Partial Order Hierarchies and Rank-Order Data

Geoff
Department of Statistics
University of Oxford, UK

Jessie J.
Department of Statistics
University of Oxford, UK

Kiwi Jessie
Mathematical Sciences
University of Southampton, UK

Kate
Department of Statistics
University of Auckland, NZ

Abstract

In rank-order data, assessors give preference orders over choice sets. An order is registered as a list giving the elements of its choice set in order from best to worst. 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 centre the random lists on a partial order. Lists are random linear extensions of a partial order or linear extensions observed with noise. We give a new hierarchical model for partial orders to handle list data which come in labeled groups. The model reduces to a Plackett-Luce model when the partial order dimension is set equal one and can be used to cluster unlabeled list data. We carry out Bayesian inference for the poset hierarchy using MCMC. Evaluation of the likelihood costs #P so applications are restricted to choice sets of up to 20 elements. Following earlier work by other authors, we give a hierarchical model for Vertex-Series-Parallel posets which admits scalable inference.

Keywords: Bayesian Inference, Partial Orders, Linear Extensions, Hierarchical Model, Clustering.

1. Modeling random orders with partial orders

Let $\mathcal{M} = [M],\ [M] = \{1, \ldots, M\}$ 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.

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

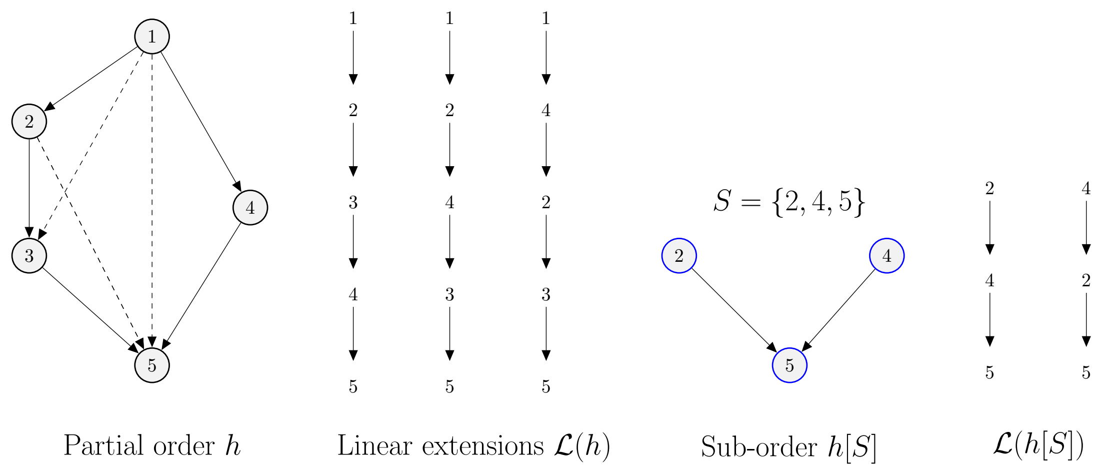

Figure 1: The example partial order h (left) on $\mathcal{M} = \{1, 2, 3, 4, 5\}$ has three linear extensions (left-centre). Dashed lines in h are relations implied by transitivity. Its suborder on $S = \{2, 4, 5\}$ (right-centre) has two linear extensions (right).

$j_2 \succ_h j_1$. 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 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 $j_1, j_2 \in S$. Let $\mathcal{H}_S$ be the set of all posets on S. When we write down models for random posets $(S, \succ_h)$, we refer to them as models for random partial orders, as the choice sets are always fixed, only the partial order component is random.

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.

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 $y_1 \succ_y y_2 \succ_y \cdots \succ_y y_m$. 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.

Definition 1 *(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 1 is weaker than context-independence in the sense of the “Luce Axiom of Choice” (Luce, 1977). We return to this in Section 2.

If preferences are context independent then we must have

$$p_S(y) = \sum_{z \in \mathcal{C}_{\mathcal{M}}} p_{\mathcal{M}}(z) \mathbb{I}_{z(S)=y}, \; S \in \mathcal{B}_{\mathcal{M}}. \tag{1}$$

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 (1) for any $S' \supseteq S$ and (1) still holds. This kind of *marginal consistency* appears in other forms below but has the special meaning of context-independence for distributions over complete preference orders.

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 this section do not, in general, satisfy (1). On the other hand, the Plackett-Luce model in the next section is context-independent in the sense of Definition 1 and in the equivalent sense of the Luce Axiom of Choice (Luce, 1959). 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).

We focus on the case where the data are realised on a choice set and return to the case of thinned data later. We model the observed orders $y \in \mathcal{C}_S$ as random linear extensions respecting partially ordered preferences $h \in \mathcal{H}_S$. 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 $\succ_h$. Denote by $\mathcal{L}[h]$ the set of all linear extensions of h. 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. In our 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 $h = (S, \succ_h)$. The likelihood for h is then

$$p_S(y|h) = |\mathcal{L}[h]|^{-1} \mathbb{I}_{y \in \mathcal{L}[h]}. \tag{2}$$

This model is motivated in Nicholls et al. (2022) from a stochastic queue process in which the random linear extension is a draw from the equilibrium of a stochastic process on $\mathcal{L}[h]$. However, it 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 $p_S(y|h) = p_S(y'|h)$. We will later modify this generative model for lists to allow some “errors” in the recorded lists.

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 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.

An observed order $y \sim p(\cdot|h)$ can be realised in a sequential way, building up the list one element at a time. Models of this kind are called *repeated selection* models. Many models for preference orders are sequential choice models: our partial order models, the Plackett-Luce model, the CRS model and some but not all Mallows models (depending on the choice of distance measures taken over $\mathcal{C}_\mathcal{M}$). Let $y_{i:m} = (y_i, \dots, y_m)$ 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

$$p_S(y|h) = \prod_{i=1}^{m-1} q_{y_{i:m}}(y_i|h[y_{i:m}]), \tag{3}$$

where

$$q_{y_{i:m}}(y_i|h[y_{i:m}]) = \frac{|\mathcal{L}_{y_i}[h[y_{i:m}]]|}{|\mathcal{L}[h[y_{i:m}]]|} \tag{4}$$

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 (3) is equal to $p_S(y|h)$ in (2) because $|\mathcal{L}_{y_i}[h[y_{i:m}]]| = |\mathcal{L}[h[y_{i+1:m}]]|$ (the number of LEs headed by y_i is just the number of LEs in the suborder that remains after y_i is removed), so $p_S(\cdot|h)$ is a repeated selection model.

The likelihood $p_S(y|h)$ can be parameterised as a Pairwise choice Markov chain (PCMC, Ragain and Ugander (2016)). In PCMC models, an order y is built sequentially by drawing y_1 from the equilibrium of a stochastic process on S with $m \times m$ rate matrix Q and repeating for $S \setminus \{y_1\}$ and so on. Taking $Q_{j_1,j_2}(\alpha) \propto |\mathcal{L}_{j_2}[h]|^{1-\alpha}|\mathcal{L}_{j_2}[h]|^{-\alpha}$ gives detailed balance, $q_S(j_1|h)Q_{j_1,j_2}(\alpha) = q_S(j_2|h)Q_{j_2,j_1}(\alpha)$, for any fixed $\alpha \in [0,1]$ so the i'th element y_i in y is selected with probability $q_{y_{i:m}}(y_i|h[y_{i:m}])$. This could be used to embed an underlying partial order structure in PCMC and thereby build transitivity into the PCMC setup, or conversely, to relax the observation model in (5), by taking a prior on entries of Q which is concentrated on $Q(\alpha)$. We have not pursued this generalisation.

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 $Y = (Y_1, \dots, Y_N)$, 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, $Y_i = (Y_{i,1}, \dots, Y_{i,m_i})$. The assessor's preferences over S_i are determined by the suborder $h[S_i] = (S_i, \succ_h)$. Assuming the lists are independent given h, the likelihood is

$$p_{S_{1:N}}(Y|h) = \prod_{i=1}^{N} p_{S_i}(Y_i|h[S_i]), \tag{5}$$

where $p_{S_i}(Y_i|h[S_i]) = |\mathcal{L}[h[S_i]]|^{-1}$ per Equation (2).

The model in (5) expresses context-dependent preferences. This is illustrated by the example in Figure 2. Suppose $h \in \mathcal{H}_\mathcal{M}$ is the poset $(\{1, \dots, M\}, \succ_h)$, in which the suborder $(\{2, \dots, M\}, \succ_h)$ 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

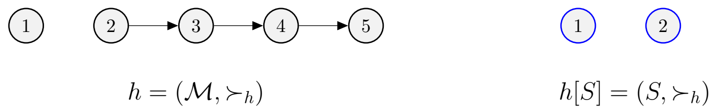

Figure 2: In the partial order (left) on $\mathcal{M} = \{1, 2, 3, 4, 5\}$, 1 is maximal in one of five possible linear extensions. In the suborder on $S = \{1, 2\}$, 1 is maximal on one of two.

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.

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 (3) at k.

If we wish to carry out Bayesian inference for h, we need a prior for h. Before giving this prior we review the Plackett-Luce model for rankings, which will be a special case of the partial order models we write down.

2. 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

$$p_S(y|\alpha_S) = \prod_{i=1}^{m} \frac{e^{\alpha_{y_i}}}{\sum_{i'=i}^{m} e^{\alpha_{y_{i'}}}}, \tag{6}$$

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

$$p_S(y|\alpha_S) = \prod_{i=1}^{m-1} q_{y_{i:m}}(y_i|\alpha_S), \tag{7}$$

where

$$q_{y_{i:m}}(y_i|\alpha_S) = \frac{e^{\alpha_{y_i}}}{\sum_{i'=i}^{m} e^{\alpha_{y_{i'}}}} \tag{8}$$

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 (6) 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 2 *(Yellott, 1977) Let $G_j \sim$ Gumbel(α_j) be independent for $j \in \mathcal{M}$, where Gumbel(α_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 α_j; the random preference 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 2. Hunter (2004) demonstrates the required relationship between marginal distributions, (9) below, by direct computation.

Theorem 3 *(Hunter, 2004) For all $S \in \mathcal{B}_{\mathcal{M}}$ and all $y \in \mathcal{C}_S$*

$$p_S(y|\alpha_S) = \sum_{z \in \mathcal{C}_{\mathcal{M}}} p_{\mathcal{M}}(z|\alpha_{\mathcal{M}})\, \mathbb{I}_{z[S]=y}. \tag{9}$$

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 2. 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 (9). 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 2, $y(G_S) \sim p_S(\cdot|\alpha_S)$, the distribution on the LHS of (9). ■

Context independence in the sense of Definition 1 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 (7), 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 (8) that is

$$\frac{q_S(j_1|\alpha_S)}{qF_S(j_2|\alpha_S)} = \frac{q_{\mathcal{M}}(j_1|\alpha_{\mathcal{M}})}{q_{\mathcal{M}}(j_2|\alpha_{\mathcal{M}})}. \tag{10}$$

This is easy to check for $q_S(j|\alpha_S)$ in (8) and in fact the converse is also true: the Plackett-Luce distribution in (6) is the only distribution over orders satisfying the LAC (Luce, 1959). Since LAC $\Rightarrow$ (6) $\Rightarrow$ Lemma 2 $\Rightarrow$ Theorem 3, it follows that LAC implies context independence in the sense of Definition 1. However, there are many marginally consistent families of distributions over orders which satisfy (1) but not (10): 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 (1) then the family $p_S,\ S \in \mathcal{B}_{\mathcal{M}}$ is marginally consistent by construction, but it is not Plackett-Luce.

Theorem 3 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 α, 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 α are a priori independent). We can drop α_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 3 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 α-prior

$$\propto \pi_{\alpha,\mathcal{M}'}(\alpha_{\mathcal{M}'}) \prod_{i=1}^{N} p_{S_i}(Y_i|\alpha_{S_i}). \tag{11}$$

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.

3. Bayesian Inference for Partial Orders

In Equation (5), we gave a likelihood $p(Y|h)$ for a poset h, when we have N lists $Y_1, \ldots, Y_N$ observed on choice sets $S_1, \ldots, S_N$. We now give a prior and posterior for h.

We start by representing h in terms of a set of continuous latent variables adapting models for random partial orders given by Winkler (1985) and Nicholls and Muir Watt (2011). An example is shown in Figure 3. 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 have $j_1 \succ_U j_2$, if and only if the two rows of weights satisfy $U_{j_1,k} > U_{j_2,k}$ for each $k = 1, \ldots, K$. 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 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 called an intersection order, and we write

$$h(U) = \bigcap_{k=1}^{K} h(U_{:,k}) \tag{12}$$

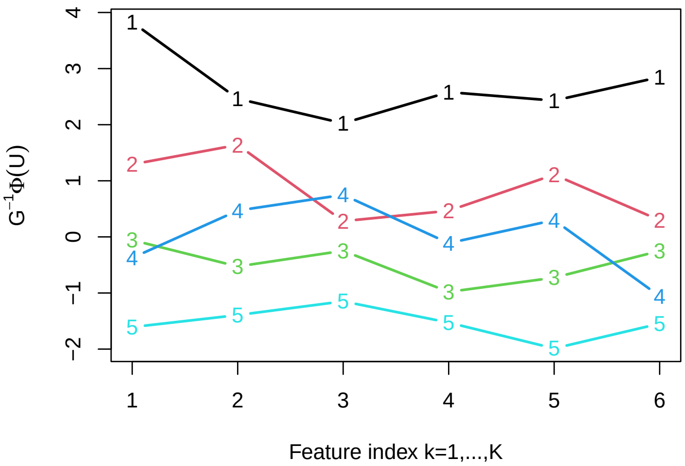

Figure 3: A U-matrix with $M = 5$ rows and $K = 6$ columns representing the partial order at left in Figure 1. Each path plots the sequence of values $G^{-1}(\Phi(U_{j,k}))$, $k = 1, \ldots, K$ in a row of feature values for choice $j \in \mathcal{M}$. The path for choice 1 lies entirely above that for 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

.

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

For any poset $h \in \mathcal{H}_{\mathcal{M}}$, the set of U matrices that map to h contains an open set in $R^{M\times K}$, so it has non-zero measure.

Lemma 4 *If $K \geq \lfloor M/2 \rfloor$, then the set of U-matrices satisfying $h = h(U)$ for any given $h \in \mathcal{H}_{\mathcal{M}}$ has infinite volume measure in $R^{M\times K}$.*

Proof 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 $U \in \mathcal{W}_{M,[K]}$. ∎

This latent variable setup makes it straightforward to add covariates. In terms of the covariate notation in Section 2, with $\alpha = X\beta$ an $M \times 1$ vector, let

$$\eta = U + \alpha 1_K^T, \tag{13}$$

where $\alpha 1_K^T$ is an outer product of α with a vector of K ones. The partial order is computed from the shifted weights, $h = h(\eta)$. As each row $\eta_{j,:} = U_{j,:} + \alpha_j 1_K^T$, a large positive effect $\alpha_j = x_j^T\beta$ for object j tends to move j up in the partial order. We do not include an intercept among the covariates as that duplicates the degree of freedom corresponding to the mean of $U_{j,:}$.

We get a prior over $h \in \mathcal{H}_{\mathcal{M}}$ by taking a prior over η. We will ask for the prior distribution over partial order "depths" to be uninformative of the depth of the reconstructed partial order as this is of interest in the inference. The depth $d(h) \in \{1, 2, \ldots, M\}$ of a partial order is the length of the longest path in the DAG representing h, that is the number of elements in the biggest complete suborder,

$$d(h) = \max_{S \in \mathcal{B}_{\mathcal{M}}} \{|S| : (S, \succ_h) \in \mathcal{C}_S\}.$$

In order to control $d(h(\eta))$, Nicholls et al. (2022) take the rows of the η-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 Σ_ρ 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 $k \neq k'$. The correlation parameter ρ is positive and controls the typical depth of $h(\eta)$. When ρ is close to one, the values $\eta_{j,k}$ don't vary much with k, so the orders $\succ_k$ are all the same and $d(h(\eta))$ is close to M. When ρ 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. Nicholls et al. (2022) show using simulation that taking $\rho \sim \mathrm{Beta}(1, 1/6)$ gives a marginal prior distribution for h which is reasonably uninformative of depth.

In order to get a model for random partial orders nesting PL, we modify this prior. 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 Φ be the CDF of a standard normal.

Theorem 5 *(Partial Order Model) For α and Σ_ρ defined above, if we take*

$$U_{j,:} \sim N(0, \Sigma_\rho), \quad \textit{independent for each } j \in \mathcal{M}, \tag{14}$$

$$\eta_{j,:} = G^{-1}(\Phi(U_{j,:})) + \alpha_j 1_K^T, \quad \textit{and} \tag{15}$$

$$y \sim p(\cdot | h(\eta(U, \beta))), \tag{16}$$

then $h(\eta_{:,k}) \sim PL(\alpha; \mathcal{M})$ for each $k = 1, \ldots, K$. In particular, if $K = 1$ then $y \sim PL(\alpha; \mathcal{M})$.

Proof The CDF of a standard normal Φ is applied to each element of $U_{j,:}$, 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 α to get $\eta_{j,k} \sim \mathrm{Gumbel}(\alpha_j),\ k = 1, \ldots, K$. 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 2. If $K = 1$ then $h(\eta(U, \beta)) = h(\eta_{:,1})$ is a complete order so it only has one linear extension, $h(\eta_{:,1})$ itself, so we must observe $y = h(\eta_{:,1})$ and hence $y \sim PL(\alpha; \mathcal{M})$. ■

Equations (14) and (15) 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 $(U_{1,1},\dots,U_{1,K},U_{2,1},\dots,U_{m,K})$ 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

$$\pi_{\mathcal{M}}(h|\rho,\beta)=E_U(\mathbb{I}_{h(\eta(U,\beta))=h}),\quad h\in\mathcal{H}_S. \tag{17}$$

Every partial order in $\mathcal{H}_{\mathcal{M}}$ has non-zero prior probability in this model.

Corollary 6 *If $K\geq\lfloor M/2\rfloor$, then $\pi_S(h|\rho,\beta)>0$ for any ρ,β, $h\in\mathcal{H}_S$ and $S\in\mathcal{B}_{\mathcal{M}}$.*

Proof If $m=|S|$ and $S\in\mathcal{B}_{\mathcal{M}}$, then $m\leq M$, so $K\geq\lfloor m/2\rfloor$. By Lemma 4 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 η and β, we invert $\eta(U,\beta)$ to get $U(\eta,\beta)=\Phi^{-1}(G(\eta-X\beta 1_K^T))$. At fixed β 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

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

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. (2022) show this for a similar timeseries model for random partial orders.

Corollary 7 *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 7 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,β (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 (11). In contrast, Theorem 3 gives marginal consistency for *observations* (a random order $y\in\mathcal{C}_{\mathcal{M}}$), so we called this property context independence there. As we saw in Section 1, we do not have context independence for lists generated from partial orders.

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

$$\begin{aligned}\pi_{\mathcal{M}'}(\rho,u,\beta|Y)&\propto\pi_R(\rho)\pi_B(\beta)\pi(u|\rho)p(Y|h(\eta(u,\beta))),\\ &\propto\pi_R(\rho)\pi_B(\beta)\left[\prod_{j\in\mathcal{M}'}N(u_{j,:};0_K,\Sigma_\rho)\right]\times\left[\prod_{i=1}^N p(Y_i|h(\eta(u_{S_i,:},\beta)))\right],\end{aligned} \tag{18}$$

where we integrated over $U_{j,:},\ j \in \mathcal{M} \setminus \mathcal{M}'$ using marginal consistency to remove degrees of freedom that are uninteresting. If we have samples $\rho^{(t)}, u^{(t)}, \beta^{(t)},\ t = 1, \ldots, T$ distributed according to $\pi(\rho, u, \beta|Y)$, and we want samples from the marginal posterior over partial orders, then we simply set $h^{(t)} = h(\eta(u^{(t)}, \beta^{(t)})),\ t = 1, \ldots, T$.

If the data are realised on the choice set $\mathcal{M}$ as $Y_i' \sim p(\cdot|h(\eta(U, \beta)))$, but we only observe $Y_i = Y_i'[S_i],\ i = 1, \ldots, N$, then we have parameters ρ and β as before, but now $U \in \mathbb{R}^{M \times K}$ and we introduce auxiliary variables $Y_i' \in \{y \in \mathcal{C}_\mathcal{M} : y[S_i] = Y_i\},\ i = 1, \ldots, N$ that agree with Y_i on the suborder. The posterior is

$$\pi_\mathcal{M}(\rho, U, \beta, Y'|Y) \propto \pi_R(\rho)\pi_B(\beta) \left[\prod_{j \in \mathcal{M}} N(U_{j,:}; 0_K, \Sigma_\rho)\right] \times \left[\prod_{i=1}^{N} p(Y_i'|h(\eta(U, \beta)))\right], \quad (19)$$

and we have to work on the full space $U \in \mathbb{R}^{M \times K}$ and further marginalise over Y'.

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

$$\pi_{\mathcal{M}'}(K, \rho, U, \beta|Y) \propto \pi_K(K)\pi_R(\rho)\pi_B(\beta)\pi(u|\rho, K)p(Y|h(\eta(u, \beta))), \quad (20)$$

where $\pi_K(K)$ is a prior for K (geometric with mean chosen so that $\pi_K(K > M'/2) \simeq 0.5$). 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.

4. A hierarchical model for grouped data

We now give a hierarchical model for list data which come in labeled groups, one group for each assessor. This was inspired by Crispino et al. (2019) who give a hierarchical model for Mallows ranking. Just as in Section 3, we learn about the partial order representing the preferences of each assessor from the lists they returned, but now we shrink the fitted partial orders towards the central partial order. The model can be represented as a tree. It has a root node with label 0 for the central poset and leaves with labels 1 through A, one leaf for each assessor. The posets on the leaves have a conditional distribution given the root poset, so we can think of the model as "shrinking" the posets at the leaves towards the central poset. In the following, we write down priors for all these objects, building in the properties we need, and then give the posterior distribution for the hierarchy of posets given the labeled list data.

4.1 Grouped data

Let $\mathcal{A} = \{1, \ldots, A\}$ be a set of A "assessors," who rank objects selected from a universe $\mathcal{M} = \{1, \ldots, M\}$. For $a \in \mathcal{A}$, assessor a is presented with N_a choice sets $S_{a,i} \in \mathcal{B}_\mathcal{M},\ i = 1, \ldots, N_a$ of varying size. They order the objects in each set from most favoured to least favoured and we assume they order on the choice set and not as a suborder. Let $m_{a,i} = |S_{a,i}|$ be the number of objects in the i'th batch ordered by assessor a and let $\mathcal{M}_a = \cup_{i=1}^{N_a} S_{a,i}$ be the set of objects assessor a was actually asked to rank with $m_a = |\mathcal{M}_a|$. Let $S_a = (S_{a,1}, \ldots, S_{a,N_a})$ and $S = (S_1, \ldots, S_A)$. Let $\mathcal{M}_0 = \cup_{a \in \mathcal{A}} \mathcal{M}_a$ be the set of objects ranked by at least one assessor and set $M_a = |\mathcal{M}_a|,\ a = 0, 1, \ldots, A$. Let $Y_{a,i} \in \mathcal{C}_{S_{a,i}}$ be the complete order assessor

a returned for choice set $S_{a,i}$. We condition on a fixed collection of choice sets which may be chosen in an arbitrary way. Let $Y_a = (Y_{a,1}, \ldots, Y_{a,N_a})$ be the data from assessor a and let $Y = (Y_1, \ldots, Y_A)$ be all data.

4.2 A hierarchy of posets

Let $U^{(0)} \in \mathbb{R}^{M \times K}$ be a matrix of preference weights with one K-dimensional row-vector $U_j^{(0)} = (U_{j,1}^{(0)}, \ldots, U_{j,K}^{(0)})$ for each object $j \in \mathcal{M}$. These are the "global" preference weights, which inform the preferences of individual assessors. Let $u^{(0)} = (U_{j,:})_{j \in \mathcal{M}_0}$ be the preference weights for objects ranked by at least one assessor. Let $h^{(0)} = h(u^{(0)}),\ h^{(0)} \in \mathcal{H}_{\mathcal{M}_0}$ be the global partial order determined by $u^{(0)}$, and let $H^{(0)} = h(U^{(0)}),\ H^{(0)} \in \mathcal{H}_{\mathcal{M}}$ be the global partial order on the universe of objects $\mathcal{M}$.

For $a \in \mathcal{A}$, let $U^{(a)} \in \mathbb{R}^{M \times K}$ be a matrix with one row for each object in $\mathcal{M}$, so that $U_j^{(a)} \in \mathbb{R}^K$ is the vector of preference-weights assigned by assessor a to object j. Let $u^{(a)} = U_{\mathcal{M}_a}^{(a)}$ with $U_{\mathcal{M}_a}^{(a)} = (U_{j,:}^{(a)})_{j \in \mathcal{M}_a}$ be the preference weights for the objects assessor a actually ranked so $u^{(a)} \in \mathbb{R}^{M_a \times K}$. Let $h^{(a)} = h(u^{(a)}),\ h^{(a)} \in \mathcal{M}_a,\ a = 0, 1, \ldots, A$ be the partial order of preferences held by assessor a for the objects they ranked and correspondingly $H^{(a)} = h(U^{(a)}),\ H^{(a)} \in \mathcal{H}_M,\ a = 0, 1, \ldots, A$, their partial orders on the universe of objects. Write $u = (u^{(0)}, u^{(1)}, \ldots, u^{(A)})$ and $h = h(u)$ (applied matrix by matrix), so that $h = (h^{(0)}, h^{(1)}, \ldots, h^{(A)})$. The objects U and H are defined accordingly. We learn about h without having to infer H. Denote by $\pi_{\mathcal{M}^{A+1}}(H|\psi)$ a prior for H in $\mathcal{H}_{\mathcal{M}}^{A+1} = \mathcal{H}_{\mathcal{M}} \times \cdots \times \mathcal{H}_{\mathcal{M}}$ ($A+1$ times), with ψ a set of prior parameters we specify later. Let $s_{0:A} = (s_0, s_1, \ldots, s_A)$ be a set of $A+1$ subsets $s_a \in \mathcal{B}_{\mathcal{M}},\ a \in \mathcal{A}$ and let $\mathcal{H}_{s_{0:A}} = \mathcal{H}_{s_0} \times \mathcal{H}_{s_1} \times \ldots \mathcal{H}_{s_A}$. Let $H[s_{0:A}] = (H^{(0)}[s_0], H^{(1)}[s_1], \ldots, H^{(A)}[s_A])$, so that $H[s_{0:A}] \in \mathcal{H}_{s_{0:A}}$ is a set of suborders and let $\mathcal{M}_{0:A} = (\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_A)$. Denote by $\pi_{\mathcal{H}_{\mathcal{M}_{0:A}}}(h|\psi),\ h \in \mathcal{H}_{\mathcal{M}_{0:A}}$ a prior for h.

4.3 A prior for the poset hierarchy

We now give a joint prior for the latent preference weight matrices in U. This will determine

$$\pi_{\mathcal{M}^{A+1}}(H|\psi) = E_{U|\psi}(\mathbb{I}_{h(U)=H}),$$

for $H \in \mathcal{H}_{\mathcal{M}}^{A+1}$. Similarly, if $s_{0:A} \in \mathcal{B}_{\mathcal{M}}^{A+1}$ is any collection of $A+1$ subsets of $\mathcal{M}$ with $s_0 = \cup_{a=1}^{A} s_a$, if $h \in \mathcal{H}_{s_{0:A}}$ and $U_{s_{0:A}} = (U_{s_0}^{(0)}, U_{s_1}^{(1)}, \ldots, U_{s_A}^{(A)})$ then

$$\pi_{s_{0:A}}(h|\psi) = E_{U_{s_{0:A}}|\psi}(\mathbb{I}_{h(U_{s_{0:A}})=h}).$$

Notice that the prior for partial orders on subsets s of the universe of choices is not defined as the marginal obtained by taking suborders of the prior on the universe of choices, but separately, and only involves latent variables for choices in s.

Before we write down the prior, we list the properties it must posses. We clearly need a model with a hyperparameter controlling the correlation of $H^{(0)}$ and $H^{(a)}$. We want $H^{(1)}, \ldots, H^{(A)}$ to be exchangeable, and we further require $H^{(0)} \sim H^{(a)}$ for all $a \in \mathcal{A}$. If we think of the unknown true partial orders for assessors as samples from a larger population, then the global order $H^{(0)}$ is representative of the same population, so it should have the same marginal prior distribution. We could not put these conditions on h as its component partial orders live in different spaces. We also require prior hierarchy to be marginally consistent.

Definition 8 *(Marginally Consistent Hierarchy) The family of probability distributions* $\pi_{s_{0:A}}(\cdot|\psi)$, $s_{0:A} \in \mathcal{B}_{\mathcal{M}}^{A+1}$ *is marginally consistent if* $H \sim \pi_{\mathcal{M}^{A+1}}(\cdot|\psi)$ *implies* $H[s_{0:A}] \sim \pi_{s_{0:A}}(\cdot|\psi)$ *for every* $s_{0:A} \in \mathcal{B}_{\mathcal{M}}^{A+1}$ *with* $s_0 = \cup_{a=1}^{A} s_a$.

This says the prior we get by taking suborders of random partial orders on the universe of choices is the same as the prior we defined separately for the suborders. This will work because $h(U)[s] = h(U_s)$. Finally, we would like the same control over the distributions of the depths $d(h^{(a)}), a = 0, 1, \dots, A$ as we had for a single partial order in Section 3.

We now give this prior and set out its properties.

Theorem 9 *(Hierarchical Partial Order prior) For* $\alpha_{\mathcal{M}} = X\beta$ *and* Σ_ρ *as in Section 3, for* $0 < \tau \le 1$ *and sets* $\mathcal{M}_a \in \mathcal{B}_{\mathcal{M}},\ a \in \mathcal{A}$, *let* $\mathcal{M}_0 = \cup_{a=1}^{A} \mathcal{M}_a$ *and*

$$U_{j,:}^{(0)} \sim N(0, \Sigma_\rho), \quad \textit{independent for each } j \in \mathcal{M}_0, \tag{21}$$

$$U_{j,:}^{(a)} | U_{j,:}^{(0)} \sim N\left(\tau U_{j,:}^{(0)}, (1-\tau^2)\Sigma_\rho\right) \quad \textit{independent for each } a \in \mathcal{A} \textit{ and } j \in \mathcal{M}_a, \tag{22}$$

$$\eta_{j,:}^{(a)} = G^{-1}(\Phi(U_{j,:}^{(a)})) + \alpha_j 1_K^T \quad \textit{for } a = 0, 1, \dots, A \textit{ and } j \in \mathcal{M}_a \textit{ and} \tag{23}$$

$$h = h(\eta(U, \beta)) \quad \textit{for } \eta(U,\beta) = (\eta^{(0)}, \eta^{(1)}, \dots, \eta^{(A)}) \textit{ and } h^{(a)} = h(\eta^{(a)}).$$

For $h \in \mathcal{H}_{\mathcal{M}_{0:A}}$ *let* $\pi_{\mathcal{M}_{0:A}}(h|\rho,\beta,\tau) = E_U(\mathbb{I}_{h(\eta(U,\beta))=h})$ *be the resulting prior for the poset hierarchy* h *and let* $\pi_{\mathcal{M}_{0:A}}(h^{(a)}|\rho,\beta,\tau) = E_U(\mathbb{I}_{h(\eta^{(a)}(U^{(a)},\beta)=h^{(a)}})$ *be the marginal for* $h^{(a)}$. *The prior* $\pi_{\mathcal{M}_{0:A}}$ *has the following properties:*

1. *(single PO marginals) For* $a = 0, 1, \dots, A$, $\pi_{\mathcal{M}_{0:A}}(h^{(a)}|\rho,\beta,\tau) = \pi_{\mathcal{M}_a}(h^{(a)}|\rho,\beta)$ *where* $\pi_{\mathcal{M}_a}$ *is the single partial-order prior given in* (17)*;*
2. *(PL hierarchy at* $K = 1$*) when* $K = 1$, $h^{(a)} \sim PL(\alpha_{\mathcal{M}_a}, \mathcal{M}_a)$ *for* $a \in \mathcal{A}$*;*
3. *(independent of* $h^{(0)}$ *at* $\tau = 0$*)* $\pi_{\mathcal{M}_{0:A}}(h|\rho,\beta,\tau=0) = \prod_{a=0}^{A} \pi_{\mathcal{M}_a}(h^{(a)}|\rho,\beta,\tau)$*;*
4. *(matching* $h^{(0)}$ *as* $\tau \to 1$*)* $\pi_{\mathcal{M}_{0:A}}(h^{(a)}|\rho,\beta,\tau) \to \mathbb{I}_{h^{(a)}=h^{(0)}}$ *as* $\tau \to 1$ *for each* $a \in \mathcal{A}$*;*
5. *(coverage) if* $K \ge \lfloor M/2 \rfloor$ *then* $\pi_{\mathcal{M}_{0:A}}(h|\rho,\beta,\tau) > 0$ *for all* $h \in \mathcal{H}_{\mathcal{M}_{0:A}}$*;*
6. *(marginal consistency) If* $\mathcal{M}_a = \mathcal{M}$ *for each* $a = 0, 1, \dots, A$ *and* $H \sim \pi_{\mathcal{M}^{A+1}}(\cdot|\rho,\beta,\tau)$ *then* $H[s_{0:A}] \sim \pi_{s_{0:A}}(\cdot|\rho,\beta,\tau)$ *for every* $s_{0:A} \in \mathcal{B}_{\mathcal{M}}^{A+1}$ *with* $s_0 = \cup_{a=1}^{A} s_a$.

Proof We can write

$$U_{j,:}^{(a)} = \tau U_{j,:}^{(0)} + \epsilon_{j,:}^{(a)} \tag{24}$$

with

$$\epsilon_{j,:}^{(a)} \sim N(0_K, (1-\tau^2)\Sigma_\rho)$$

independent of everything else so the (marginal) covariance of $U_{j,:}^{(a)}$ is

$$\mathrm{cov}(\tau U_{j,:}^{(0)}) + (1-\tau^2)\Sigma_\rho = \Sigma_\rho,$$

and the mean of $U^{(a)}_{j,:}$ is zero. It follows that marginally,

$$U^{(a)}_{j,:} \sim N(0_{m_a}, \Sigma_\rho).$$

Since this is the distribution of $U_{j,:}$ in the single-PO model in Theorem 5 and all else is the same, this establishes result 1 of Theorem 9 and result 2 then follows from Theorem 5. Result 3 follows from (24) also, as $\mathrm{cov}(U^{(a)}_{j_1,:}, U^{(0)}_{j_2,:}) = \tau\Sigma_\rho$ and $\mathrm{cov}(U^{(a)}_{j_1,:}, U^{(a')}_{j_2,:}) = \tau^2\Sigma_\rho$ so $h(\eta^{(a)}(U^{(a)}, \beta))$ and $h(\eta^{(a')}(U^{(a')}, \beta))$ are functions of jointly normal independent random variables for $a \neq a'$. Also, $\eta^{(a)}_{j,k} - \eta^{(0)}_{j,k} = U^{(a)}_{j,k} - U^{(0)}_{j,k}$ so (given $U^{(0)}$) we have

$$(\eta^{(a)}_{j,k} - \eta^{(0)}_{j,k} | U^{(0)}) \sim N\big((\tau - 1)U^{(0)}_{j,k},\, (1 - \tau^2)\big).$$

When $\tau \to 1$ this gives $\Pr(\max_{j,k} |\eta^{(a)}_{j,k} - \eta^{(0)}_{j,k}| > \delta | U^{(0)}) \to 0$ for all $\delta > 0$ so $h^{(a)} \xrightarrow{P} h^{(0)}$ also giving result 4. Results 5 and 6 may be shown using the same reasoning as Lemma 4 and Theorem 3 respectively. ■

The relation between $U^{(0)}_{j,:}$ and $U^{(a)}_{j,:}$ is equivalent to running a multivariate Ornstein-Uhlenbeck process $dX_t = -X_t + V^{1/2}dW_t$ with $V = 2\Sigma_\rho$ starting at $X_0 = U^{(0)}_{j,:}$ to get $X_t \sim U^{(a)}_{j,:}$ at $t = -2\log(\tau)$. The stationary distribution of this process is $N(0_K, \Sigma_\rho)$ so if we run the process for no time at all ($\tau = 1$) we get back $U^{(0)}_{j,:}$ and if we run it for an infinite amount of time ($\tau \to 0$) we get an independent draw from $N(0_K, \Sigma_\rho)$.

4.4 Observation Model and posterior distribution

The model parameters are ρ, β, τ and U. The priors for ρ and β are unchanged from Section 3. The prior for τ is $\tau \sim U(0, 1)$. The prior for $U = (U^{(0)}, U^{(1)}, \ldots, U^{(A)})$ is given in (21) and (22).

The likelihood for these parameters given the grouped data is

$$p_S(Y|\beta, U) = \prod_{a=1}^{A} p_{S_a}(Y_a | h(\eta(U^{(a)}, \beta)))$$

where

$$p_{S_a}(Y_a | h(\eta(U^{(a)}, \beta))) = \prod_{i=1}^{N_a} p_{S_{a,i}}(Y_{a,i} | h^{(a)}[S_{a,i}]) \tag{25}$$

with $p_{S_{a,i}}(Y_{a,i} | h^{(a)}[S_{a,i}])$ given in (2) and

$$h^{(a)}[S_{a,i}] = h(\eta(U^{(a)}_{S_{a,i}}, \beta)).$$

This depends only on β and U, highlighting the role of ρ and τ as prior hyper-parameters.

The posterior is

$$\pi_S(\rho,\beta,\tau,U|Y) \propto \pi_R(\rho)\pi_B(\beta)\pi_T(\tau)\pi(U|\rho,\tau)p_S(Y|h(U,\beta)) \tag{26}$$

$$= \pi_R(\rho)\pi_B(\beta)\pi_T(\tau)\prod_{j=1}^{m_0} N\left(U^{(0)}_{j,:};0_K,\Sigma_\rho\right)\times$$

$$\prod_{a=1}^{A}\left[p_{S_a}(Y_a|h(\eta(U^{(a)},\beta)))\prod_{j=1}^{m_a} N\left(U^{(a)}_{j,:};\tau U^{(0)}_{j,:},(1-\tau^2)\Sigma_\rho\right)\right]. \tag{27}$$

As for the single poset analysis in Section 3, inference for K may be added, as in (20).

5. Extending the observation model to allow for noise

In this section we give models for "noise" on observations. For ease of exposition, we drop the assessor labels and return to the "single-partial-order" setup of Sections 1 and 3.

5.1 Queue jumping noise

We allow for noise in the observation models in Sections 1 and 4.4 by allowing individuals to "jump the queue", exploiting the sequential selection structure of the poset likelihood in (3). A generic list $y_{1:m}$, written as a permutation of the elements of a choice set $S=[m]$ and constrained by a poset $h=(S,\succ_h)$, is formed by taking items from the top of a queue, which continues to mix rapidly, constrained by the suborder on those remaining. Before the i'th entry of $y_{1:m}$ is chosen, there are $m-i+1$ elements of S, with labels $y_{i:m}$, yet to be placed. With probability p the next entry (i.e., y_i) is chosen at random, ignoring any order constraints. Otherwise, y_i is the maximal element in a random linear extension of the suborder $h[y_{i:m}]$ for the elements remaining, so the probability y_i is chosen next is $q_{y_{i:m}}(y_i|h[y_{i:m}])$ in (4). Working from the top down,

$$p^{(D)}_S(y_{1:m}|h,p)=\prod_{i=1}^{m-1} q(y_{i:m}|h[y_{i:m}],p) \tag{28}$$

where

$$q(y_{i:m}|h[y_{i:m}],p)=\frac{p}{m-j+1}+(1-p)\,q_{y_{i:m}}(y_i|h[y_{i:m}]). \tag{29}$$

When we use this model in the posterior, we add a parameter p in (27) with prior $\mathrm{Beta}(\alpha_p,\beta_p)$ and replace the likelihood $p_{S_{a,i}}(Y_{a,i}|h^{(a)}[S_{a,i}])$ in (25) with $p^{(D)}_{S_{a,i}}(Y_{a,i}|h^{(a)}[S_{a,i}],p)$ in (28).

5.2 Mallows noise

Jiang and Nicholls (2021) give a model for observations in the single-poset model of Section 1 in which the "unknown true" data $\ell\in\mathcal{C}_S$ is a linear extension of a poset $h=(S,\succ_h)$ with $|S|=m$, and we observe $y\in\mathcal{C}_S$ with "Mallows noise" on ℓ. In the Mallows-ϕ model the probability $p_S^{(M)}(y|\ell,\theta)$ to observe y when the true order is ℓ is given in terms of a symmetric divergence $d(\ell,y)\geq 0$ between ℓ and y. Let $\sigma(\ell,j)=\{k\in[m]:\ell_k=j\}$ and let

$$d(y,\ell)=\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\mathbb{I}_{\sigma(\ell[y_{i:m}],y_i)>\sigma(\ell[y_{i:m}],y_j)}.$$

The Mallows ϕ-model is the distribution over y centred on ℓ,

$$p(y|\ell,\theta)=\frac{\exp(-\theta d(\ell,y))}{\Psi_m(\theta)},\quad y\in\mathcal{C}_S, \tag{30}$$

where $\Psi_m(\theta)=\sum_{z\in\mathcal{C}_S}e^{-\theta d(\ell,z)}$ is a closed-form normalising constant $\Psi_m(\theta)=\prod_{i=1}^m\psi_i$ with $\psi_i=\sum_{j=1}^i e^{-(i-1)\theta}$. This is is a sequential choice model as

$$\begin{aligned}p(y|\ell,\theta)&=\prod_{i=1}^{m-1}\frac{\exp(-\theta\sum_{j=i+1}^m\mathbb{I}_{\sigma(\ell[y_{i:m}],y_i)>\sigma(\ell[y_{i:m}],y_j)})}{\psi_{m-i+1}(\theta)} &(31)\\ &=\prod_{i=1}^{m-1}q(y_i|\ell[y_{i:m}],\theta), &(32)\end{aligned}$$

where $q(y_i|\ell[y_{i:m}],\theta)$ is the probability y_i is selected next from the remaining choices.

We give an algorithm to evaluate the likelihood, which is a weighted sum over linear extensions. We adapt a recursive counting algorithm (Knuth and Szwarcfiter, 1974), following Jiang and Nicholls (2021). The likelihood itself is the marginal probability to observe y given the partial order $h=(S,\succ_h)$,

$$p_S^{(M)}(y|h,\theta)=\sum_{\ell\in\mathcal{L}[h]}p(y|\ell,\theta)p_S(\ell|h) \tag{33}$$

using (2) and the symmetry of $d(\ell,y)$ in its arguments

$$=\frac{1}{|\mathcal{L}[h]|}\sum_{\ell\in\mathcal{L}[h]}p(\ell|y,\theta),$$

then since $\mathcal{L}[h]=\cup_{j\in\max(h)}\mathcal{L}_j[h]$,

$$=\frac{1}{|\mathcal{L}[h]|}\sum_{j\in\max(h)}q^{(M)}(j|y_{-j},\theta)\sum_{\ell'\in\mathcal{L}[h_{-j}]}p(\ell'|y_{-j},\theta), \tag{34}$$

where $h_{-j}=(S\setminus\{j\},\succ_h)$ and similarly for y_{-j}. This gives a recursion in which a function $f(h,y,\theta)$ returns the sum $\sum_{\ell\in\mathcal{L}[h]}p(\ell|y,\theta)$ over orders of m elements, by evaluating the sums in the last line and calling itself to evaluate $\sum_{\ell'\in\mathcal{L}[h_{-j}]}p(\ell'|y_{-j},\theta)$ on orders of length $m-1$. The recursion stops if f is called with $h\in\mathcal{C}_S$ (one extension, so return (30)) or if h is the empty order (then $\mathcal{L}[h]=\mathcal{C}_S$ so the sum is one) which means it stops with $m\geq 2$. We evaluate $|\mathcal{L}[h]|$ in the same pass, as $|\mathcal{L}[h]|=m!f(h,y,\theta=0)$.

When we use this noise model in the posterior we add a parameter θ in (27) with prior $\mathrm{Exp}(\mu_p)$ and replace $p_{S_{a,i}}(Y_{a,i}|h^{(a)}[S_{a,i}])$ in (25) with $p^{(M)}_{S_{a,i}}(Y_{a,i}|h^{(a)}[S_{a,i}],\theta)$ in (33).

5.3 η-weighted queue-jumping

Paired data, in which the choice set for each each ordered data list has just two elements, highlight a feature of our observation model. Let $h \in \mathcal{H}_{\mathcal{M}}$ be a generic partial order and let $S = \{j_1, j_2\}$ be a choice set with $m = 2$ elements. In this setting, the noise models in Sections 5.1 and 5.2 are the same. Referring to (28), if $h[S]$ is the empty order then $p_S^{(D)}(y|h,p) = p/2 + (1-p)/2 = p/2$ (two elements in the choice set and two extensions) for $y = y^{(1)} = (j_1, j_2)$ and $y = y^{(2)} = (j_2, j_1)$. If $j_1 \succ_h j_2$ then we observe $y^{(1)}$ with probability $p/2 + (1-p) = 1 - p/2$ (since $h[S]$ has just one extension). The Mallows distance in (32) is just $d(y,\ell) = \mathbb{I}_{y=\ell}$ so $p(y|\ell,\theta) = (1+\exp(-\theta))^{-1}$ if $y = \ell$ and otherwise $\exp(-\theta)/(1+\exp(-\theta))$. It follows from (33) that $p_S^{(M)}(y|h,\theta) = 1/2$ if $h[S]$ is unordered and $p_S^{(M)}(y^{(1)}|h,\theta) = (1+\exp(-\theta))^{-1}$, so the parameter mapping is $p = 2\exp(-\theta)/(1+\exp(-\theta))$.

Further, our observation models do not take into account how far apart j_1 and j_2 were in h before we took the suborder. In some settings we might expect that an error is less likely if the suborder is $j_1 \succ_h j_2$ but j_1 was far above j_2 in h (for example, if h is a total order and $\max(h) = j_1$ and $\min(h) = j_2$). We can account for this in the Queue-Jumping error model. Referring to (29), an error occurs with probability p. When this happens the next element in the list is chosen uniformly at random from this that remain. We now let this choice depend on $\eta(U,\beta)$: for $j \in \mathcal{M}$ let $\bar{\eta}_j = K^{-1}\sum_{k=1}^{K} \eta_{j,k}$; when an error occurs, the next element y_i in the list is chosen with probability proportional to $\exp(\bar{\eta}_{y_i})$ as in (8); when there is no error the next element is chosen according to the noise-free partial-order model in (4); the modified queue-jumping likelihood is

$$p_S^{(D,\eta)}(y_{1:m}|\eta,p) = \prod_{i=1}^{m-1} q(y_{i:m}|\eta_{y_{i:m},:},p) \tag{35}$$

where

$$q(y_{i:m}|\eta_{y_{i:m}},p) = p\, q(y_i|\bar{\eta}_{y_{i:m}}) + (1-p)\, q_{y_{i:m}}(y_i|h(\eta_{y_{i:m}})) \tag{36}$$
$$\tag{37}$$

and

$$q(y_i|\bar{\eta}_{y_{i:m}}) = \frac{e^{\bar{\eta}_{y_i}}}{\sum_{j=i}^{m} e^{\bar{\eta}_{y_j}}} \tag{38}$$

is the one-step Plackett-Luce probability to select y_i from $y_{i:m}$ given weights $\bar{\eta}_{y_{i:m}}$. The choice probability for the next element in the list is a mixture of Plackett-Luce and the noise-free partial order observation model (2). When $p = 0$ we recover the partial order model and when $p = 1$ we recover Plackett-Luce.

This has some nice features. It captures the idea that errors become less likely as the distance between j_1 and j_2 on h increases. For example, when there are just two choices and $j_1 \succ_h j_2$,

$$p_S^{(D,\eta)}(y^{(2)}|\eta,p) = p\frac{1}{1+e^{\bar{\eta}_{j_1}-\bar{\eta}_{j_2}}}.$$

When j_1 has a much higher position in h than j_2 it will typically hold that $\bar{\eta}_{j_1} \gg \bar{\eta}_{j_2}$ (we must have $\bar{\eta}_{j_1} > \bar{\eta}_{j_2}$ as $j_1 \succ_h j_2$ requires $\eta_{j_1,k} > \eta_{j_2,k}$ for each $k = 1, \dots, K$ and if there are many $j \in \mathcal{M}$ ordered between j_1 and j_2 then the paths $\eta_{j_1,:}$ and η_{j_2} must be well separated as many paths lie between them). The probability for an error goes to zero as $\bar{\eta}_{j_1} - \bar{\eta}_{j_2} \to \infty$. When j_1 is close to j_2 in h then it will typically hold that $\bar{\eta}_{j_1} - \bar{\eta}_{j_2}$ is small and in this case the probability for error approaches $p/2$.

When $p = 1$ the mixture in (35) reduces to Plackett-Luce in (8) with weights $\alpha_j = \bar{\eta}_j$ and in particular when $K = 1$ and $\eta_{j,:} = \lambda_j + x_j\beta$ with $\lambda_j = G^{-1}(\Phi(U_{j,:}))$ we have Plackett-Luce in familiar form. This gives an alternative way to nest Plackett-Luce within our model. We are not restricted to taking $U_{j,:} \sim N(0_K, \Sigma_\rho)$ and $\eta_{j,:} = G^{-1}(\Phi(U_j)) + x_j\beta$. If, for example, take $\eta_{j,:} = U_{j,:} + 1_K^T x_j \beta$ then the observation model is Plackett-Luce in (6) when $p = 1$ with weights $\bar{\eta}_j,\ j \in \mathcal{M}$ and prior $\bar{\eta}_j \sim N(0, 1/K)$.

6. Clustering unlabeled orders

In Section 4 the data came in A groups Y_a with assessor labels $a \in \mathcal{A}$. If we simply have a collection of N lists $Y = (Y_1, \dots, Y_N)$ with $Y_i \in \mathcal{C}_{S_i}$ and we think there may be some latent group structure, we add N unknown label parameters $c = (c_1, \dots, c_N)$ with $c_i \in \mathcal{A}$ for each $i \in [N]$. Take a multinomial-Dirichlet distribution for the grouping, assuming the number of groups A is fixed and the probability i is assigned to group $a \in \mathcal{A}$ is w_a with $w = (w_1, \dots, w_A)$ and $w \sim \text{Dirichlet}(\gamma/A, \dots, \gamma/A)$ (this parameterisation recovers the Chinese Restaurant Process (CRP) with parameter $\gamma > 0$ for non-empty groups as $A \to \infty$). Let $N_a(c) = \sum_{i=1}^N \mathbb{I}_{c_i = a}$. On integrating w the prior for c is

$$\pi_C(c|\gamma) = \frac{\Gamma(\gamma)}{\Gamma(\gamma/A)^A} \frac{\prod_{a=1}^A \Gamma(N_a(c) + \gamma/A)}{\Gamma(N + \gamma)}. \tag{39}$$

The likelihood now depends on the list-labels c, that is

$$p_S(Y|c, h(U, \beta)) = \prod_{i=1}^N p_{S_i}(Y_i | h^{(c_i)}[S_i]), \tag{40}$$

where $h^{(c_i)} = h(\eta^{(c_i)}(U^{(c_i)}, \beta))$. The posterior becomes

$$\pi_S(\rho, \beta, \tau, U, c|Y) \propto \pi_R(\rho)\pi_B(\beta)\pi_T(\tau)\, \pi(U|\rho, \tau)\, \pi_C(c|\gamma)\, p_S(Y|c, h(U, \beta)), \tag{41}$$

with π_C given in (39) and $p_S(Y|c, h(U, \beta))$ in (40) and otherwise as (27). As for the single poset analysis in Section 3 and the hierarchical model for labeled lists in Section 4, the likelihood may be extended to allow for noise via queue-jumping (28) or Mallows (30), adding a posterior variable p or θ respectively, and K may also be variable in the posterior, as in (20).

7. Hierarchies of Vertex-Series-Parallel Posets

The Likelihoods in (3) and (28) are expensive to evaluate as we must count the linear extensions of a partial order on the elements of S. This limits the applicability of the method to choice sets with up to around 20 elements. Jiang et al. (2023) work with a large sub-class of posets called Vertex Series Parallel posets (VSPs). These posets, characterised in Valdes (1978) and Valdes et al. (1982), admit linear-time evaluation of $|\mathcal{L}[h[S]]|$. This allowed Jiang et al. (2023) to fit single-poset models like the one in Section 3 to list data with hundreds of elements.

A VSP is a poset $v = (S, \succ_v)$ which has no suborders of the form $(\{j_1, j_2, j_3, j_4\}, \{j_1 \succ_v j_3, j_1 \succ_v j_4, j_2 \succ_v j_4\})$ (the "forbidden sub-graph"). If $\mathcal{V}_S$ is the set of all VSPs on a choice set S, then $\mathcal{V}_S \subset \mathcal{H}_S$ for $m > 3$. VSPs can be formed by a sequence of series and parallel merging operations on VSPs, and this leads to a Binary Decomposition Tree (BDT) representation. In some settings restriction to VSPs would actually be a natural modeling choice, rather than a way of simplifying calculations. See Jiang et al. (2023) for further discussion of these points and an explanation of how counting linear extensions works for VSPs.

Jiang et al. (2023) put a prior on VSPs by taking a prior over BDTs. This works well, as far as it goes, but it is unclear how to add covariates to the model, or construct timeseries or hierachical models for VSPs in the BDT parameterisation. We use a latent U-matrix representation, like that given above. Every VSP $v \in \mathcal{V}_S$ can be written in the form $v = \ell^{(1)} \cap \ell^{(2)}$ for some pair of complete orders $\ell^{(1)}, \ell^{(2)} \in \mathcal{C}_S$ (that is to say, their dimension is at most two). It follows that Lemma 4 holds for $K = 2$ - for any VSP $v \in \mathcal{V}_S$ the set $\{U \in \mathbb{R}^{2m} : h(U) = v\}$ has infinite volume.

Although every VSP has dimension at most two, not all two-dimensional posets are VSPs. We therefore take as our prior for a hierarchy of VSPs the same prior we took for $h = h(\eta(U, \beta))$ in Section 4, but set $K = 2$ and condition U on $h(\eta(U, \beta)) \in \mathcal{V}_{\mathcal{M}_{0:A}}$ with $\mathcal{V}_{\mathcal{M}_{0:A}} = \mathcal{V}_{\mathcal{M}_0} \times \mathcal{V}_{\mathcal{M}_1}, \ldots, \mathcal{V}_{\mathcal{M}_A}$. In an MCMC setting, we can initialise the chain with empty orders or complete orders (which are VSPs) and reject if we create a forbidden subgraph (so $h(\eta(U, \beta)) \notin \mathcal{V}_{\mathcal{M}_{0:A}}$). Testing a poset for the VSP property can be done in linear time (Valdes et al., 1982) and in our setting in constant time, as we maintain data structures which support the query.

8. Examples

8.1 HPO

Perhaps the strange-sounds dataset discussed by the Oslo group would work here. WBG to find a few more.

8.1.1 Poset-hierarchy

Use queue jumping and variable K of possible.

8.1.2 VSP-hierarchy

Use queue jumping or Mallows noise. Maybe add Mallows noise to Section 5.

8.1.3 Model comparison

8.2 Clustering

Many data sets. Anything that was previously used for PL-clustering

8.2.1 Poset-hierarchy

Use queue jumping and variable K of possible.

8.2.2 VSP-hierarchy

Use queue jumping or Mallows noise.

8.2.3 Model comparison

Acknowledgments

We thank the Oslo group for useful conversations!

References

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

Marta Crispino, Elja Arjas, Valeria Vitelli, Natasha Barrett, and Arnoldo Frigessi. A Bayesian Mallows approach to nontransitive pair comparison data: How human are sounds? *The Annals of Applied Statistics*, 13(1):492 – 519, 2019. doi: 10.1214/18-AOAS1203. URL `https://doi.org/10.1214/18-AOAS1203`.

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/.

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.

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, 2022.

Stephen Ragain and Johan Ugander. Pairwise choice markov chains. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/bcb41ccdc4363c6848a1d760f26c28a0-Paper.pdf.

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

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.

Jacobo Valdes. *Parsing Flowcharts and Series-Parallel Graphs.* PhD thesis, Stanford University, Stanford, CA, USA, 1978. AAI7905944.

Jacobo Valdes, Robert E. Tarjan, and Eugene L. Lawler. The recognition of series parallel digraphs. *SIAM Journal on Computing*, 11(2):298–313, 1982. doi: 10.1137/0211023. URL https://doi.org/10.1137/0211023.

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.