Statistical Models for Rank Data using Partial Orders: an Introduction

Presenters: Geoff Nicholls & Kate Lee

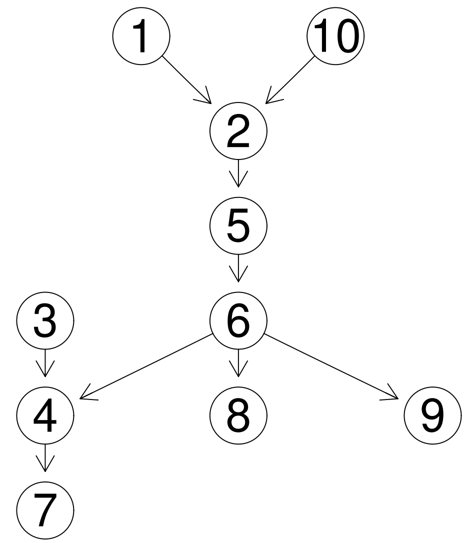
PrefStat 25 - Oslo 04/07/25

Spoiler alert - a look ahead

We will have data which are ordered lists.

We want to estimate a partial order* summarising relations between items in the lists.

5	6	4	7		
3	4				
1	2				
10	1	2	5	6	8
1	10	2	5	6	9
1	2				
10	1	2			



Assume assessor constructs lists on varying choice set.

Above right all order relations attested and not contradicted.

We will estimate the PO using a statistical model and quantify uncertainty in reconstruction.

*A partial order is no longer part of a complete ranking (as earlier this week).

What is a partial order?

A partial order is an incomplete set of relations.

Example: I like fruit which is sweet and crisp.

$$\text{fruit} = (\text{sweetness}, \text{crispness})$$

$$\text{fruit}_1 \succ \text{fruit}_2$$
$$\Leftrightarrow$$
$$\text{sweetness}_1 > \text{sweetness}_2 \quad \& \quad \text{crispness}_1 > \text{crispness}_2$$

so if say apple $= (8, 9)$, pear $= (7, 4)$, quince $= (2, 8)$ then

$$\text{apple} \succ \text{pear} \quad \& \quad \text{apple} \succ \text{quince}$$

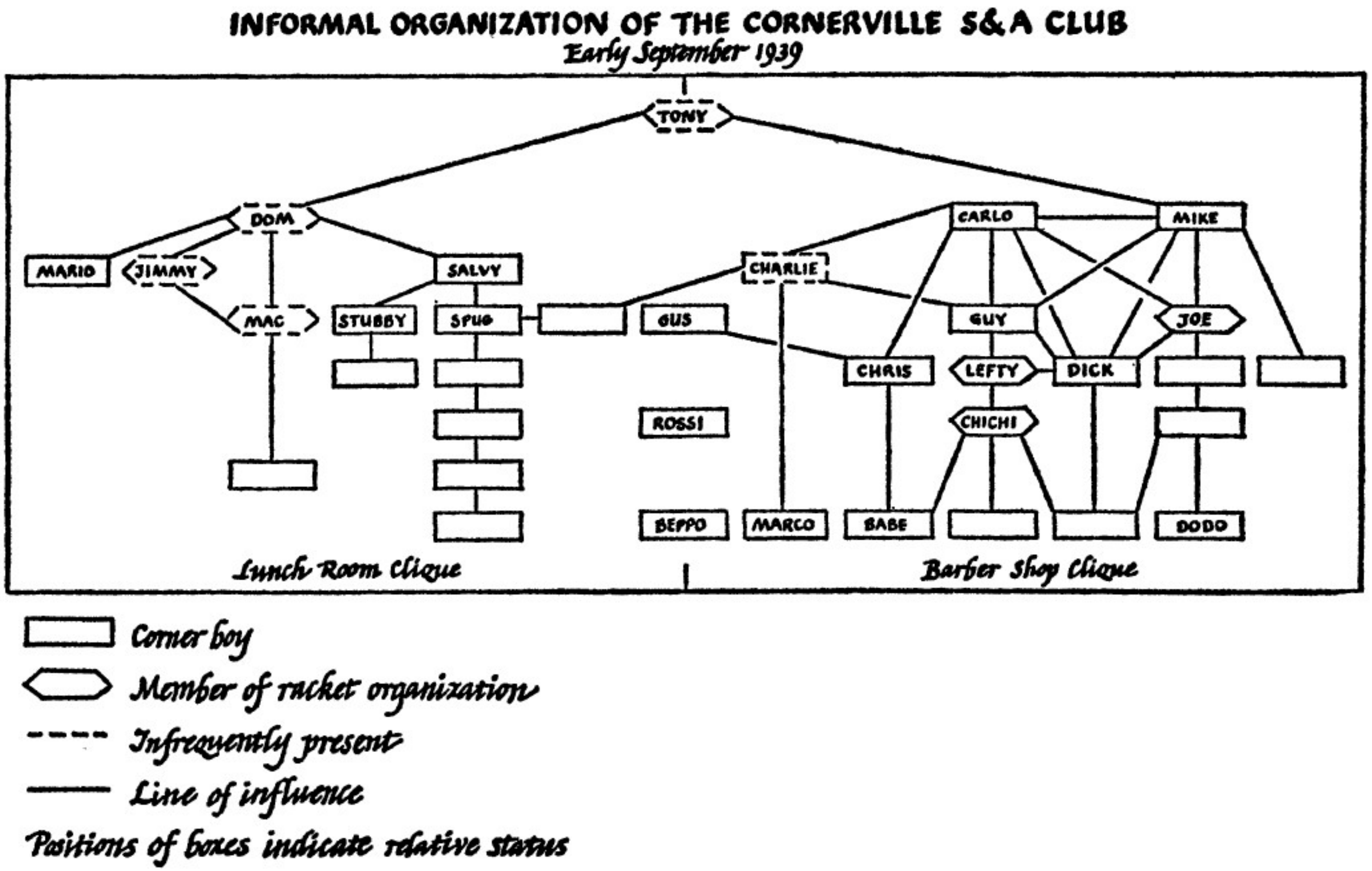
but no order between pear and quince as $7 > 2$ but $4 < 8$.

$$\begin{tikzcd} & \text{apple } (8,9) \arrow[dl] \arrow[dr] & \\ \text{pear } (7,4) & & \text{quince } (2,8) \end{tikzcd}$$

Unordered is not the same as equal.

When are partial orders useful summaries?

Social hierarchy[3,14]



Model for mutations[1,15], seriation in archaeology[4,11,12]
Potentially any ordered list data[7,8]

Notation

$\mathcal{M} = \{1, 2, \ldots, M\}$ is universe of choices.

$$\mathcal{M} = (\text{apple}, \text{pear}, \text{quince})$$

$\succ_h$ is a set of order relations on items in $\mathcal{M}$

$$\succ_h = \{\text{apple} \succ_h \text{pear},\ \text{apple} \succ_h \text{quince}\}$$

$h = (\mathcal{M}, \succ_h)$ is a *partial order* or *poset*.

- irreflexive and strict: no $j \succ_h j$ and no $j_1 \sim_h j_2$
- anti-symmetric: if $j_1 \succ_h j_2$ then not $j_2 \succ_h j_1$
- transitive: if $j_1 \succ_h j_2$ and $j_2 \succ_h j_3$ then $j_1 \succ_h j_3$

See [7] for weak POs with ties.

Let $\mathcal{H}_\mathcal{M}$ be the set of all posets on $\mathcal{M}$.

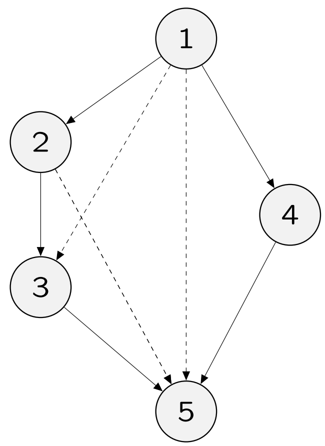
Complete order: $h \in \mathcal{C}_\mathcal{M} \ \Rightarrow j_1 \succ_h j_2$ or $j_2 \succ_h j_1 \quad \forall j_1 \neq j_2$.

Empty order: $h = (\mathcal{M}, \emptyset)$

Suborders: Let $\mathcal{B}_{\mathcal{M}}$ be all non-empty subsets of $\mathcal{M}$.

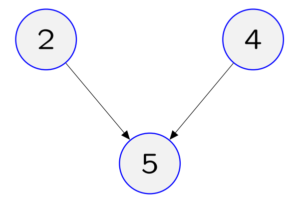
If $h = (\mathcal{M}, \succ_h)$ and $S \in \mathcal{B}_{\mathcal{M}}$ then $h[S] = (S, \succ_h)$ is *suborder*

$$j_1 \succ_h j_2 \text{ for } j_1, j_2 \in S.$$



Partial order h

Sub-order $h[S]$



$S = \{2, 4, 5\}$

Plotting: POs $1 \leftrightarrow 1$ *transitively closed* Directed Acyclic Graphs.

$$h = (S, \succ_h) \leftrightarrow g = (V, E)$$

Transitive reduction: remove edges implied by transitivity (dashed). Plot reduction as easier to read.

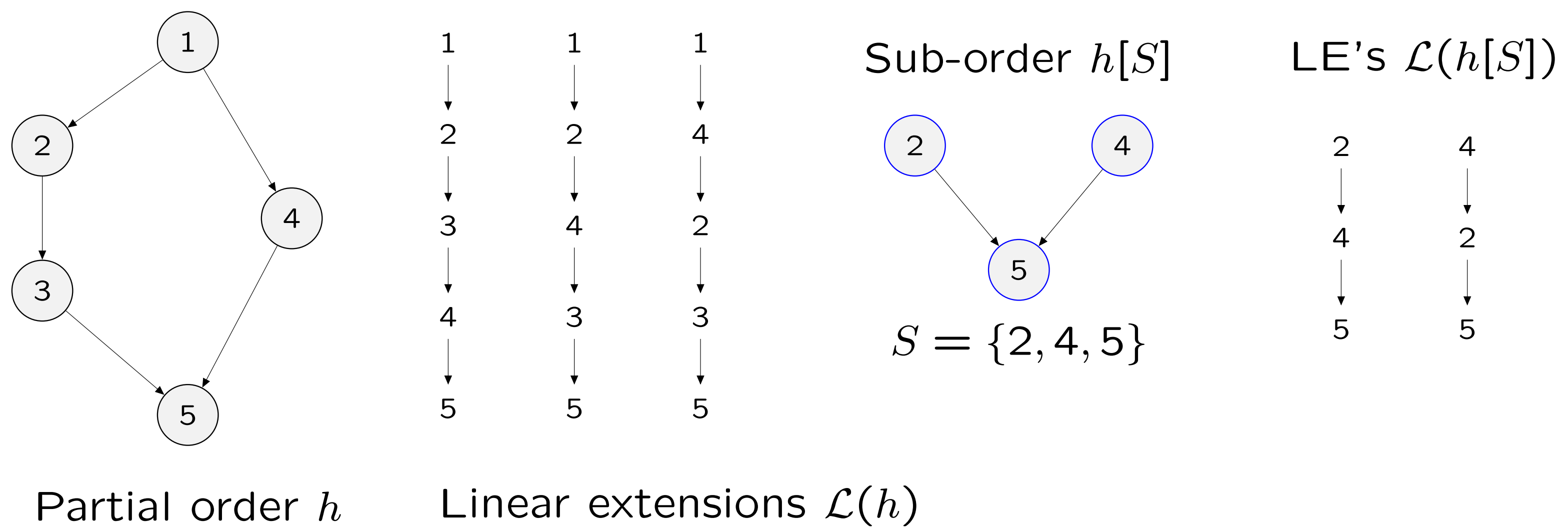
Depth $d(h)$, length of longest chain. Above $d(h) = 4$, $d(h[S]) = 2$.

If $h = (\mathcal{M}, \emptyset)$ then $d(h) = 1$ and if $h \in \mathcal{C}_{\mathcal{M}}$ then $d(h) = M$.

Linear extension of $h = (S, \succ_h)$ is a "completion" of h:

$$(S, \succ_\ell) \in \mathcal{C}_S \quad \text{satisfying} \quad j_1 \succ_h j_2 \Rightarrow j_1 \succ_\ell j_2$$

Denote by $\mathcal{L}[h]$ the set of all linear extensions of $h = (S, \succ_h)$.



Intersection order: if $\ell_i \in \mathcal{C}_\mathcal{M}$, $i = 1, \ldots, N$ are complete orders

$$h = \left(\mathcal{M}, \cap_{i=1}^N \succ_{\ell_i}\right)$$

is the *intersection order* and we write $h = \cap_i \ell_i$.

If $\ell_i \in \mathcal{C}_{S_i}$, $S_i \in \mathcal{B}_\mathcal{M}$ so ℓ_i, $i = 1, \ldots, N$ are complete suborders,

$$\succ_h = \{j_1 \succ j_2 : \exists\ i \in [N],\ j_1 \succ_{\ell_i} j_2\ \&\ \nexists\ i' \in [N],\ j_2 \succ_{\ell_{i'}} j_1\}.$$

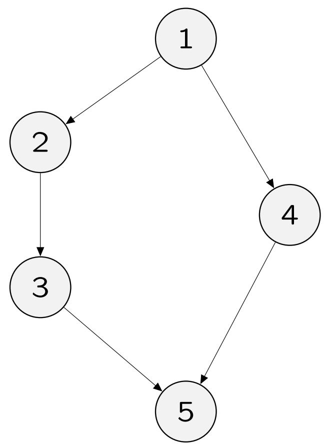
"attested and not contradicted".

Dimension of a partial order

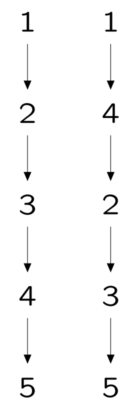
If for some PO $h \in \mathcal{H}_S$ and complete orders $L \subseteq \mathcal{C}_S$ we have

$$\bigcap_{\ell \in L} \ell = h$$

then L is a *realiser* for h.



Partial order h



Realiser for h

Dimension: $D(h)$ is the number of elements in the realiser containing the smallest number of orders.

The example above has $D(h) = 2$.

The dimension of partial order on M items is at most $\lfloor M/2 \rfloor$ [5]. This will help us parameterise our prior for partial orders.

Noise-free observations ... are linear extensions of suborders

Assessor has unknown true PO $h = (\mathcal{M}, \succ_h)$.

Given *choice set* $S \in \mathcal{B}_\mathcal{M}$

- assessor forms suborder $h[S]$;
- returns y chosen uniformly at random from $\mathcal{L}[h[S]]$;

The **likelihood for a single list**:

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

Interpretation 1: *random queue model*.

Suppose $|S| = m$ and $X_t \in \mathcal{L}[h]$ is a Markov chain with $t \geq 0$.

Say $X_t = \ell$ with $\ell = (\ell_1, \ldots, \ell_m)$ in ordered list notation:

- pick $i \in \{1, \ldots, m-1\}$ at random
- swap ℓ_i, ℓ_{i+1} giving ℓ'
- if $\ell' \in \mathcal{L}[h]$ set $X_{t+1} = \ell'$ else $X_{t+1} = \ell$

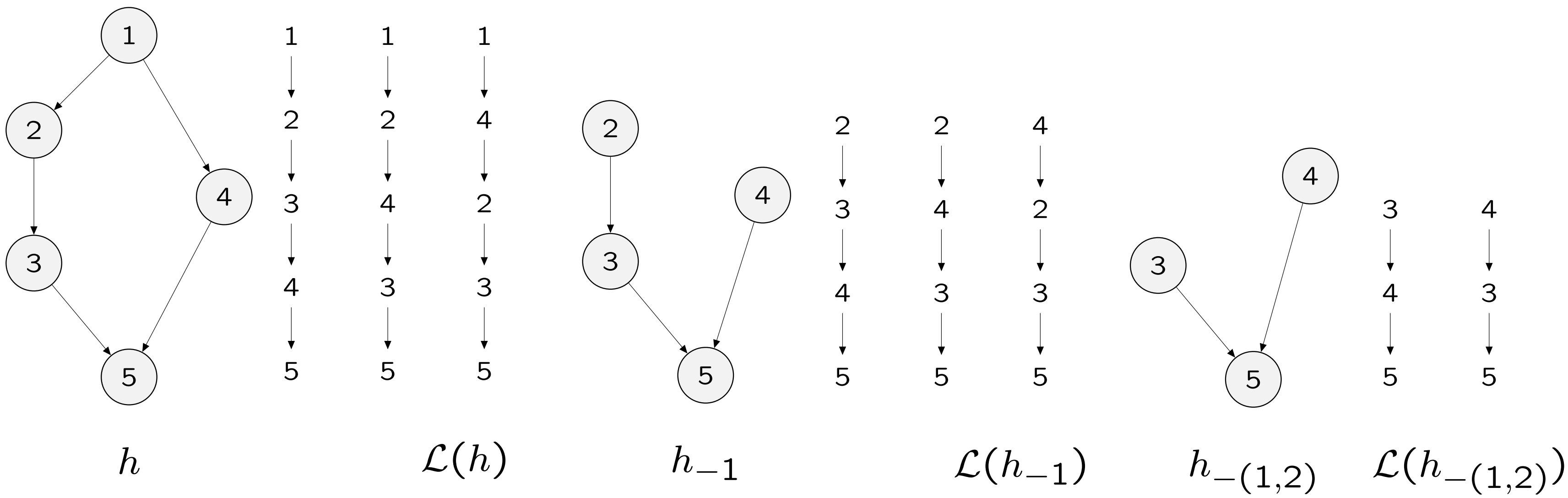
Irreducible, aperiodic, doubly stochastic so $X_t \xrightarrow{D} p_S(\cdot|h)$.

Interpretation 2: *sequential choice model*.

Suppose $|S| = m$ and $y = (y_1, \ldots, y_m)$ (order in list notation).

Let $\mathcal{L}_j[h] = \{\ell \in \mathcal{L}[h] : \max(\ell) = j\}$ be all LEs with j at top.

Useful fact: $|\mathcal{L}_{y_1}[h]| = |\mathcal{L}[h[y_{2:m}]]|$



$$p_S(y|h) = \prod_{i=1}^{m-1} q_{y_{i:m}}(y_i | h[y_{i:m}]) \tag{2}$$

$$= \frac{|\mathcal{L}_{y_1}[h]|}{|\mathcal{L}[h]|} \times \frac{|\mathcal{L}_{y_2}[h[y_{2:m}]]|}{|\mathcal{L}[h[y_{2:m}]]|} \times \prod_{i=3}^{m-1} \frac{|\mathcal{L}_{y_i}[h[y_{i:m}]]|}{|\mathcal{L}[h[y_{i:m}]]|}$$

$p(1, 2, 3, 4, 5|h) = 1 \times 2/3 \times 1/2 \times 1 \times 1$

Full likelihood - noise free model

If choice sets $S_i \in \mathcal{B}_{\mathcal{M}},\ i = 1, \ldots, N$ are given and $Y_i \sim p_{S_i}(\cdot|h)$ with $h = (\mathcal{M}, \succ_h)$ then

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

with $Y = (Y_1, \ldots, Y_N)$.

The MLE

If $S_i = \mathcal{M},\ i = 1, \ldots, N$ so assessor ranks full choice set then

$$\widehat{h}_N = \bigcap_{i=1}^{N} Y_i$$

is the MLE. Adding a relation will conflict the data and removing any will increase $\mathcal{L}[h]$ (decrease LKD).

MLE is *consistent* ($\lim_{N\to\infty} \widehat{h}_N \overset{P}{\to} h$) as realiser is subset of LEs and the probability Y contains a realiser $\to 1$ as $N \to \infty$.

Exercise: consider general S_i. Show the "attested and not contradicted" rule is not the MLE but is consistent if every pair j_1, j_2 appears infinitely often across all $S_i,\ i \geq 1$.

Calculating $|\mathcal{L}[h]|$

Computing $|\mathcal{L}[h]|$ is #P-complete[2]. Fast for $m \leq 20$.

Feasible for $m \leq 45$ using `lecount()` [9].

For $m > 45$ restrict to vertex series-parallel PO[8,16] - linear time.

Observation model with noise: Data not perfect LEs[13,14].

Queue jumping (QJ) up queue: wp p select next UAR,

$$p_S(y|h,p) = \prod_{j=1}^{m-1} \left(\frac{p}{m-j+1} + (1-p) \frac{|\mathcal{L}(h[y_{j+1:m}])|}{|\mathcal{L}(h[y_{j:m}])|} \right)$$

Now $p_S(y|h,p) > 0$ for all $y \in \mathcal{P}_{\mathcal{M}}$ and $\hat{h}_N \xrightarrow{P} (\mathcal{M}, \emptyset)$ is no longer MLE or consistent if $p > 0$.

Posterior for h, p:

For $i \in [N]$, choice sets $S_i \in \mathcal{B}_{\mathcal{M}}$ and order data $Y_i = (S_i, \succ_{Y_i})$,

$$Y = (Y_1, \ldots, Y_N), \quad p_{S_{1:N}}(Y|h,p) = \textstyle\prod_{i=1}^{N} p_{S_i}(Y_i|h,p).$$

Give priors for $h \in \mathcal{H}_{\mathcal{M}}$ and $p \in [0,1]$ then

$$\pi_{S_{1:N}}(h, p|y) \propto \pi_{\mathcal{M}}(h)\, \pi(p)\, p_{\mathcal{M}}(Y|h,p).$$

Prior for PO - background requirements

(1) *Marginal Consistency* for $\pi_{\mathcal{M}}(h)$: for $S \in \mathcal{B}_{\mathcal{M}}$ and $g \in \mathcal{H}_S$,

$$h \sim \pi_{\mathcal{M}}(\cdot) \;\Rightarrow\; h[S] \sim \pi_S(\cdot).$$

Example: Suppose $\mathcal{M} = \{1, 2, 3\}$ and $S = \{1, 2\}$.

Let $g = (S, \{1 \succ 2\})$. MC requires $\Pr\{h[S] = g\} = \pi_S(g)$.

Get $h[S] = g$ iff $\succ_h$ is in the preimage

$$h(g) = \{\{1 \succ 2 \succ 3\}, \{1 \succ 3 \succ 2\}, \{3 \succ 1 \succ 2\}, \{1 \succ 2\}, \{1 \succ 2,\ 1 \succ 3\}\}$$

so for MC we require

$$\sum_{h \in h(g)} \pi_{\mathcal{M}}(h) = \pi_S(g)$$

MC need not hold if we simply write down $\pi_S(\cdot)$ for each $S \in \mathcal{B}_{\mathcal{M}}$.

(2) *Control of depth*: seek prior non-informative WRT PO depth. The depth $d(h)$ is an important object of inference. Prior should be non-informative WRT key scientific hypothesis.

Prior for PO - why not just uniform?

Marginal consistency: $U(\mathcal{H}_{\mathcal{M}})$ is not MC.

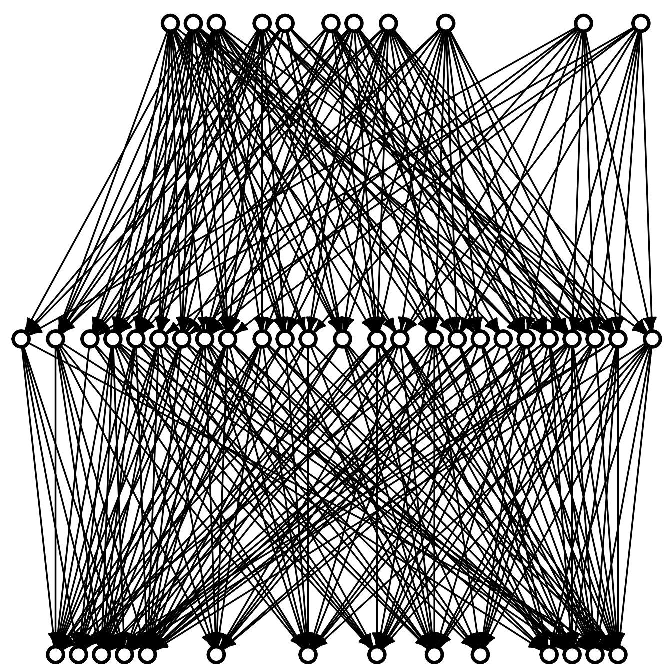
Example: There are 19 PO's if $\mathcal{M} = \{1, 2, 3\}$ and 3 if $S = \{1, 2\}$.

If $h \sim U(\mathcal{H}_{\mathcal{M}})$ then get $h[S] = (S, \{1 \succ 2\})$ if $\succ_h$ is one of

$$\{1 \succ 2 \succ 3\}, \{1 \succ 3 \succ 2\}, \{3 \succ 1 \succ 2\}, \{1 \succ 2\}, \{1 \succ 2,\ 1 \succ 3\}$$

so $\Pr\{h[S] = (S, \{1 \succ 2\})\} = 5/19 \neq 1/3$.

Prior depth distribution: recall seek prior $d(h) \sim U\{1, \ldots, M\}$. But if $h \sim U(\mathcal{H}_{\mathcal{M}})$ then[10] $\Pr(d(h) = 3) \to 1$ as $M \to \infty$.



Prior for PO - Recall Gumbel construction[6,18] for PL

Gumbel CDF

$$F(g) = \exp(-\exp(-g)), \;\; g \in \mathbb{R}.$$

Suppose for $j \in \mathcal{M}$ covariates x_j inform status of actor/item j in hierarchy/poset.

Let $\alpha_j = x_j^T \beta$ with $\beta \in \mathbb{R}^{d_\beta}$ be linear predictor.

Take $V_j \sim F$ iid for $j \in S$ and

$$G_j = V_j + \alpha_j$$

and $G = (G_1, \ldots, G_m)$. If $y = y(G)$ is the complete order

$$j_1 \succ_y j_2 \Leftrightarrow G_{j_1} > G_{j_2}$$

then $y \sim \mathrm{PL}(\alpha, S)$:

$$\Pr(y(G) = y|\alpha) = \prod_{i=1}^{m-1} \frac{e^{\alpha_{y_i}}}{\sum_{j=i}^{m} e^{\alpha_{y_j}}}.$$

Prior for PO (inspired by [17])

Fix $K \geq 1$ and $0 \leq \rho < 1$. Latent variables for choice $j \in \mathcal{M}$:

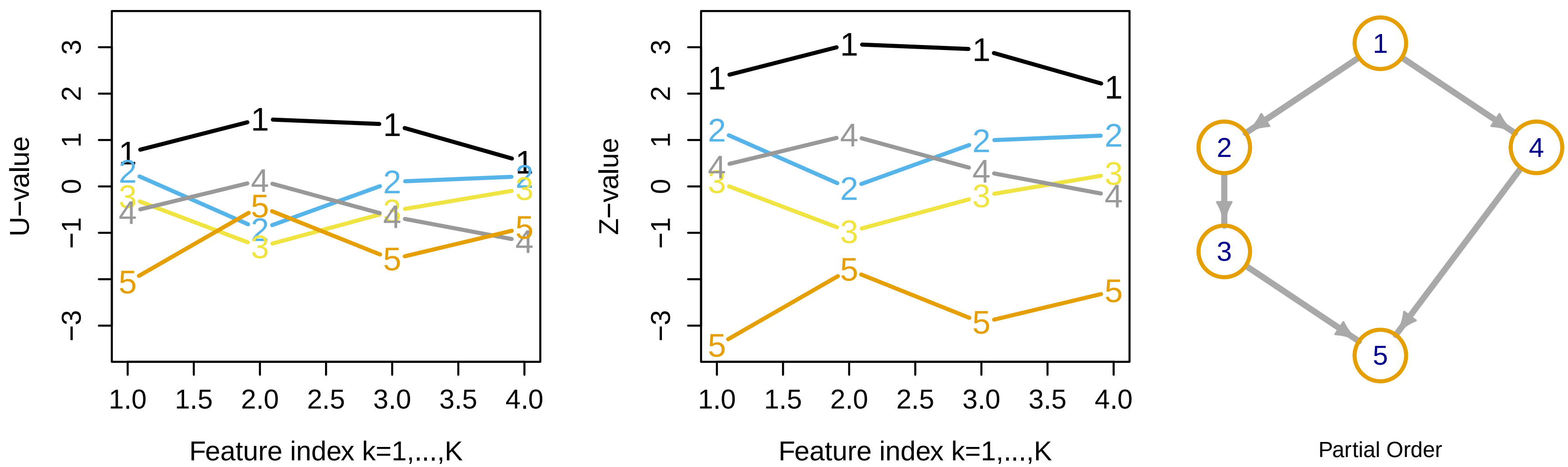
$$(U_{j,1}, \ldots, U_{j,K}) \sim N(0_K, \Sigma^{(\rho)})$$

with $\Sigma^{(\rho)}$ a $K \times K$ covariance with $\Sigma^{(\rho)}_{k,k} = 1$ and $\Sigma^{(\rho)}_{k,k'} = \rho,\ k \neq k'$;

Linear predictor $\alpha_j = x_j^T \beta,\ j \in \mathcal{M}$ and features

$$Z_{j,:} = F^{-1}(\Phi(U_{j,:})) + \alpha_j 1_M;$$

$h(Z) = (\mathcal{M}, \succ_Z)$ has $j_1 \succ_Z j_2 \ \Leftrightarrow Z_{j_1,k} > Z_{j_2,k}\ \forall\ k = 1, \ldots, K.$



Prior on h: $\qquad \pi_{\mathcal{M},K}(h|\rho,\alpha) = \Pr(h(Z(U,\alpha)) = h|\rho,\alpha)$

Properties of Latent Variable Prior

Marginally consistent for h: for all $S \in \mathcal{B}_{\mathcal{M}}$

$$h \sim \pi_{\mathcal{M},K}(\cdot|\rho,\alpha) \;\Rightarrow\; h[S] \sim \pi_{S,K}(\cdot|\rho,\alpha).$$

Proof: The rows $Z_{j,:}$ of Z are independent. Removing rows (ie going from h to $h[S]$) doesn't change the distribution of the relations between the rows remaining.

$$h(Z) \sim \pi_{\mathcal{M},K},\; h(Z_{S,:}) \sim \pi_{S,K}$$

$$\text{but } h(Z)[S] = h(Z_{S,:}) \text{ so } h(Z)[S] \sim \pi_{S,K}. \qquad \square$$

Gives PL when $K = 1$: if $h \sim \pi_{\mathcal{M},1}(\cdot|\rho,\alpha)$ then $h \sim \mathrm{PL}(\alpha, S)$

Proof: If $K = 1$ then $h \in \mathcal{C}_S$ is a complete order and

$$(Z_{1,1}, \ldots, Z_{M,1}) \sim (G_1, \ldots, G_M)$$

in the Gumbel construction of PL. $\square$

Properties of Latent Variable Prior...(cont)

Expresses any PO: $K \geq \lfloor M/2 \rfloor \;\Rightarrow \pi_{\mathcal{M},K}(h|\rho,\alpha) > 0$ all $h \in \mathcal{H}_S$.

Proof: (1) The rule $h = h(Z)$ is the same as taking the intersection of the column orders of Z.

If $\ell_k = h(Z_{:,k})$ then $j_1 \succ_{\ell_k} j_2 \;\Leftrightarrow Z_{j_1,k} > Z_{j_2,k}$ and $h = h(Z)$ then

$$h = \bigcap_{k=1}^{K} \ell_k.$$

(2) Every $h \in \mathcal{H}_\mathcal{M}$ has a realiser $\ell_1, \ldots, \ell_D$ with $D \leq \lfloor M/2 \rfloor$.

(3) If $K = D$ then we choose the Z-values in each column so that $h(Z_{:,k}) = \ell_k$ (Z's have the same order as ℓ_k). If $K > D$ then repeat column D - copies dont change intersection. $\square$

Depth dbn: experiment shows taking a prior $\rho \sim \text{Beta}(1, 1/6)$ gives depth $d(h)$ uniform (approx).

Posterior for Partial Orders (using $\alpha = X\beta$)

$$\begin{aligned}
\pi_{S_{1:N},K}(U,\beta,\rho,p|Y) &\propto \pi_{B,R,P}(\beta,\rho,p)\,\pi(U|\rho)\; p_{S_{1:N}}(Y|h(Z(U,X\beta)),p) \\
&\propto N(\beta; 0_{d_\beta}, \mathbb{I}_{d_\beta}) \;\times\; \text{Beta}(\rho; 1, 1/6) \;\times\; \text{Beta}(p; 1, 9) \\
&\quad \times \quad \prod_{j\in\mathcal{M}} N(U_{j,:}; 0_K, \Sigma_\rho) \\
&\quad \times \quad \prod_{i=1}^{N} p_{S_i}(Y_i|h(Z(U,X\beta)),p).
\end{aligned}$$

These subjective priors are typical for the applications we have seen to date.

Sample this using simple classical random-walk MCMC.

Get samples $U^{(t)}, \beta^{(t)}, \rho^{(t)}, p^{(t)},\ t = 1,\dots,T$

Now to get PO samples from the posterior set

$$h^{(t)} = h(Z(U^{(t)}, X\beta^{(t)}))$$

and this gives samples from the marginal PO posterior

$$h^{(t)} \xrightarrow{D} \pi_{S_{1:N},K}(\cdot|Y).$$

*References

[1] Niko Beerenwinkel, Nicholas Eriksson, and Bernd Sturmfels. "Conjunctive bayesian networks". In: *Bernoulli* 13.4 (2007), pp. 893–909.

[2] Graham Brightwell and Peter Winkler. "Counting linear extensions". In: *Order* 8.3 (1991), pp. 225–242.

[3] Morris F. Friedell. "Organizations as Semilattices". In: *American Sociological Review* 32.1 (1967), pp. 46–54. ISSN: 00031224. URL: `http://www.jstor.org/stable/2091717` (visited on 09/15/2024).

[4] Aristides Gionis et al. "Algorithms for discovering bucket orders from data". In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2006, pp. 561–566.

[5] Toshio Hiraguchi. "On the dimension of partially ordered sets". In: *The science reports of the Kanazawa University* 1 (2 1951), pp. 77–94.

[6] David R. Hunter. “MM algorithms for generalized Bradley-Terry models”. In: *The Annals of Statistics* 32.1 (2004), pp. 384–406. DOI: 10.1214/aos/1079120141. URL: https://doi.org/10.1214/aos/1079120141.

[7] Chuxuan Jiang and Geoff K. Nicholls. *Non-Parametric Bayesian Inference for Partial Orders with Ties from Rank Data observed with Mallows Noise*. https://arxiv.org/abs/2408.14661. 2024. arXiv: 2408.14661 [stat.ME]. URL: https://arxiv.org/abs/2408.14661.

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

[9] Kustaa Kangas et al. “Counting Linear Extensions of Sparse Posets”. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. IJCAI’16. New York, New York, USA: AAAI Press, 2016, pp. 603–609. ISBN: 9781577357704. URL: `https://www.cs.helsinki.fi/u/jwkangas/lecount/`.

[10] Daniel J Kleitman and Bruce L Rothschild. “Asymptotic enumeration of partial orders on a finite set”. In: *Transactions of the American Mathematical Society* 205 (1975), pp. 205–220.

[11] H. Mannila. “Finding Total and Partial Orders from Data for Seriation”. In: *Discovery Science*. Ed. by J.-F. Boucault. Vol. 5255. LNAI. Berlin Heidelberg: Springer-Verlag, 2008, pp. 16–25.

[12] Heikki Mannila and Christopher Meek. “Global Partial Orders from Sequential Data”. In: *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’00. Boston, Massachusetts, USA: Association for Computing Machinery, 2000, pp. 161–168. ISBN: 1581132336. DOI: `10.1145/347090.347122`. URL: `https://doi.org/10.1145/347090.347122`.

[13] Geoff K Nicholls and Alexis Muir Watt. "Partial Order Models for Episcopal Social Status in 12th Century England". In: *IWSM 2011* (2011), p. 437.

[14] Geoff K. Nicholls et al. "Bayesian inference for partial orders from random linear extensions: Power relations from 12th century royal acta". In: *The Annals of Applied Statistics* 19.2 (2025), pp. 1663–1690. DOI: 10.1214/24-AOAS2002. URL: https://doi.org/10.1214/24-AOAS2002.

[15] Thomas Sakoparnig and Niko Beerenwinkel. "Efficient sampling for Bayesian inference of conjunctive Bayesian networks". In: *Bioinformatics* 28.18 (2012), pp. 2318–2324. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bts433. eprint: https://academic.oup.com/bioinformatics/article-pdf/28/18/2318/698477/bts433.pdf. URL: https://doi.org/10.1093/bioinformatics/bts433.

[16] Jacobo Valdes, Robert E. Tarjan, and Eugene L. Lawler. "The Recognition of Series Parallel Digraphs". In: *SIAM Journal on Computing* 11.2 (1982), pp. 298–313. DOI: 10.1137/0211023. eprint: https://doi.org/10.1137/0211023. URL: https://doi.org/10.1137/0211023.

[17] Peter Winkler. "Random orders". In: *Order* 1.4 (1985), pp. 317–331.

[18] John I. Yellott. "The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution". In: *Journal of Mathematical Psychology* 15.2 (1977), pp. 109–144. ISSN: 0022-2496. DOI: `https://doi.org/10.1016/0022-2496(77)90026-8`. URL: `https://www.sciencedirect.com/science/article/pii/0022249677900268`.