R^{m\times K} \to R^{m\times K}$, so the preimage $\mathcal{W}^{(U)}_{m,[K]} = U(\mathcal{W}^{(\eta)}_{m,[K]},\beta)$ is an open set in $R^{m\times K}$. Now $U \in \mathcal{W}^{(U)}_{m,[K]} \Rightarrow h(U,\beta) = h$, so

$$E_U(\mathbb{I}_{h(\eta(U,\beta))=h}) \ge E_U(\mathbb{I}_{U\in\mathcal{W}^{(U)}_{m,[K]}}) > 0$$

as $N(0_{Km}, I_m \otimes \Sigma_\rho)$ puts non-zero probability mass on every open set in $R^{m\times K}$. ■

The latent variable setup for the partial order prior ensures that the family of prior distributions $\pi_S(\cdot|\rho,\beta),\ S \in \mathcal{B}_{\mathcal{M}}$ is marginally consistent. Following Winkler (1985), Nicholls et al. (2025) show this for a similar timeseries model for random partial orders.

**Corollary 5** *The family of prior distributions $\pi_S(\cdot|\rho,\beta),\ S \in \mathcal{B}_{\mathcal{M}}$, is marginally consistent, that is if $h \sim \pi_{\mathcal{M}}(\cdot|\rho,\beta)$, then $h[S] \sim \pi_S(\cdot|\rho,\beta)$.*

**Proof** For a $M \times K$ matrix $X$ let $X_{S,:}$ denote the sub-matrix with rows $X_{j,:},\ j \in S$. Since the maps are all applied element by element, $h[S] = h(\eta(U,\beta)_{S,:})$, and

$$\eta(U,\beta)_{S,:} = G^{-1}(\Phi(U_{S,:})) + X_{S,:}\beta\, 1_K^T.$$

As $U_{j,:} \sim N(0,\Sigma_\rho), j \in \mathcal{M}$ are iid over $j$, so are $U_{j,:},\ j \in S$ (no dependence on index $j \in S$) so these are just the random variables and maps defining $\pi_S(\cdot|\rho,\beta)$. ■

Corollary 5 gives marginal consistency for a prior distribution over $h$, which is a *parameter* of the observation model. When we write down the posterior for $U, \beta$ (or in other words $h$) we can drop from the analysis parameters for actors not present in any choice set, just as we did for PL in (24) in Appendix B. In contrast, Theorem 8 in Appendix B gives marginal consistency for *observations* (that is, list data) drawn using the Plackett-Luce model, so we called this property context independence there. As we saw in Section 3, we do not have context independence for lists generated from partial orders.

### 4.6 Posterior for latent variable model for posets

If the data are realised directly on the choice sets $S_i,\ i = 1,\dots,N$, with $\mathcal{M} = \cup_i S_i$ then the posterior parameters are $p \in [0,1]$, $\beta \in \mathbb{R}^p,\ \rho \in [0,1)$ and $U = (U_{j,:})_{j\in\mathcal{M}},\ U \in \mathbb{R}^{M\times K}$. The posterior is

$$\pi_{S_{1:N}}(\rho,U,\beta,p|Y) \propto \pi_R(\rho)\pi_B(\beta)\pi_P(p)\pi(U|\rho)p_{S_{1:N}}(Y|h(\eta(u,\beta)),p), \tag{16}$$
$$\propto \pi_R(\rho)\pi_B(\beta)\pi_P(p)\left[\prod_{j\in\mathcal{M}} N(U_{j,:};0_K,\Sigma_\rho)\right] \times \left[\prod_{i=1}^{N} p_{S_i}(Y_i|h(\eta(U_{S_i,:},\beta)),p)\right].$$

If we have samples $\rho^{(t)}, u^{(t)}, \beta^{(t)}, p^{(t)},\ t = 1,\dots,T$ distributed according to $\pi_{S_{1:N}}(\rho,U,\beta,p|Y)$, and we want samples from the marginal posterior over partial orders, then we simply set $h^{(t)} = h(\eta(U^{(t)},\beta^{(t)})),\ t = 1,\dots,T$.

In the analysis above, the number of columns $K$ is a fixed hyper-parameter of the prior over partial orders. It is straightforward to estimate $K$ and work with

$$\pi_{S_{1:N}}(K, \rho, U, \beta, p|Y) \propto \pi_K(K)\pi_R(\rho)\pi_B(\beta)\pi_P(p)\pi(U|\rho, K)p_{S_{1:N}}(Y|h(\eta(U, \beta)), p), \quad (17)$$

where $\pi_K(K)$ is a prior for $K$ (geometric with mean chosen so that $\pi_K(K \geq M/2)$ isn't too small). The dimension of the $M \times K$ matrix $u$ is now random. This is handled in the inference using a relatively straightforward application of reversible-jump MCMC.

## 5. Example of Bayesian inference for a partial order

In this section we give a small example of Bayesian inference for a poset. An example like this is given in Nicholls and Muir Watt (2011). A similar example (in the setting of a time series model for an evolving poset) is given in Nicholls et al. (2025) and further examples can be found in Jiang and Nicholls (2021) and Jiang et al. (2023).

In this example we have $N = 15$ witness lists recorded for $M = 16$ bishops who witnessed $N$ legal documents in the year 1135. The universe of choices $\mathcal{M} = \{1, \ldots, 16\}$ is just the set of bishops present in the lists. The choice sets vary from list to list (in this context the choice set is just the set of bishops who showed up to witness a given legal document). If $Y_i$ is a witness list (so, an ordered set) then we know $S_i$ is an unordered set with the same elements as $Y_i$. The lists are given in Appendix D. The aim of the analysis is to reconstruct the unknown true social hierarchy/poset that constrained the order of the bishops in witness lists when they witnessed legal documents.

Referring to the posterior in (16) we can assume that all the bishops appear in at least one choice set so no redundant rows of $U$ need to be integrated out using marginal consistency of the prior, as they all play a role in the likelihood. Also, in this example we will ignore the covariate data we have for these bishops (equivalently fix $\beta = 0_r$ and take $\eta = U$). We do actually have a covariate for the bishops. It is a categorical covariate $x_j \in \{1, \ldots, M\}$ indicating "seniority": the longest serving bishop has seniority equal 1 and the most recent appointment has seniority $M$. In theor time series analysis, Nicholls et al. (2025) show that seniority is informative of position in the status hierarchy. However, since we are not making a time series analysis, but just taking data from a single year, the seniority of a bishop is the same in every list in which they appear, so the effect from $U$ and the effect from $X\beta$ are not identifiable (bishops and covariate levels are one to one). We can't distinguish an effect due to bishop (expressed in $U_{j,:}$) from an effect due to their seniority (expressed in $x_j\beta$).

We therefore target the posterior

$$\pi_{S_{1:N}}(\rho, U, p|Y) \propto \pi_R(\rho)\pi_P(p)\pi(U|\rho)\, p_{S_{1:N}}(Y|h(U), p).$$

The prior for $\rho$ is Beta$(1, 1/6)$ as this gives a prior depth distribution which is reasonably flat. The prior for $U$ is

$$\pi(U|\rho) = \prod_{j \in \mathcal{M}} N(U_{j,:}; 0_K, \Sigma_\rho).$$

The prior for the queue jumping probability/noise effect $p$ is Beta$(1, 9)$ so the prior mean for $p$ is 0.1. The likelihood, $p_{S_{1:N}}(Y|h(U), p)$ is given in (6) using the queue jumping likelihood set out in Section 3.6 and in particular, below (8).

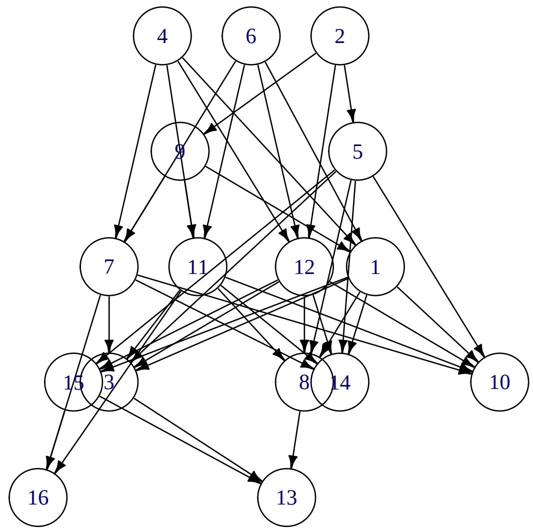



Figure 8: Posterior consensus poset thresholded at $\xi = 0.5$ posterior probability. To retrieve the bishop names see Appendix D.

We take $K = \lfloor M/2 \rfloor = 8$ fixed so we can represent any poset/social hierarchy on the $M = 16$ bishops. We don't know the dimension of the true poset. Experiments in Nicholls et al. (2025) show that results are quite insensitive to the choice of $K$ and in fact taking $K = 2$ looses little in the quality of reconstructed posets. This is presumably because the dimension of the true poset is quite small. If it has a realiser with less than $K$ complete orders then $K$ is "large enough".

We use MCMC to get samples $\rho^{(t)}, U^{(t)}, p^{(t)},\ t = 1, \ldots, 1000$ (subsampling every 10 sweeps over all parameters) distributed according to $\pi_{S_{1:N}}(\rho, U, p|Y)$. This is actually quite a straightforward application of MCMC. For the most part we use simple random walk Metropolis updates. The MCMC mixes quite well because we are using the queue jumping likelihood so there are no hard constraints of the form $Y_i \in \mathcal{L}[h[S_i]]$ for $i = 1, \ldots, N$: for the queue jumping likelihood, the posterior density is never zero as any list can arise from any poset with enough queue jumping events. The run we present here took about an hour on a 6 year old laptop. The effective sample size for the log likelihood, for $p$, $\rho$ and the log of the $U$ prior were 500, 1000, 63 and 34. There might be a case for making a longer run, say 10 times longer.

Once we have the samples our main interest is in the reconstructed poset. We have the problem of summarising the $T$ sampled posets $h^{(t)} = h(U^{(t)}),\ t = 1, \ldots, T$. One natural point estimate is the *consensus* partial order, $\hat{h}(\xi)$. Here $\xi \in [0, 1]$ is a threshold probability. We include a relation $j_1 \succ_{\hat{h}} j_2$ in $\hat{h}$ if it appears in at least $\xi T$ of the sampled posets $h^{(1:T)}$ so the relation is probably present. Experiments with synthetic data, looking at ROC curves, suggest $\xi = 0.5$ is a reasonable threshold.

The consensus poset for the data in Appendix D is given in Figure 8. To historians there are some plausible elements here. Bishop 2 was Henry de Blois, brother of King Stephen. It is not surprising to see high status there. Length of time in office also contributed to status in this context. Bishops 4 (Roger, Bishop of Salisbury) and 6 (John, Bishop of Lisieux) had been in post for around 30 years at this point and were the two longest serving bishops.

## Appendix A. Context Independence

### A.1 Definition of Context Independence

**Definition 6** *(Context-independent preference) Let* $\{p_S(y),\ y \in \mathcal{C}_S\},\ S \in \mathcal{B}_\mathcal{M}$ *be a family of probability distributions over orders and let* $y \sim p_\mathcal{M}(\cdot)$*. If* $y[S] \sim p_S(\cdot)$ *for all* $S \in \mathcal{B}_\mathcal{M}$*, then the family expresses context-independent preference, and otherwise, preference is context-dependent.*

Context independence in the sense of Definition 6 is weaker than context-independence in the sense of the "Luce Axiom of Choice" (Luce, 1977). We return to this in Appendix B.

If preferences are context independent then we must have

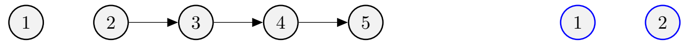
$$p_S(y) = \sum_{z \in \mathcal{C}_\mathcal{M}} p_\mathcal{M}(z) \mathbb{I}_{z(S)=y},\ S \in \mathcal{B}_\mathcal{M}. \tag{18}$$

It further holds that all the marginals are consistent, by countable additivity of $p_\mathcal{M}(\cdot)$, so we can replace $\mathcal{M}$ by $S'$ in (18) for any $S' \supseteq S$ and (18) still holds. This kind of *marginal consistency* but has the special meaning of context-independence for distributions over complete preference orders. It is important to distinguish between marginal consistency in the prior (which the poset model in Section 4.4 posesses) and marginal consistency for the observations (which doesnt hold for the poset observation model in Section 3.6).

Context independence need not hold if we simply write down $p_S(\cdot)$ separately for each $S \in \mathcal{B}_\mathcal{M}$. For example, the Mallows model (Mallows, 1957), the contextual repeated selection (CRS) model (Seshadri et al., 2020) and the partial order model in Section 4.4 do not, in general, satisfy (18). On the other hand, the Plackett-Luce model in Appendix B is context-independent in the sense of Definition 6 and in the stronger sense of the Luce Axiom of Choice (Luce, 1959) in (23). Ragain and Ugander (2018) and Seshadri et al. (2019, 2020) develop models for context-dependent choice, which generalise the Plackett-Luce model. Our approach builds transitivity into a context-dependent preference model whilst transitivity plays no role in Seshadri et al. (2020).

### A.2 Random linear extensions are context dependent

The model in (6) expresses context-dependent preferences. This is illustrated by the example in Figure 9. Suppose $h \in \mathcal{H}_\mathcal{M}$ is the poset $(\{1, \ldots, M\}, \succ_h)$, in which the suborder $(\{2, \ldots, M\}, \succ_h)$ is a complete order $2 \succ_h 3 \succ_h \cdots \succ_h M$, but item 1 has no order relation to any other elements. Now $h$ has $M$ linear extensions (item 1 can go in any position in the complete order, and the order of the rest is fixed), so item 2 is the maximal element in $M-1$ out of $M$ linear extensions. If the choice set is $S = \{1, 2\}$, and the list is realised as a suborder on $S$, then we pick a random linear extension $y' \sim p_\mathcal{M}(\cdot|h)$ and get $\Pr(1 \succ_{y'[S]} 2) = 1/M$. However, the suborder $h[S] = (\{1, 2\}, \succ_h)$ is empty (it has two unordered elements), so it has two linear extensions, $\ell^{(1)}$ and $\ell^{(2)}$, with $1 \succ_{\ell^{(1)}} 2$ and $2 \succ_{\ell^{(2)}} 1$. If the list is realised on $S$, then $y \sim p_S(\cdot|h[S])$ and $\Pr(1 \succ_y 2) = 1/2$, so $y'[S]$ and $y$ do not have the same distribution so preferences are context dependent.




$h = (\mathcal{M}, \succ_h)$ $\qquad$ $h[S] = (S, \succ_h)$

Figure 9: In the partial order (left) on $\mathcal{M} = \{1, 2, 3, 4, 5\}$, 1 is maximal in one of five possible linear extensions. In the suborder on $S = \{1, 2\}$, 1 is maximal on one of two.

## Appendix B. The Plackett-Luce Model

In a Plakett-Luce (PL) model with preference weights $\alpha_{\mathcal{M}} = (\alpha_1, \ldots, \alpha_M)$, the weight for choice $i \in \mathcal{M}$ is $\alpha_i \in \mathbb{R}$ and the observation model for an ordering $y \in \mathcal{C}_S$ of a choice set $S \in \mathcal{B}_{\mathcal{M}}$ with $m$ elements is

$$p_S(y|\alpha_S) = \prod_{i=1}^{m} \frac{e^{\alpha_{y_i}}}{\sum_{i'=i}^{m} e^{\alpha_{y_{i'}}}}, \tag{19}$$

where $\alpha_S = (\alpha_j)_{j \in S}$. We write $y \sim \mathrm{PL}(\alpha_S; S)$. This is another sequential choice model in which the order is built up element by element: we can write

$$p_S(y|\alpha_S) = \prod_{i=1}^{m-1} q_{y_{i:m}}(y_i|\alpha_S), \tag{20}$$

where

$$q_{y_{i:m}}(y_i|\alpha_S) = \frac{e^{\alpha_{y_i}}}{\sum_{i'=i}^{m} e^{\alpha_{y_{i'}}}} \tag{21}$$

is the probability to choose $y_i$ from the choice set $\{y_i, y_{i+1}, \ldots, y_m\}$.

We can add covariates to this model. For $j \in \mathcal{M}$, let $x_j = (x_{j,1}, \ldots, x_{j,p})$ be a vector of $p$ covariates associated with object $j$, and let $X = (x_{j,c})_{j \in \mathcal{M}}^{c \in [p]}$ be the $M \times p$ matrix of all covariates. Let $\beta \in \mathbb{R}^p$ be a vector of effects. If we include an intercept in the covariate vector, then the vector of preference weights in (19) is $\alpha_{\mathcal{M}} = X\beta$.

The generative model for orders in the Plackett-Luce model can be given in terms of latent Gumbel random variables (Yellott, 1977).

**Lemma 7** *(Yellott, 1977) Let $G_j \sim$ Gumbel$(\alpha_j)$ be independent for $j \in \mathcal{M}$, where Gumbel$(\alpha_j)$ is a distribution with CDF $F_{\alpha_j}(g) = \exp(-\exp(-(g - \alpha_j)))$, $g \in \mathbb{R}$. Let $G = (G_1, \ldots, G_M)$ and let $y(G) = (\mathcal{M}, \succ_G)$ be the corresponding complete order for the elements of $G$, that is $j_1 \succ_G j_2 \Leftrightarrow G_{j_1} > G_{j_2}$. Yellott (1977) shows that $y(G) \sim PL(\alpha_{\mathcal{M}}; \mathcal{M})$.*

We can think of $G_j$ as a random grade for object $j$, which is biased by their preference weight or *skill score* $\alpha_j$; the random order $y(G)$ gives the objects ordered by grade.

The PL model is context independent: if $y \sim p_{\mathcal{M}}(\cdot|\alpha_{\mathcal{M}})$ then $y[S] \sim p_S(\cdot|\alpha_S)$. We prove this using Lemma 7. Hunter (2004) demonstrates the required relationship between marginal distributions, (22) below, by direct computation.

**Theorem 8** *(Hunter, 2004) For all $S \in \mathcal{B}_{\mathcal{M}}$ and all $y \in \mathcal{C}_S$*

$$p_S(y|\alpha_S) = \sum_{z \in \mathcal{C}_{\mathcal{M}}} p_{\mathcal{M}}(z|\alpha_{\mathcal{M}}) \, \mathbb{I}_{z[S]=y}. \tag{22}$$

**Proof** This follows from the Gumbel construction. If $G_j \sim \text{Gumbel}(\alpha_j)$ independent for $j = 1, \ldots, M$, then $y(G) \sim \text{PL}(\alpha_{\mathcal{M}}; \mathcal{M})$ with $y(G) \in \mathcal{C}_{\mathcal{M}}$ by Lemma 7. In our notation, $y(G)[S] \in \mathcal{C}_S$ is the $S$-suborder of $y(G)$, so $\Pr(y(G)[S] = y)$ is given by the RHS of (22). Let $G_S = \{G_j : j \in S\}$. Removing elements from $G$ does not change the relative ordering for the elements that remain, so $y(G)[S] = y(G_S)$ with $G_j \sim \text{Gumbel}(\alpha_j),\ j \in S$ jointly independent and so by Lemma 7, $y(G_S) \sim p_S(\cdot|\alpha_S)$, the distribution on the LHS of (22). ■

Context independence in the sense of Definition 6 is a weaker condition than context-independence in the sense of the "Luce Axiom of Choice" (LAC, Luce (1977)). As we build up a list by choosing the sequence of elements one at a time according to the factorisation in (20), the odds of choosing $j_1$ next over $j_2$ in the sequence is the same for every choice set $S$ containing $j_1$ and $j_2$. In the notation of (21) that is

$$\frac{q_S(j_1|\alpha_S)}{q_S(j_2|\alpha_S)} = \frac{q_{\mathcal{M}}(j_1|\alpha_{\mathcal{M}})}{q_{\mathcal{M}}(j_2|\alpha_{\mathcal{M}})}. \tag{23}$$

This is easy to check for $q_S(j|\alpha_S)$ in (21) and in fact the converse is also true: the Plackett-Luce distribution in (19) is the only distribution over orders satisfying the LAC (Luce, 1959). Since LAC $\Rightarrow$ (19) $\Rightarrow$ Lemma 7 $\Rightarrow$ Theorem 8, it follows that LAC implies context independence in the sense of Definition 6. However, there are many marginally consistent families of distributions over orders which satisfy (18) but not (23): for example if we take $p_{\mathcal{M}}(z) = |\mathcal{L}[h]|^{-1}\mathbb{I}_{z\in\mathcal{L}[h]}$ for some $h \in \mathcal{H}_{\mathcal{M}}$ and *define* $p_S(y)$ by (18) then the family $p_S,\ S \in \mathcal{B}_{\mathcal{M}}$ is marginally consistent by construction, but it is not Plackett-Luce.

Theorem 8 is helpful when data $Y = (Y_1, \ldots, Y_N)$ are observed as suborders with $Y_i = Y'_i[S_i]$, and $Y'_i \sim p_{\mathcal{M}}(\cdot|\alpha)$ jointly independent for $i = 1, \ldots, N$ given $\alpha$, as this is the same as generating $\tilde{Y}_i$ on the choice set $S_i$. Suppose the family of priors $\{\pi_{\alpha,S}(\alpha_S),\ \alpha_S \in \mathbb{R}^m\}_{S\in\mathcal{B}_{\mathcal{M}}}$ for $\alpha \in \mathbb{R}^M$ is marginally consistent, so $\alpha \sim \pi_{\alpha,\mathcal{M}}$ implies $\alpha_S \sim \pi_{\alpha,S}$ (for example, when the components of $\alpha$ are a priori independent). We can drop $\alpha_j$ from the analysis if $j$ does not appear in at least one $S_i$. Let $\mathcal{M}' = \cup_i S_i$. The marginal posterior for $\alpha \in \mathbb{R}^{|\mathcal{M}'|}$ is

$$\begin{aligned}\pi_{\alpha,\mathcal{M}'}(\alpha_{\mathcal{M}'}|Y) &= \int \pi_{\alpha,\mathcal{M}}(\alpha_{\mathcal{M}}|Y)d\alpha_{\mathcal{M}\setminus\mathcal{M}'} \\ &\propto \int \pi_\alpha(\alpha_{\mathcal{M}}) \prod_{i=1}^{N} \left[ \sum_{Y'_i\in\mathcal{C}_{\mathcal{M}}} p_{\mathcal{M}}(Y'_i|\alpha)\mathbb{I}_{Y'[S_i]=Y_i} \right] d\alpha_{\mathcal{M}\setminus\mathcal{M}'}\end{aligned}$$

after applying Theorem 8 to do the sum, the $\alpha_j,\ j \in \mathcal{M} \setminus \mathcal{M}'$ are no longer in the product,

$$\propto \int \pi_\alpha(\alpha_{\mathcal{M}})\, d\alpha_{\mathcal{M}\setminus\mathcal{M}'} \prod_{i=1}^{N} p_{S_i}(Y_i|\alpha_{S_i})$$

so by the assumed marginal consistency of the $\alpha$-prior

$$\propto \pi_{\alpha,\mathcal{M}'}(\alpha_{\mathcal{M}'}) \prod_{i=1}^{N} p_{S_i}(Y_i|\alpha_{S_i}). \tag{24}$$

The posterior $\pi_{\alpha,\mathcal{M}'}(\alpha_{\mathcal{M}'}|Y)$ only involves parameters for objects in $\mathcal{M}$ that we have data for, so sampling and estimation will be more efficient than if we had to target $\pi_{\alpha,\mathcal{M}}(\alpha_{\mathcal{M}}|Y)$. Another advantage is that the posterior is robust to the observation model: we don't need to know whether the assessor realised their preference order on the choice set $S$ or made a full ranking on $\mathcal{M}$ and then thinned the list down to the suborder for $S$.

## Appendix C. Sampling partial orders uniformly at random

It isnt so easy to sample posets uniformly at random to illustrate the result of Kleitman and Rothschild (1975) that $d(h) \to 3$ in probability as $M \to \infty$. Muir Watt (2015) gives a sampler which is a Markov chain on DAGs. Posets are hard to sample because if you change one relation (or edge of the transitively closed DAG) it can lead to a cascade of new relations implied by transitivity. Simple operations won't be reversible. It is easier to sample DAGs as you can just randomly turn on and off edges, and the only constraint is to avoid loops (so the graph stays a DAG). The trick is to target a distribution on DAGs which "closes" to the uniform distribution on posets in the sense that if $G \in \mathcal{D}_{\mathcal{M}}$ is a random DAG then its transitive closure is uniform on $\mathcal{H}_{\mathcal{M}}$, that is, $C(G) \sim \tilde{\pi}_{\mathcal{M}}(\cdot)$.

For any poset $h \in \mathcal{H}_{\mathcal{M}}$, how many DAGs close to give $h$? In the notation of Section 2.1 we want the number of elements in the set $\{g \in \mathcal{D}_{\mathcal{M}} : C(g) = h\}$. We can remove from the closure any edge which is not in the reduction, as these edges are all implied by transitivity. If $c(g)$ is the number of edges in $C(g)$ and $r(g)$ is the number of edges in the transitive reduction $R(g)$ then we can switch on and off any of the $c(g) - r(g)$ edges which are in the closure but not the reduction. It follows that there are

$$|\{g \in \mathcal{D}_{\mathcal{M}} : C(g) = h\}| = 2^{\,c(g)-r(g)}$$

DAGs $g$ that close to give $h = C(g)$. Let

$$\tilde{\pi}^{\mathcal{D}}_{\mathcal{M}}(g) = 2^{-(c(g)-r(g))}/|\mathcal{H}_{\mathcal{M}}|.$$

If $G \sim \tilde{\pi}^{\mathcal{D}}_{\mathcal{M}}(\cdot)$ and $H = C(G)$ then

$$\begin{aligned}
\Pr(H = h) &= \sum_{g:C(g)=h} \Pr(G = g) \\
&= \sum_{g:C(g)=h} 2^{-(c(g)-r(g))}/|\mathcal{H}_{\mathcal{M}}| \\
&= \tilde{\pi}_{\mathcal{M}}(h)
\end{aligned}$$

(where $\pi_{\mathcal{M}}(h) = |\mathcal{H}_{\mathcal{M}}|^{-1}$ is the uniform distribution of posets defined in Section 4.1) because $c(g)$ and $r(g)$ dont depend on $g$ as long as $C(g) = h$: every poset with closure $C(g)$ has reduction $R(g)$.

It follows that if we simulate a Markov chain $\tilde{X}_t,\ t \geq 0$ on $\mathcal{D}_{\mathcal{M}}$ targeting $\tilde{\pi}^{\mathcal{D}}_{\mathcal{M}}(\cdot)$ and take $H_t = C(\tilde{X}_t)$ then $H_t,\ t \geq 1$ is a Markov chain targeting $\tilde{\pi}_{\mathcal{M}}(\cdot)$.

A Metropolis Hastings MCMC update for a chain targeting $\tilde{\pi}^{\mathcal{D}}_{\mathcal{M}}(\cdot)$ is given in Figure 10. Since we are working with DAGs we use the notation for directed graphs, so instead of writing $j_1 \succ j_2$ we write $\langle j_1, j_2 \rangle$. The proposal just switches on and off edges in the DAG.

If $\tilde{X}_t = g$ with $g = (\mathcal{M}, E)$ a DAG in $\mathcal{D}_\mathcal{M}$ with vertex set $\mathcal{M}$ and edge set $E$ then $\tilde{X}_{t+1}$ is determined in the following way.

1. Choose a distinct pair of vertices $j_1, j_2 \in \mathcal{M}$ uniformly at random.
   (a) If $\langle j_1, j_2 \rangle \in E$ set $E' \leftarrow E \setminus \{\langle j_1, j_2 \rangle\}$.
   (b) If $\langle j_1, j_2 \rangle \notin E$ set $E' \leftarrow E \cup \{\langle j_1, j_2 \rangle\}$.
2. Let $g' = (\mathcal{M}, E')$. With probability

$$\alpha(g'|g) = \min\left\{1,\ \frac{\tilde{\pi}^{\mathcal{D}}_{\mathcal{M}}(g')}{\tilde{\pi}^{\mathcal{D}}_{\mathcal{M}}(g)}\right\} \mathbb{I}_{g' \in \mathcal{D}_\mathcal{M}}$$

   set $\tilde{X}_{t+1} = g'$ and otherwise $\tilde{X}_{t+1} = g$.

Figure 10: Metropolis-Hastings MCMC algorithm targeting $\tilde{\pi}^{\mathcal{D}}_{\mathcal{M}}(g) = 2^{-(c(g)-r(g))}/|\mathcal{H}_\mathcal{M}|$. Setting $H_t = C(\tilde{X}_t)$, the transitive closure of the DAG $\tilde{X}_t$, gives a Markov chain $H_t,\ t \geq 0$ converging in distribution to $\tilde{\pi}_\mathcal{M}(\cdot)$, the uniform distribution on $\mathcal{H}_\mathcal{M}$.

This might create a loop so we reject if the proposed graph is not a DAG. This is enforced by the indicator $\mathbb{I}_{g' \in \mathcal{D}_\mathcal{M}}$ in the acceptance probability. If the current state is $g$ and we propose $g'$ then the probability $q(g'|g)$ to propose $g'$ from $g$ is the equal to the probability $q(g|g')$ to propose $g$ from $g'$ so the proposal probabilities cancel.

## Appendix D. Data for example analysis

The following data are taken from the analysis in Nicholls et al. (2025) and ultimately come from Sharpe et al. (2014).

We give here the preference orders. These are witness lists from old legal documents. We give each list separated by `/`.

```
2 4 / 5 15 / 2 12 / 2 11 14 / 9 14 / 2 4 / 12  3 / 2 15 / 6 8 /
4  1  3 13 / 4  2  7 10 / 6 8 / 6 14 / 6 10 / 9 15 16
```

The mapping from these bishop labels to actual historical bishop names is given below.

```
1, Simon, Bishop of Worcester
2, Henry, de Blois, Bishop of Winchester
3, Bernard, Bishop of St David's
4, Roger, Bishop of Salisbury
5, Everard, bishop of Norwich
6, John, Bishop of Lisieux
7, Alexander, Bishop of Lincoln
8, Ouen, Bishop of Evreux
9, Nigel, Bishop of Ely
```

```
10, Geoffrey, Rufus, Bishop of Durham
11, Algar, Bishop of Coutances
12, Seffrid, Bishop of Chichester
13, Roger, de Clinton, Bishop of Chester
14, Adelulf, Bishop of Carlisle
15, Richard, de Beaufeu, Bishop of Avranches
16, Robert, Bishop of Exeter
```

## References

Niko Beerenwinkel, Nicholas Eriksson, and Bernd Sturmfels. Conjunctive bayesian networks. *Bernoulli*, 13(4):893–909, 2007.

G. Brightwell. *Surveys in Combinatorics*, volume 187 of *London Mathematical Society Lecture Note Series*, chapter Models of random partial orders, pages 53–83. Cambridge Univeristy Press, 1993.

Graham Brightwell and Peter Winkler. Counting linear extensions. *Order*, 8(3):225–242, 1991.

Toshio Hiraguchi. On the dimension of partially ordered sets. *The science reports of the Kanazawa University*, 1:77–94, 1951.

David R. Hunter. MM algorithms for generalized Bradley-Terry models. *The Annals of Statistics*, 32(1):384 – 406, 2004. doi: 10.1214/aos/1079120141. URL `https://doi.org/10.1214/aos/1079120141`.

Chuxuan Jiang and Geoff K. Nicholls. Bayesian inference for partial orders with ties from ranking data with a placket-luce distribution centred on random linear extensions, 2021.

Chuxuan Jiang, Geoff K. Nicholls, and Jeong Eun Lee. Bayesian inference for vertex-series-parallel partial orders. In Robin J. Evans and Ilya Shpitser, editors, *Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence*, volume 216 of *Proceedings of Machine Learning Research*, pages 995–1004. PMLR, 31 Jul–04 Aug 2023. URL `https://proceedings.mlr.press/v216/jiang23b.html`.

Kustaa Kangas, Teemu Hankala, Teppo Niinimäki, and Mikko Koivisto. Counting linear extensions of sparse posets. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 603–609. AAAI Press, 2016. ISBN 9781577357704. URL `https://www.cs.helsinki.fi/u/jwkangas/lecount/`.

Daniel J Kleitman and Bruce L Rothschild. Asymptotic enumeration of partial orders on a finite set. *Transactions of the American Mathematical Society*, 205:205–220, 1975.

Donald E Knuth and Jayme L Szwarcfiter. A structured program to generate all topological sorting arrangements. *Information Processing Letters*, 2(6):153–157, 1974.

R Duncan Luce. On the possible psychophysical laws. *Psychological review*, 66(2):81, 1959.

R.Duncan Luce. The choice axiom after twenty years. *Journal of Mathematical Psychology*, 15(3):215–233, 1977. ISSN 0022-2496. doi: https://doi.org/10.1016/0022-2496(77)90032-3. URL https://www.sciencedirect.com/science/article/pii/0022249677900323.

Colin L Mallows. Non-null ranking models. i. *Biometrika*, 44(1/2):114–130, 1957.

H. Mannila. Finding total and partial orders from data for seriation. In J.-F. Boucault, editor, *Discovery Science*, volume 5255 of *LNAI*, pages 16–25, Berlin Heidelberg, 2008. Springer-Verlag.

A. Muir Watt. *Inference for partial orders from random linear extensions*. PhD thesis, University of Oxford, 2015. URL https://ora.ox.ac.uk/objects/uuid:c4bfa987-d5e0-4f11-b714-4a137e0464b4.

Geoff K Nicholls and Alexis Muir Watt. Partial order models for episcopal social status in 12th century England. *IWSM 2011*, page 437, 2011.

Geoff K. Nicholls, Jeong Eun Lee, Nicholas Karn, David Johnson, Rukuang Huang, and Alexis Muir-Watt. Bayesian inference for partial orders from random linear extensions: Power relations from 12th century royal acta. *The Annals of Applied Statistics*, 19(2):1663 – 1690, 2025. doi: 10.1214/24-AOAS2002. URL https://doi.org/10.1214/24-AOAS2002.

Stephen Ragain and Johan Ugander. Choosing to rank, 2018. URL https://arxiv.org/abs/1809.05139. ARXIV.1809.05139.

Thomas Sakoparnig and Niko Beerenwinkel. Efficient sampling for Bayesian inference of conjunctive Bayesian networks. *Bioinformatics*, 28(18):2318–2324, 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts433. URL https://doi.org/10.1093/bioinformatics/bts433.

Arjun Seshadri, Alex Peysakhovich, and Johan Ugander. Discovering context effects from raw choice data. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5660–5669. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/seshadri19a.html.

Arjun Seshadri, Stephen Ragain, and Johan Ugander. Learning rich rankings. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9435–9446. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/6affee954d76859baa2800e1c49e2c5d-Paper.pdf.

R. Sharpe, D. Carpenter, H. Doherty, M. Hagger, and N. Karn. The charters of William II and Henry I. Online: Last accessed 27 October 2022, 2014.

Peter Winkler. Random orders. *Order*, 1(4):317–331, 1985.

John I. Yellott. The relationship between luce's choice axiom, thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2):109–144, 1977. ISSN 0022-2496. doi: https://doi.org/10.1016/0022-2496(77)90026-8. URL `https://www.sciencedirect.com/science/article/pii/0022249677900268`.