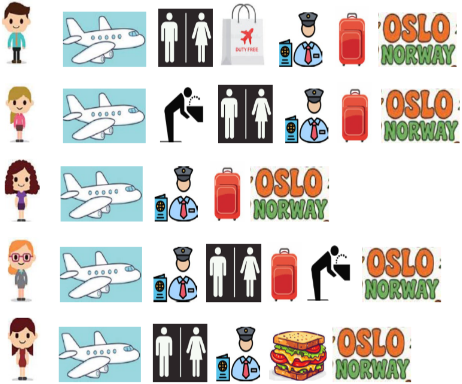
Bayesian Inference for Partial Orders from Rank-Order Data (part II)

Kate Lee* Geoff, Nicholls+

University of Auckland, NZ*
University of Oxford, UK+.

PrefStat: Preference Statistics summer school
30 June - 4 July, 2025 (University of Oslo)

velkommen til Oslo. Which path did you take?

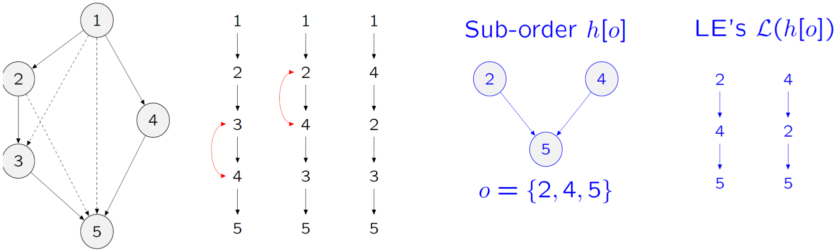


....

- Revision
- Bayesian inference framework
- R-simulation
- Other PO work
- Summary

Queues and partial orders (PO)

Queue of M actors $[M] = \{1, \ldots, M\}$ constrained by PO $h \in \mathcal{H}_{[M]}$.
$S \subseteq [M]$ be a suborder with the size of $m = |S|$.



Partial order h Linear extensions $\mathcal{L}(h)$

A random queue $y_{1:m} = (y_1, ..., y_m)$ occurs at random

$$p(y_{1:m}|h) = \frac{1}{|\mathcal{L}(h[y_{1:m}])|} \mathbb{1}_{y_{1:m} \in \mathcal{L}_{[h(y_{1:m})]}}$$

where $|\mathcal{L}(h[y_{1:m}])|$ is a number of linear extensions respecting PO h.
e.g., $|\mathcal{L}(h[(1, 2, 3, 4, 5)])| = 3$ and $|\mathcal{L}(h[(2, 4, 5))])| = 2$.

Noise free likelihood

A random queue $y_{1:m}$ respecting $h[S]$ is also expressed sequentially. e.g, top-down

$$p(y_{1:m}|h) = \frac{1}{|\mathcal{L}(h[y_{1:m}])|}\mathbb{1}_{y_{1:m}\in\mathcal{L}_{[h(y_{1:m})]}} = \prod_{i=1}^{m-1}\frac{|\mathcal{L}_{y_i}[h[y_{i:m}]]|}{|\mathcal{L}[h[y_{i:m}]]|}$$

where
$\mathcal{L}[h[y_{i:m}]]$ is a list of linear extensions of $[y_{i:m}]$
$\mathcal{L}_{y_i}[h[y_{i:m}]]$ is a list of linear extensions of $[y_{i:m}]$ starting with y_i.

This is the noise-free likelihood. With a prior $\pi(h)$, the posterior density for h is

$$\pi(h|y_{1:m}) \propto \pi(h)p(y_{1:m}|h)\,.$$

Noisy lists & likelihood

Queue jumping (QJ) noise: With the noise probability p, the next one is selected at random. e.g., top-down QJ

$$p^{(D)}(y_{1:m}|h,p) = \prod_{i=1}^{m-1} \frac{p}{m-i+1} + (1-p)\frac{|\mathcal{L}_{y_i}[h[y_{i:m}]]|}{|\mathcal{L}[h[y_{i:m}]]|}$$

Y_i observed on corresponding subsets $S_i \subseteq [M]$ for $i = 1, ..., N$. With a prior $\pi(h)$ and $\pi(p)$, the **posterior** distribution is

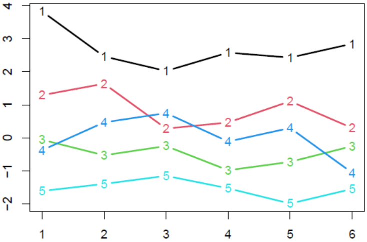
$$\pi(h,p|Y_1,...,Y_N) \propto \pi(h)\,\pi(p) \prod_{i=1}^{N} p^{(D)}(Y_i|h[S_i],p)\,.$$

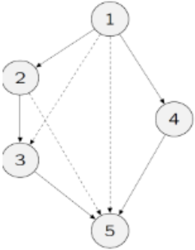
Latent variable U [1]

K-dimensional latent variable $U_{j,:} \in \mathbb{R}^K$ for each object $j \in [M]$ defines order relations $\succ_U$.

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

This is called an intersection order $h(U) = \bigcap_{k=1}^{K} h(U_{:,k})$ as $h(U)$ contains just the order relations shared by all $h(U_{:,k}),\ k = 1, \ldots, K$.

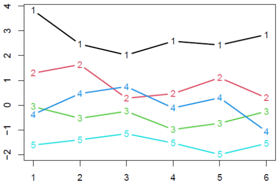




Latent variable U [1]

K-dimensional latent variable $U_{j,:} \in \mathbb{R}^K$ for each object $j \in [M]$ defines order relations $\succ_U$.

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



$$U_{j,:} \sim N([0,..,0], \Sigma_\rho) \text{ independent for each } j \in M$$

where $Diag(\Sigma_\rho) = 1$ and off-diagonals are ρ.

- Deeper order with higher ρ & Lower order with lower ρ.

- K is at least $M/2$ (conservative choice) [2]

Latent variable η [1]

With d-dimensional covariates X and their effects β,

$$\eta = X\beta + U.$$

K-dimensional latent variable $\eta \in \mathbb{R}^K$ for each object $j \in [M]$ defines order relations $\succ_\eta$.

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

This is called an intersection order $h(\eta) = \bigcap_{k=1}^{K} h(\eta_{:,k})$ as $h(\eta)$ contains just the order relations shared by all $h(\eta_{:,k}), \ k = 1, \ldots, K$.

Alternatively, each $\sigma(\eta_{:,k}) \sim PL(\alpha_{\mathcal{M}} + X\beta)$ if $U_{j,:} \sim Gumbel(\alpha_{\mathcal{M}})$.

In this tutorial, for simplicity, $\eta = U$. No covariate.

In this tutorial, we will consider the two models.

1. PO with a fixed K

$$\pi(U, p, \rho | Y_1, ..., Y_N) \propto \pi(\rho)\pi(p)\pi(U|\rho)L(h(U); Y_1, ...Y_N)\,.$$

2. PO with a variable K

$$\pi(U, p, \rho, K | Y_1, ..., Y_N) \propto \pi(\rho)\pi(p)\pi(U|\rho, K)\pi(K)L(h(U); Y_1, ...Y_N)\,.$$

where $L(h(U), p; Y_1, ...Y_N) = \prod_{i=1}^{N} p^{(D)}(Y_i | h[S_i], p)$.

Bayesian inference for h (fixed K)

Latent variable $U_{j,:} \sim N([0,..,0],\Sigma_\rho)$ where $Diag(\Sigma_\rho) = 1$ and off-diagonals are ρ.

The random PO $h(U)$ has prior distribution

$$\pi(h|\rho) = \int \mathbb{1}_{h(U)=h}\pi(U|\rho)dU$$

where $\pi(U|\rho) = \prod_{j=1}^{M} N(U_{j,:};[0,...,0],\Sigma_\rho)$.

With priors $\pi(\rho)$, $\pi(U|\rho)$, $\pi(p)$, the joint posterior with latent variable U, is

$$\pi(U,p,\rho|Y_1,...,Y_N) \propto \pi(p)\pi(\rho)\pi(U|\rho)\prod_{i=1}^{N} p^{(D)}(Y_i|h(U),p)\,.$$

Bayesian inference for h (fixed K)

The joint posterior is

$$\pi(U, p, \rho | Y_1, ..., Y_N) \propto \pi(p)\pi(\rho)\pi(U|\rho) \prod_{i=1}^{N} p^{(D)}(Y_i|h(U), p) \,.$$

The posterior simulation using the MCMC method is a natural approach.

The likelihood $p^{(D)}(Y_i|h(U), p)$ is a discrete problem and no derivatives with respect to parameters are available. No benefit from the gradient-based sampler.

The Gibbs within Metropolis-Hastings algorithm is used to simulate the conditional posterior.

Bayesian inference for h (fixed K)

At the t-th iteration, we sample the following.

1. $U^{(t+1)} \sim \pi(U|p^{(t)}, \rho^{(t)}, Y_1, ..., Y_N)$
   Set $U' = U^{(t)}$. For a randomly selected j, a proposal $U'_{j,:}|U^{(t)}_{j,:} \sim q_u$ is accepted according to

$$a_{U^{(t)} \to U'} = \left\{1, \frac{L(h(U'), p^{(t)}; Y_1, ... Y_N)\pi(U'|\rho^{(t)})}{L(h(U^{(t)}), p^{(t)}; Y_1, ... Y_N)\pi(U^{(t)}|\rho^{(t)})} \frac{q_u(U^{(t)}|U')}{q_u(U'|U^{(t)})}\right\}$$

2. $\rho^{(t+1)} \sim \pi(\rho|U^{(t+1)}, p^{(t)}, Y_1, ..., Y_N)$
   A proposal $\rho'|\rho^{(t)} \sim q_\rho(\cdot)$ is accepted according to

$$a_{\rho^{(t)} \to \rho'} = \left\{1, \frac{L(h(U^{(t+1)}), p^{(t)}; Y_1, ... Y_N)\pi(\rho')}{L(h(U^{(t+1)}), p^{(t)}; Y_1, ... Y_N)\pi(\rho^{(t)})} \frac{q_\rho(\rho^{(t)}|\rho')}{q_\rho(\rho'|\rho^{(t)})}\right\}$$

3. $p^{(t+1)} \sim \pi(p|U^{(t+1)}, \rho^{(t+1)}, Y_1, ..., Y_N)$
   A proposal $p'|p^{(t)} \sim q_p(\cdot)$ is accepted according to

$$a_{p^{(t)} \to p'} = \left\{1, \frac{L(h(U^{(t+1)}), p'; Y_1, ... Y_N)\pi(p')}{L(h(U^{(t+1)}), p^{(t)}; Y_1, ... Y_N)\pi(p^{(t)})} \frac{q_p(p^{(t)}|p')}{q_p(p'|p^{(t)})}\right\}$$

Bayesian inference for h (variable K)

If K changes, the dimension of U changes ($M \times K$). With a prior for K, $\pi(K)$ and $\pi(U|\rho, K)$, the conditional posterior for (K, U) is

$$\pi(K, U|p, \rho, Y_1, ..., Y_N) \propto \pi(K)\pi(U|\rho, K) \prod_{i=1}^{N} p^{(D)}(Y_i|h(U), p) .$$

The reversible-jump MCMC method is implemented [3]

- Birth move $K \to K+1$; A vector v is added at the randomly chosen i-th column with a probability of q_k.
- Death move $K \to K-1$; A randomly chosen column is deleted with a probability of $1 - q_k$.

Delete Column

Insert Column

Bayesian inference for h (variable K)

Birth move $K \to K+1$ with a probability of q_k.

A vector v is added at the randomly chosen i-th column.
i.e., $U' = (U^{(t)}_{:,<i}, v, U^{(t)}_{:,\geq i})$ and $v_j = U'_{j,i}$, $j = 1, ..., M$.

Given $U'_{j,:}$ is normally distributed,

$$U'_{j,i}|U_{j,\neq i} \sim N(\Sigma_{i,\neq i}\Sigma^{-1}_{\neq i,\neq i}U'_{j,\neq i}, \Sigma_{i,i} - \Sigma_{i,\neq i}\Sigma^{-1}_{\neq i,\neq i}\Sigma_{\neq i,i})\,.$$

A proposal $(K+1, U')$ is accepted according to

$$a = \left\{1, \frac{L(h(U'), p^{(t)}; Y_1, ... Y_N)\pi(U'|K+1, \rho^{(t)})\pi(K+1)}{L(h(U^{(t)}), p^{(t)}; Y_1, ... Y_N)\pi(U^{(t)}|K, \rho^{(t)})\pi(K)} \frac{1-q_k}{q_k}\right\}$$

Bayesian inference for h (variable K)

Death move $K \rightarrow K-1$ with a probability of $1-q_k$.

A randomly chosen i-th column is deleted.
i.e., $U' = (U^{(t)}_{:,<i}, U^{(t)}_{:,>i})$

A proposal $(K-1, U')$ is accepted according to

$$a = \left\{1, \frac{L(h(U'), p^{(t)}; Y_1, \dots Y_N)\pi(U'|K-1, \rho^{(t)})\pi(K-1)}{L(h(U^{(t)}), p^{(t)}; Y_1, \dots Y_N)\pi(U^{(t)}|K, \rho^{(t)})\pi(K)} \frac{q_k}{1-q_k}\right\}$$

Bayesian inference for h (variable K)

Dimension of PO = min # of linear extensions
$\leq K$ (latent variable dimension)

i.e., K is relevant to the dimension of PO.

- If K is too small, higher prior mass on deeper orders. e.g., $K = 1$ means the total order.
- If K is too large, higher prior mass on shallow orders. e.g., $K = \infty$ means no order.

K and ρ trade-off if K is large enough. (see page 7)

Deeper order; small K & high ρ
Shallow order; high K & low ρ

Output summary

We have the posterior samples.

- p Noise probability
- ρ Order strictness; $\rho = 0$ means no order and $\rho = 1$ means the total order.
- U Latent variables; $M \times K$-matrix.
- $h(U)$ PO; $M \times M$-binary matrix
- K PO dimension relevant parameter
  K posterior does not yield the uncertainty quantification of the dimension of the true PO. However, it will put ≈ 0 mass for K less than the true PO dimension.

Output summary

From T-iterations, we get $\left(U^{(1)}, ..., U^{(T)}\right), \left(\rho^{(1)}, ..., \rho^{(T)}\right), \left(p^{(1)}, ..., p^{(T)}\right)$.

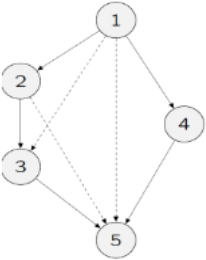
The posterior samples for PO, $(h^{(1)}, ..., h^{(T)})$ is obtained.

The PO h is a $M \times M$ matrix; $h_{j_{w1}, j_{w2}} = 1$ if $j_{w1} \succ_h j_{w2}$ and otherwise $h_{j_{w1}, j_{w2}} = 0$.

For example,

$$h = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$h_{j_{w1}, j_{w2}}$ is a probability for $j_{w1} \succ_h j_{w2}$.



Output summary

Evidence of order relationships

- The posterior mean probability for the order relationship $j_{w1} \succ j_{w2}$ is

$$\bar{h}_{j_{w1},j_{w2}} = \frac{1}{T}\sum_t h^{(t)}_{j_{w1},j_{w2}} \approx \mathbb{E}[h_{w1,w2}|Y_1,.,,,Y_N]$$

i.e., $\bar{h}_{j_{w1},j_{w2}} \approx 1$ means strong evidence for $j_{w1} \succ j_{w2}$.

- Extension to the order relationship of $S \subseteq M$. For example, evidence for $j_{w1} \succ j_{w2} \succ j_{w3}$ is estimated as $\bar{h} = \frac{1}{T}\sum_t \mathbb{1}(h^{(t)}_{j_{w1},j_{w2}} = h^{(t)}_{j_{w2},j_{w3}} = 1)$.
- Credible intervals of order relationships can be computed.
- Consensus PO shows the order relationships with significant evidence. i.e., posterior probabilities over the threshold (0.5 or 0.9).

Output summary

Summary of other parameters

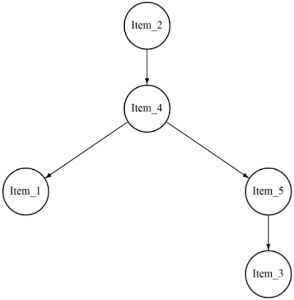
Posterior samples for ρ, p (and K) are summarized as usual.

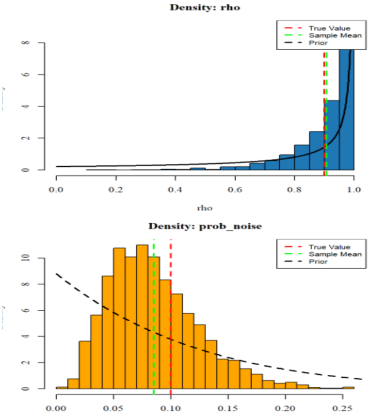
Diagnostics

Usual diagnostic techniques; rate of acceptance, trace plot or some diagnostics.

R-code

Let's implement the synthetic data and do the uncertainty quantification [1]





$$p \sim Beta(1,9),$$

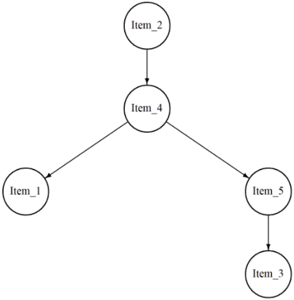
$$\rho \sim Beta(1,1/6)$$

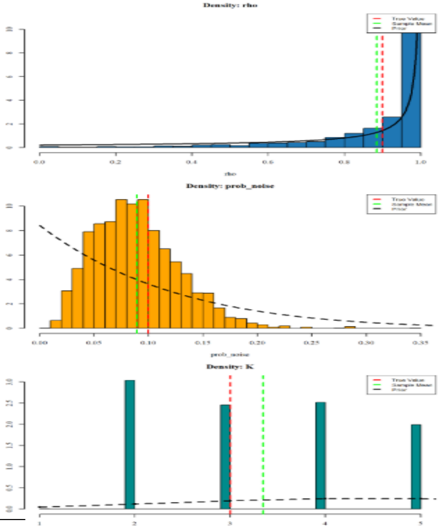
$$K = 3$$

[1] https://github.com/hollyli-dq/BayesianPO_R MCMC_Simulation_Tutorial.Rmd

Simulation study

Let's implement the synthetic data and simulate MCMC [2]





$$p \sim Beta(1, 9),$$

$$\rho \sim Beta(1, 1/6),$$

$$K \sim U[1, M]$$

[2] `https://github.com/hollyli-dq/BayesianPO_R` MCMC_Simulation_Tutorial.Rmd

Scalability

Counting LE is #P-complete task and is easily expensive with M (number of actors) and N (data size).

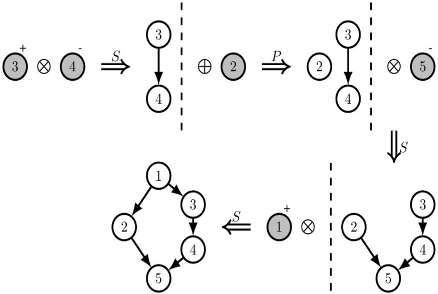
- Simpler model
- Faster LE counting algorithm
  - LEcount[3] with $O(t!M^{(t+3)})$ for a tree width t [4]

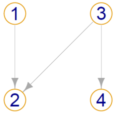
Restricted class of PO, "Vertex-Series-Parallel partial orders" [5]

[3] https://www.rforge.net/lecount/

Vertex-Series-Parallel partial orders (VSPs)

Sub-class of partial orders be formed by repeated series and parallel operations (left fig)



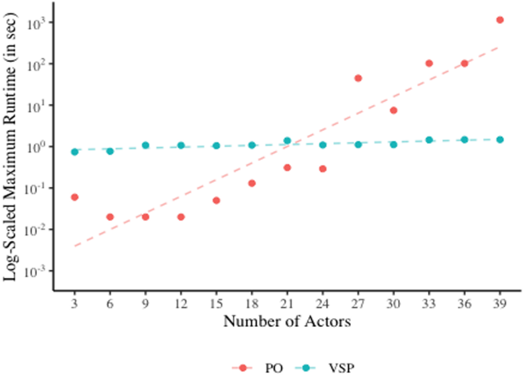


Two-dimensional POs ($K = 2$) except the forbidden graph order (right fig)

Binary decomposition trees (BDTs) representation with $O(M)$ **cost!**

With a prior for BDT, the posterior for h is formulated to do the uncertainty quantification.

Linear extension counting time comparison (Runtime for 20 lists)

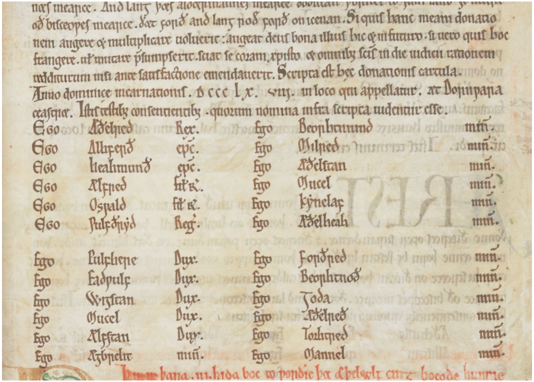


VSP counting method via the tree representation was scailable more than 200 actors while *LEcount* (PO) failed.

Other work with applications

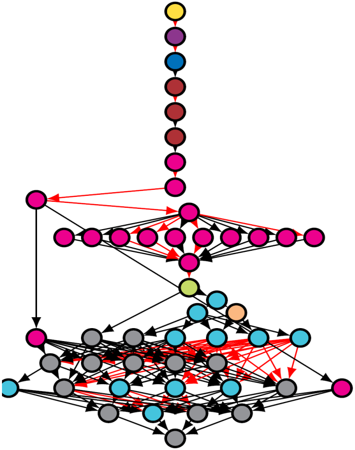
I found a clip explaining it better :)

Social hierarchy



The witness list reflects the social hierarchy.

Social hierarchy of $M = 49$ witnesses for 1134-1138 in England using VSP [6]



VSP/QJ-U model. Consensus order. Significant/strong order relations are indicated by black/red edges, respectively.

Witness lists of legal documents from 20 Anglo-Norman dioceses (yr 1080-1155) [7]

After some data cleaning; $N = 371$ lists over $T = 76$ years.

Historian's questions

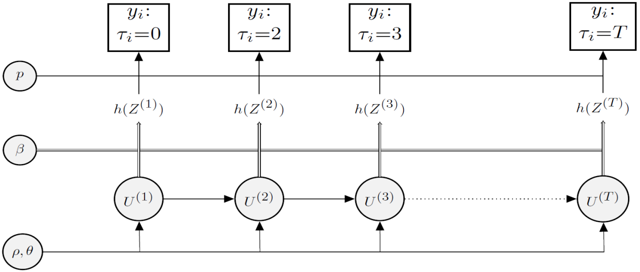
- Can we construct the social hierarchy of bishops from the lists?
- What affects ranks? e.g., seniority, individual power, individual behaviour.
- How does the social hierarchy structure change in time?

How to model?

Unknown social hierarchy need not be a total order (confirmed by historians) and this motivates us to represent the hierarchy as a **partial order**.

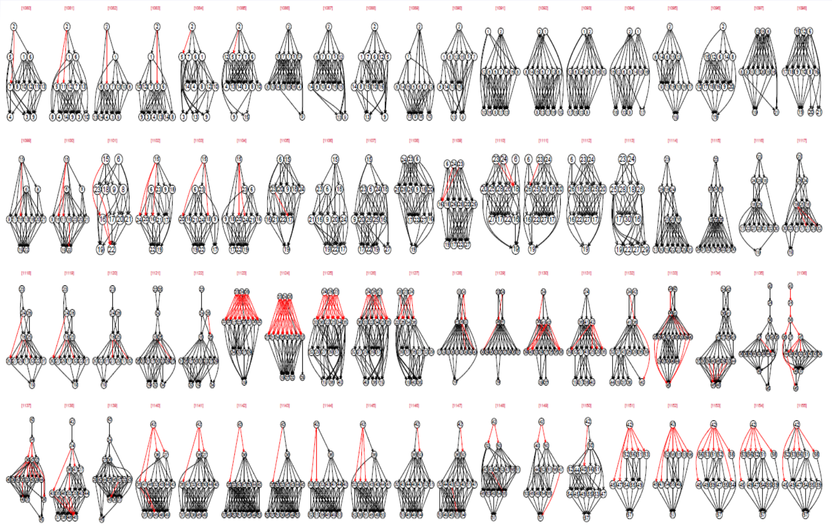
Other PO work

Time-varying PO [8]
List data y_i come with time stamp $\tau_i \in \{1, \ldots, T\}$. Extend to HMM using $U = (U^{(t)}\ t = 1, \ldots, T)$ and $U \sim VAR_{\rho,\theta}(1)$.

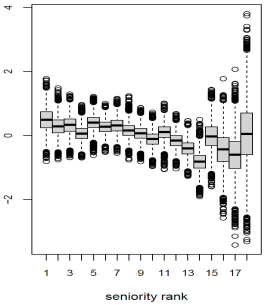


PO with covariate effect [8]
PO is $h(\eta)$ where the actor j specific covariant is X_j and latent variable is

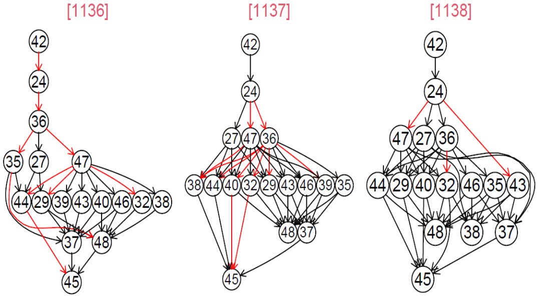
$$\eta_{j,\cdot} = U_{j,\cdot} + \beta X_j\,.$$



Consensus partial orders for all years 1080-1155. Significant (black) and strong (red) orders.



Seniority effect (β)



Consensus PO in year 1136-1138 (significant (black) and strong (red) orders.)

Summary

- PO offers flexible order structures; the biggest class of posets.
- PO includes a total order, bucket order, no order and VSP.
- Free from the total order assumption.
- VSP is an alternative option for high-dimensional PO.

Future work

- Noise model - Realistic noise, Mallows type noise, etc
- PO models for real-world problems
- Summary of uncertainty quantification
- Scalable methods

References I

[1] P. Winkler, "Random orders," *Order*, vol. 1, no. 4, pp. 317–331, 1985.

[2] T. Hiraguchi, "On the dimension of partially ordered sets," *The science reports of the Kanazawa University*, vol. 1, pp. 77–94, 2 1951.

[3] C. Jiang and G. K. Nicholls, *Bayesian inference for partial orders with ties from ranking data with a placket-luce distribution centred on random linear extensions*, 2021.

[4] K. Kangas, M. Koivisto, and S. Salonen, "A faster tree-decomposition based algorithm for counting linear extensions," *Algorithmica*, pp. 1–18, 2019.

[5] H. Mannila and C. Meek, "Global partial orders from sequential data," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '00, Boston, Massachusetts, USA: Association for Computing Machinery, 2000, pp. 161–168, ISBN: 1581132336. DOI: 10.1145/347090.347122. [Online]. Available: https://doi.org/10.1145/347090.347122.

[6] C. Jiang, G. K. Nicholls, and J. E. Lee, "Bayesian inference for vertex-series-parallel partial orders," in *Volume 216: Uncertainty in Artificial Intelligence, 31-4 August 2023, Pittsburgh, PA, USA*, 2023.

[7] 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.

[8] G. K. Nicholls, J. E. Lee, N. Karn, D. Johnson, R. Huang, and A. Muir-Watt, "Bayesian inference for partial orders from random linear extensions: Power relations from 12th century royal acta," *Ann. Appl. Stat.*, no. 2, pp. 1663–1690, 2025.