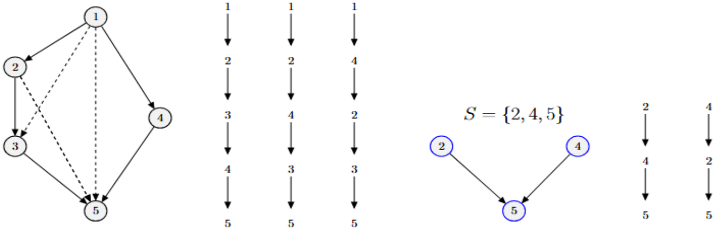
Bayesian Inference for Partial Orders from Rank-Order Data (part II)

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

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

Queues and partial orders (PO)

Queue of M actors $[M] = \{1, \dots, M\}$ 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 $\mathcal{L}(h)$ Sub-order $h[S]$ $\mathcal{L}(h[S])$

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

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

Noise free likelihood

If just $S \subseteq [M]$ are queuing, then the constraining suborder is $h[S]$. A random queue $y_{1:m}$ respecting $h[S]$ is

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

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.

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

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

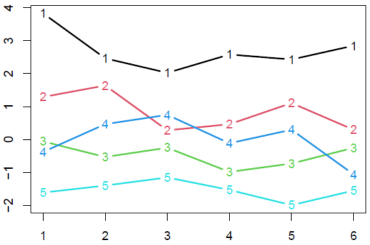
$$\pi(h,p|Y_1,...,Y_N) \propto \pi(h)\,\pi(p) \prod_{i=1}^{N} p^{(D)}(Y_i|h[S_i],p)\,.$$

Latent variable U [1]

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

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

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 $h(U_{:,k}), \ k = 1, \dots, K$.



Deeper order with higher ρ & Lower order with lower ρ.
Dimension of PO, K is at least $M/2$.

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 $h(\eta_{:,k}),\ k = 1, \ldots, K$.

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

Alternatively, each $\eta_{:,k} \sim PL(\alpha_{\mathcal{M}} + X\beta)$ if $U_{j,:} \sim Gumbel(\alpha_{\mathcal{M}})$.
In this tutorial, for simplicity, $\eta = U$.

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

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 Metropolis-Hastings algorithm was used to simulate the conditional posterior.

Bayesian inference for h (variable K)

If K changes, the dimension of U changes. With a prior for K, $\pi(K)$ and $\pi(U|\rho, K)$, the conditional posterior 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. Jacobian = 1.

- If K is too small, deeper orders are likely to be generated. e.g., $K = 1$ means the total order.
- If K is too large, shallow orders are likely to be generated. e.g., $K = \infty$ means no order.

K and ρ trade-off if K is large enough. (see page 6)

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 parameter
  K posterior mean is not necessarily the true dimension of PO. However, it will put ≈ 0 mass for K less than the true PO dimension.

Output summary

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

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

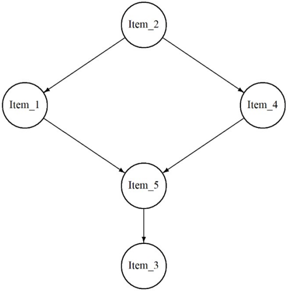
Posterior samples for ρ, p (and K) are summarized as usual.

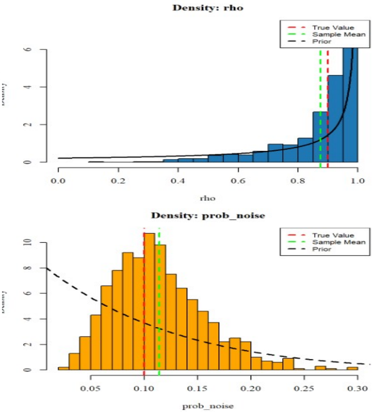
Diagnostics

Once the MC is simulated, the rate of acceptance, trace plot or some diagnostics.

R-code

Let's implement the synthetic data and simulate MCMC [1]





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

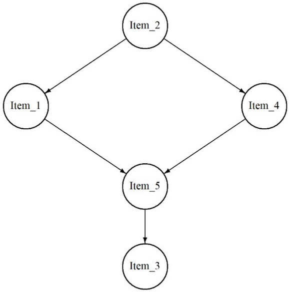
$$\rho \sim Beta(1,1/6)$$

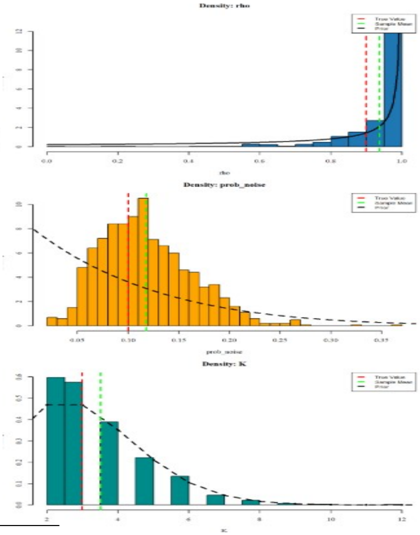
$$K = 3$$

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

Simulation study

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





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

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

$$K \sim \overline{Pois}(3)$$

[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).

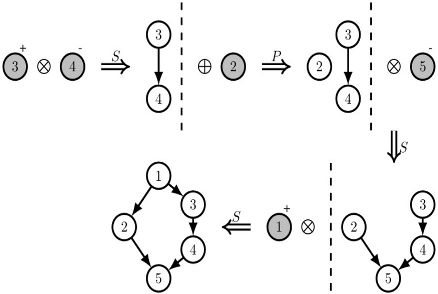
- Simpler model
- Faster LE counting algorithm
  - LEcount[3] with $O(t!M^{(t+3)})$ for a tree width t (Kangas et al (2020))
- Posterior approximation
  - *h*-matrix approximation??

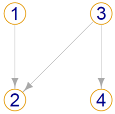
Restricted class of PO, "Vertex-Series-Parallel partial orders" (Mannila & Meek (2000))

[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)





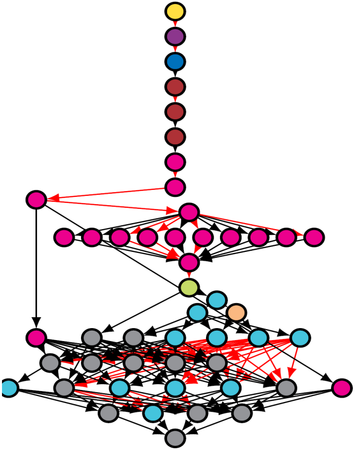
Two-dimensional POs ($K = 2$) except the forbidden graph order (right fig)

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.

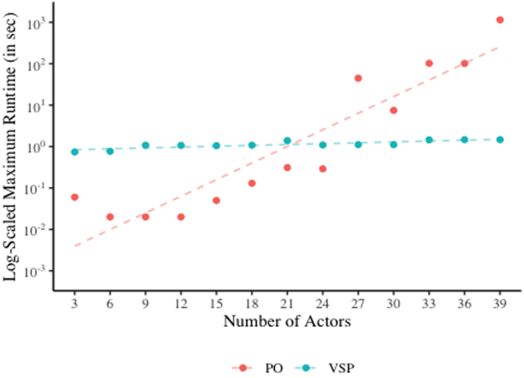
Vertex-Series-Parallel partial orders (VSPs)

Case study : Social hierarchy of $M = 49$ witnesses for 1134-1138 in England (Jiang et al (2023))



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

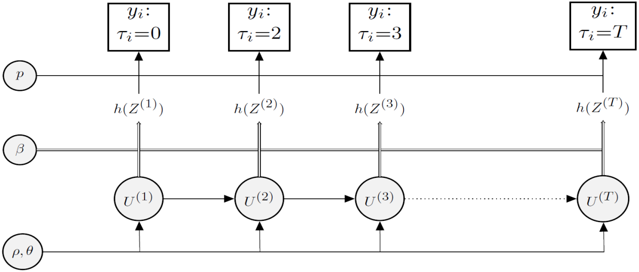
Linear extension counting time comparison (Runtime for 20 lists)



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

Other PO work

Time-varying PO (Nicholls et al (2025))
List data y_i come with time stamp $\tau_i \in \{1, \dots, T\}$. Extend to HMM using $U = (U^{(t)}\ t = 1, \dots, T)$ and $U \sim VAR_{\rho,\theta}(1)$.



PO with covariate effect (Nicholls et al (2025))
PO is $h(\eta)$ where the actor j specific covariant is X_j and latent variable is

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

Warning: A fixed covariate will yield an identifiable issue.

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

Reference

Jiang, C. and G. K. Nicholls (2024). Non-Parametric Bayesian Inference for Partial Orders with Ties from Rank Data observed with Mallows Noise. https://arxiv.org/abs/2408.14661.

Jiang, C. et al. (2023). Bayesian inference for vertex-series-parallel partial orders. Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence, pp. 995–1004. PMLR.

Nicholls, G. et al. (2025). Bayesian inference for partial orders from random linear extensions: power relations from 12th Century Royal Acta. Ann. Appl. Stat. 19(2): 1663-1690.

Kangas, K. et al. (2020). A faster tree-decomposition based algorithm for counting linear extensions. Algorithmica 82 2156–2173. MR4132887 https://doi.org/10.1007/s00453-019-00633-1

Mannila, H. and Meek, C. (2000) Global partial orders from sequential data. In Proceedings of the sixth ACMSIGKDDinternational conference on Knowledge discovery and data mining, pages 161-168.