Supplementary material for

On Bayesian inference for the Extended Plackett-Luce model

1 Simulating data from the Extended Plackett-Luce model

The method for generating data from the Extended Plackett-Luce model is similar to that for the standard (forward ranking) Plackett-Luce model however there is a subtle but yet important difference. As discussed in the paper, for the standard Plackett-Luce model, the entity that is chosen first is also that which is the most preferred. In contrast, for the Extended Plackett-Luce model the entity that is chosen first is instead considered to have rank σ_1. Recall that $\boldsymbol{x}^*$ denotes a permuted ranking defined by $x_j^* = x_{\sigma_j}$ for $j = 1, \ldots, K$, where $\boldsymbol{x}$ is the corresponding rank ordering and $\boldsymbol{\sigma}$ is the choice order parameter. Then it follows that, by construction, a permuted ranking $\boldsymbol{x}^*$ gives the ordering of the entities according to the *order in which they were chosen* (and not the preference of the entities). We can therefore simulate these (permuted) orderings from the standard Plackett-Luce model as they follow the forward ranking process by definition. Further given a permuted ranking $\boldsymbol{x}^*$ it is straightforward to recover the rank ordering $\boldsymbol{x}$ by noting that $x_{\sigma_j} = x_j^*$ for $j = 1, \ldots, K$. It is worth recalling that additional attention must be placed on the choice of the entity parameters when considering the Extended Plackett-Luce model; in particular the interpretation of the λ parameters depends explicitly on the choice of $\boldsymbol{\sigma}$. The strategy we suggest is to first specify the preference of the entities (via the λ values) under the standard Plackett-Luce interpretation and then, once a choice order value has been chosen (or simulated), use $\boldsymbol{\lambda}^{(\boldsymbol{\sigma})}$ (with jth element $\lambda_j^{(\boldsymbol{\sigma})} = \boldsymbol{\lambda}_{\eta_j^{(\boldsymbol{\sigma})}}$) to simulate the data using the mechanism below. This strategy will ensure that the modal rank ordering is preserved; this is discussed in detail within Section 2.2 of the paper.

Given a choice order parameter $\boldsymbol{\sigma}$ and an appropriate choice of entity parameters $\boldsymbol{\lambda}^{(\boldsymbol{\sigma})}$ a collection $X = \{\boldsymbol{x}_i\}_{i=1}^n$ of n *rank orderings* are generated from the Extended Plackett-Luce data generating mechanism as follows.

For $i = 1, \ldots, n$,

1. Sample $\nu_{ij} \overset{indep}{\sim} \text{Exp}(\lambda_j^{(\boldsymbol{\sigma})})$ for $j = 1, \ldots, K$.
2. Set $x_{ij}^* = \underset{q \in S_{ij}}{\text{argmin}}\ \nu_{iq}$ where $S_{ij} = \mathcal{K} \setminus \{x_{i1}^*, \ldots, x_{ij-1}^*\}$ for $j = 1, \ldots, K$.
3. Let $x_{i\sigma_j} = x_{ij}^*$ for $j = 1, \ldots, K$.

2 Simulation study

This section contains additional figures and tables corresponding to the simulation studies described in Section 4 of the paper. Recall that we consider $K \in \{5, 10, 15, 20\}$ entities and generate 500 orderings for each choice of K. Further we took several subsets of each of these datasets by taking the first $n \in \{20, 50, 200, 500\}$ orderings thus giving rise to 16 (nested) datasets. Table 1 gives the parameter values $(\boldsymbol{\lambda}', \boldsymbol{\sigma}')$ from which these data were generated for each choice of K. Note that these were drawn from the prior distribution outlined in Section 3.1 with $a_k = q_k = 1$ for $k = 1, \ldots, K$ and to ease comparison we have rescaled the entity parameter values so that $\lambda'_1 = 1$ in each case; the values given are also on the log scale. Table 2 shows posterior probability $\Pr(\boldsymbol{\sigma}'|\mathcal{D})$ along with the mean squared error between the (log) values $\boldsymbol{\lambda}'$ used to generate the data and the posterior expectation of the (log) entity parameters (conditional on the $\boldsymbol{\sigma}'$ used to generate the data) as in the paper (Table 1). However here the table also includes $\mathrm{E}(\log \boldsymbol{\lambda}|\mathcal{D}) \equiv \mathrm{E}_{\boldsymbol{\lambda},\boldsymbol{\sigma}=\boldsymbol{\sigma}'|\mathcal{D}}(\log \boldsymbol{\lambda})$; the posterior expectation of the (log) entity parameters conditional on $\boldsymbol{\sigma}'$ for those analyses where $\boldsymbol{\sigma}'$ was observed within the posterior samples. Again the values have been rescaled (offline) so that $\lambda'_1 = 1$ in each case.

Section 2.1 shows traceplots of the log (observed data) likelihood (top), marginal posterior distribution specified by the probabilities $\Pr(\boldsymbol{\sigma}|\mathcal{D})$ (middle) and the marginal posterior of the log entity parameters $\pi(\log \boldsymbol{\lambda}|\boldsymbol{\sigma} = \boldsymbol{\sigma}', \mathcal{D})$ (bottom) for each $K \in \{5, 10, 15, 20\}$ and $n \in \{20, 50, 200, 500\}$. Note that traceplots show the log likelihood given the posterior samples of $\boldsymbol{\sigma}, \boldsymbol{\lambda}$ obtained from the chain with $T = 1$ and not the log likelihood under the joint Markov chain. This plot is useful for diagnosing convergence/mixing issues and the plots shown suggest our chains have both converged and are mixing well in each case. Also when many unique choice order values are observed the figures showing the marginal posterior for $\boldsymbol{\sigma}$ only shows the 25 values with largest posterior support. In these figures (middle) the cross ($\times$) highlights the posterior probability attached to the choice order parameter used to simulate the respective dataset, that is, $\Pr(\boldsymbol{\sigma} = \boldsymbol{\sigma}'|\mathcal{D})$. The marginal posterior of the log entity parameters $\pi(\log \boldsymbol{\lambda}|\boldsymbol{\sigma} = \boldsymbol{\sigma}', \mathcal{D})$, conditional on the $\boldsymbol{\sigma}'$ used to generate the respective datasets, are depicted as boxplots (bottom). Again the entity parameter values have been rescaled (offline) so that $\lambda_1 = 1$ and so the boxplot for λ_1 is omitted. Here the cross ($\times$) on the left hand side of each boxplot shows the $\log \boldsymbol{\lambda}'$ values used to generate each dataset. For those analyses where the choice order $\boldsymbol{\sigma}'$ from which the data were generated is not observed in the posterior samples the boxplots of $\pi(\log \boldsymbol{\lambda}|\boldsymbol{\sigma} = \boldsymbol{\sigma}'|\mathcal{D})$ are omitted. For these analyses we provide tabulated values of the marginal distribution for each position in $\boldsymbol{\sigma}$ within Section 2.2, that is, the marginal posterior distributions given by $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$. These show that, even with limited information, we are able to infer the lower entries in $\boldsymbol{\sigma}$ fairly well and much of the uncertainty resides within the first few stages of the ranking process. Finally Section 2.3 shows the marginal distributions given by $\Pr(\sigma_j = k|\mathcal{D})$ as heat maps for each analysis; here the cross ($\times$) highlights $\Pr(\sigma_j = \sigma'_j|\mathcal{D})$.

K		Parameter values
5	$\boldsymbol{\sigma}' =$	$(5, 4, 2, 3, 1)$
	$\log \boldsymbol{\lambda}' =$	$(0, 0.49, -0.94, -2.43, -2.38)$
10	$\boldsymbol{\sigma}' =$	$(2, 10, 5, 4, 3, 1, 9, 8, 7, 6)$
	$\log \boldsymbol{\lambda}' =$	$(0, 3.78, 2.74, 1.89, 1.77, 4.72, 2.74, 1.26, 0.26, 2.92)$
15	$\boldsymbol{\sigma}' =$	$(1, 5, 14, 15, 13, 8, 9, 3, 11, 6, 10, 7, 12, 2, 4)$
	$\log \boldsymbol{\lambda}' =$	$(0, -0.94, -1.51, -2.20, -0.36, -1.12, -1.43, -0.46, 0.73, -1.70, -1.16, 0, -2.30, -0.48, -3.14)$
20	$\boldsymbol{\sigma}' =$	$(15, 2, 17, 12, 20, 6, 13, 19, 18, 7, 1, 16, 8, 5, 4, 9, 3, 11, 10, 14)$
	$\log \boldsymbol{\lambda}' =$	$(0, -2.41, -1.24, -0.72, -1.21, -0.12, -0.84, 0.47, 0.31, -2.36, 0.99, 1.11, -0.45, -2.25, 0.11, -1.28, -1.12, -0.36, -0.52, -0.57)$

Table 1: Simulated $(\boldsymbol{\lambda}', \boldsymbol{\sigma}')$ pairs for each $K \in \{5, 10, 15, 20\}$

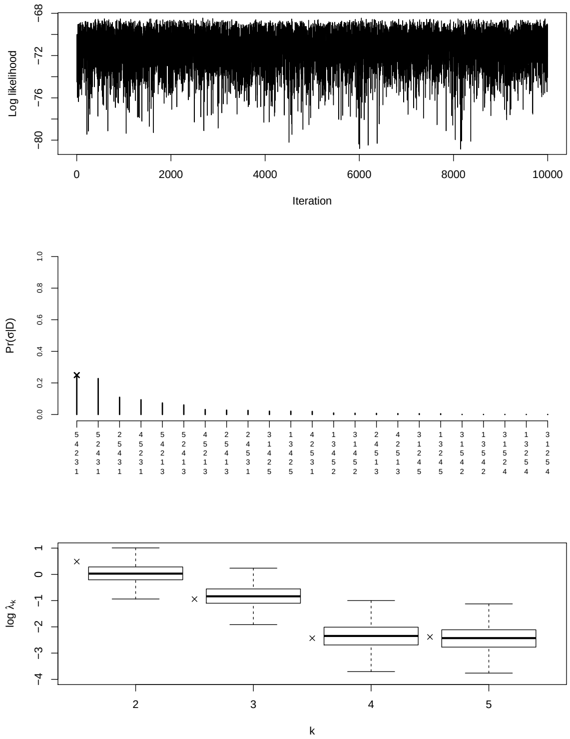
		n			
K		20	50	200	500
5	$\Pr(\boldsymbol{\sigma}'\|\mathcal{D})$	0.294*	0.716*	1.000*	1.000*
	$\mathrm{E}(\log \boldsymbol{\lambda}\|\mathcal{D})$	$(0, 0.03, -0.84, -2.36, -2.44)$	$(0, 0.14, -0.84, -2.51, -2.32)$	$(0, 0.50, -0.77, -2.26, -2.08)$	$(0, 0.43, -0.89, -2.35, -2.20)$
	$\frac{1}{K}\sum_k \left[\mathrm{E}(\log \lambda_k\|\mathcal{D}) - \log \lambda'_k\right]^2$	0.045	0.028	0.030	0.009
10	$\Pr(\boldsymbol{\sigma}'\|\mathcal{D})$	0.156*	0.604*	1.000*	1.000*
	$\mathrm{E}(\log \boldsymbol{\lambda}\|\mathcal{D})$	$(0, 4.31, 3.20, 2.55, 2.63,$ $5.02, 3.26, 2.07, 0.49, 3.03)$	$(0, 3.63, 2.68, 2.22, 1.98,$ $4.49, 2.58, 1.66, 0.67, 2.88)$	$(0, 3.47, 2.54, 1.74, 1.71,$ $4.47, 2.52, 1.15, 0.27, 2.80)$	$(0, 3.67, 2.63, 1.85, 1.72,$ $4.64, 2.66, 1.25, 0.39, 2.92)$
	$\frac{1}{K}\sum_k \left[\mathrm{E}(\log \lambda_k\|\mathcal{D}) - \log \lambda'_k\right]^2$	0.276	0.059	0.030	0.006
15	$\Pr(\boldsymbol{\sigma}'\|\mathcal{D})$	0.000	0.006	0.072	0.548*
	$\mathrm{E}(\log \boldsymbol{\lambda}\|\mathcal{D})$	—	$(0, -0.86, -1.29, -1.98, -0.36,$ $-1.29, -1.46, -0.34, 0.85, -1.64,$ $-1.00, 0.19, -2.14, -0.27, -3.11)$	$(0, -0.97, -1.37, -2.02, -0.31,$ $-1.17, -1.39, -0.28, 0.77, -1.65,$ $-1.17, 0.09, -2.17, -0.31, -3.12)$	$(0, -1.04, -1.51, -2.16, -0.34,$ $-1.2, -1.46, -0.43, 0.68, -1.79,$ $-1.23, 0.03, -2.29, -0.45, -3.13)$
	$\frac{1}{K}\sum_k \left[\mathrm{E}(\log \lambda_k\|\mathcal{D}) - \log \lambda'_k\right]^2$	—	0.020	0.010	0.002
20	$\Pr(\boldsymbol{\sigma}'\|\mathcal{D})$	0.000	0.000	0.035	0.313*
	$\mathrm{E}(\log \boldsymbol{\lambda}\|\mathcal{D})$	—	—	$(0, -2.52, -1.30, -0.92, -1.33,$ $-0.07, -0.90, 0.50, 0.27, -2.39,$ $0.89, 0.88, -0.51, -2.35, -0.08,$ $-1.32, -1.18, -0.43, -0.55, -0.64)$	$(0, -2.68, -1.31, -0.89, -1.28,$ $-0.12, -0.89, 0.46, 0.27, -2.36,$ $0.92, 0.98, -0.51, -2.31, -0.03,$ $-1.28, -1.11, -0.39, -0.48, -0.64)$
	$\frac{1}{K}\sum_k \left[\mathrm{E}(\log \lambda_k\|\mathcal{D}) - \log \lambda'_k\right]^2$	—	—	0.010	0.009

Table 2: Posterior probability $\Pr(\boldsymbol{\sigma}'|\mathcal{D})$ of the choice order used to generate each dataset along with $\mathrm{E}(\log \boldsymbol{\lambda}|\mathcal{D}) \equiv \mathrm{E}_{\boldsymbol{\lambda},\boldsymbol{\sigma}=\boldsymbol{\sigma}'|\mathcal{D}}(\log \boldsymbol{\lambda})$; the posterior expectation of the (log) entity parameters conditional on $\boldsymbol{\sigma}'$. *indicates that $\boldsymbol{\sigma}'$ is also the (posterior) modal observed choice order.

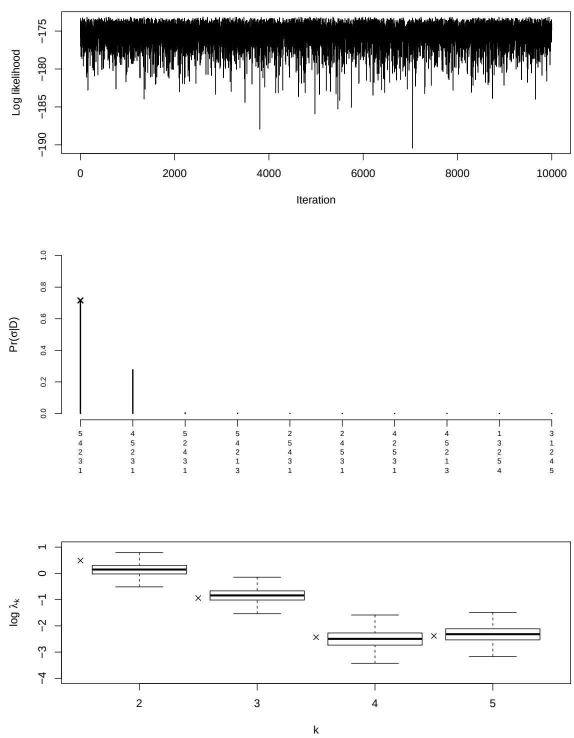
2.1 Marginal posterior summaries

$\boldsymbol{K = 5}$

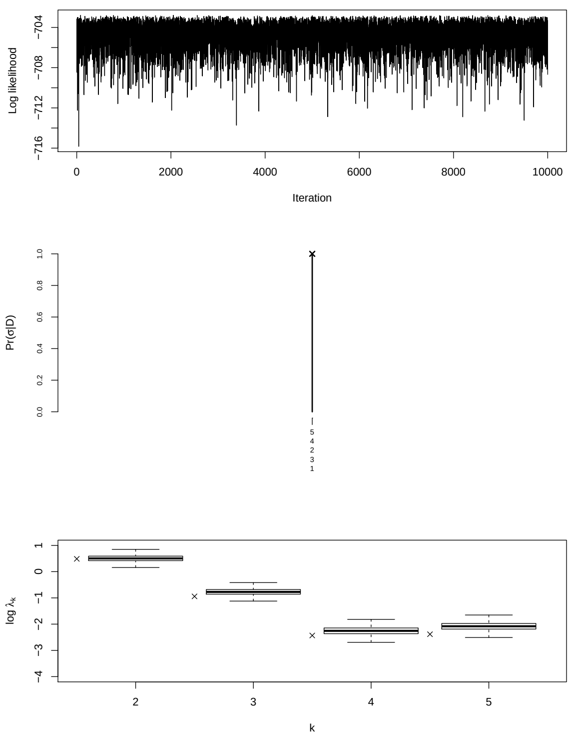
$n = 20$



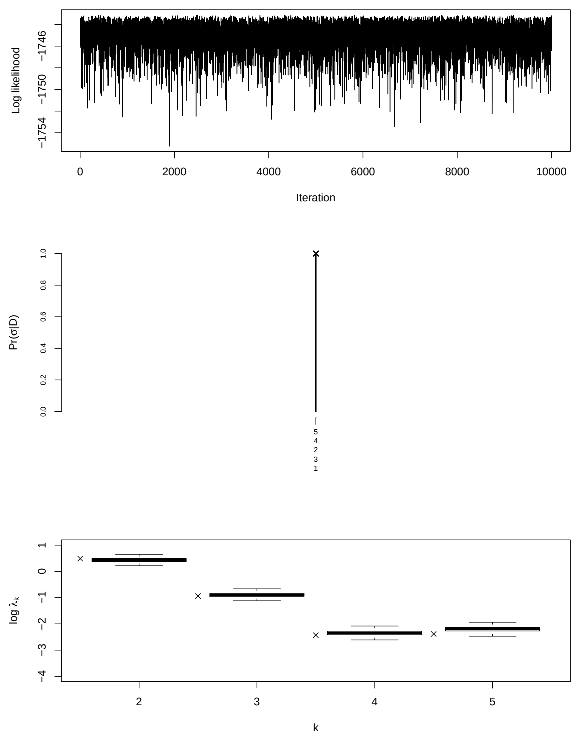
$n = 50$



$n = 200$

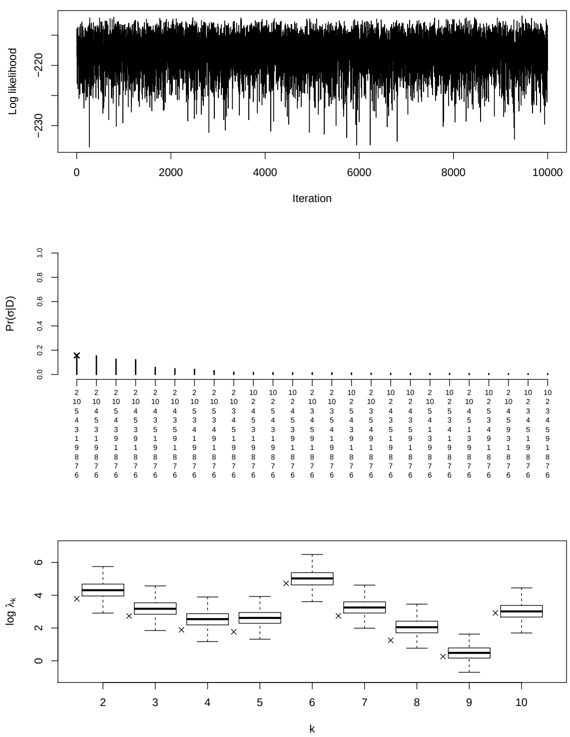


$n = 500$

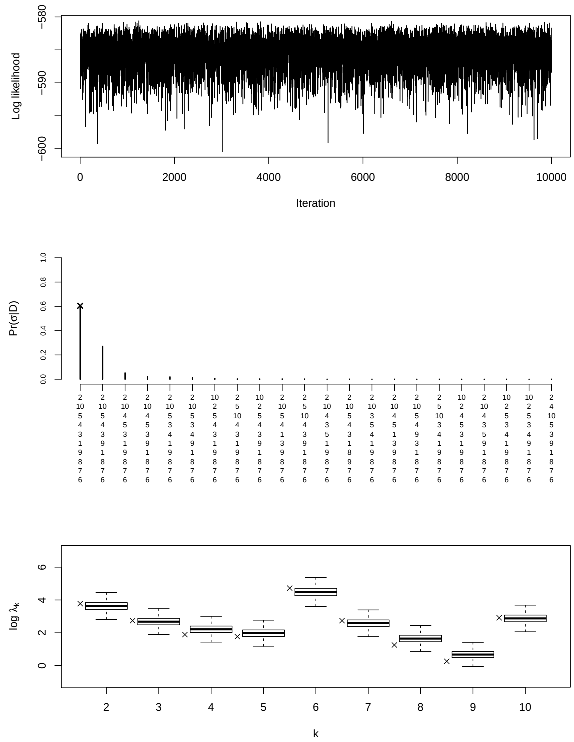


$\boldsymbol{K} = \mathbf{10}$

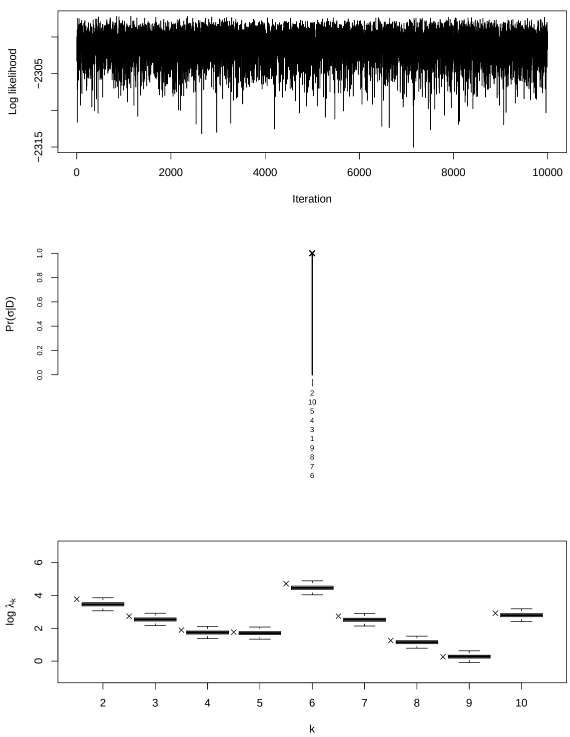
$n = 20$



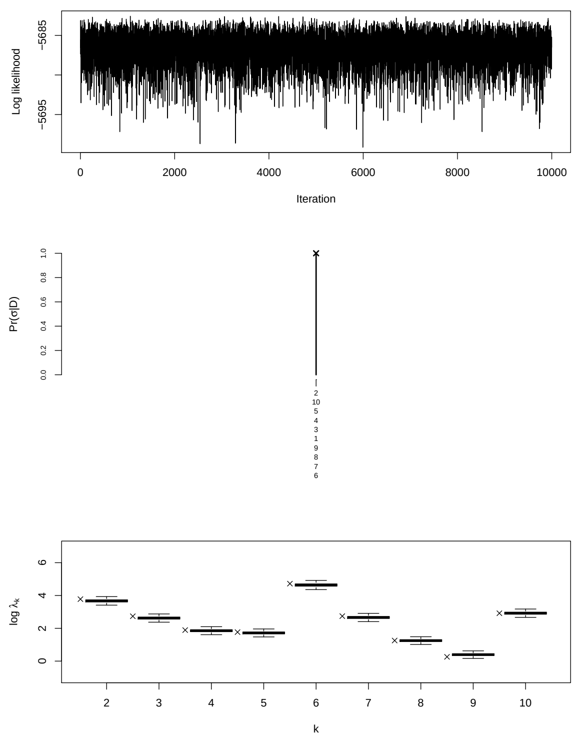
$n = 50$



$n = 200$

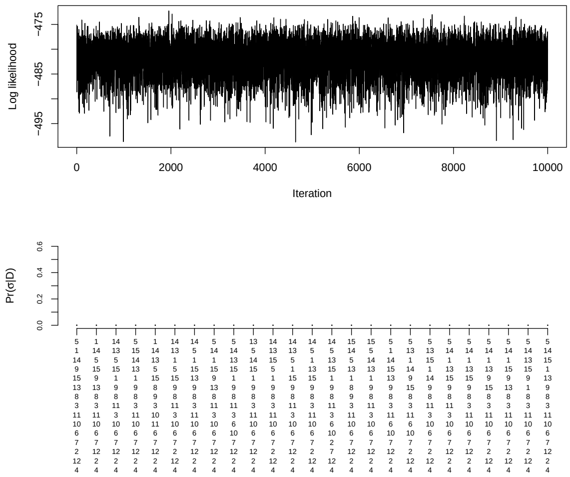


$n = 500$

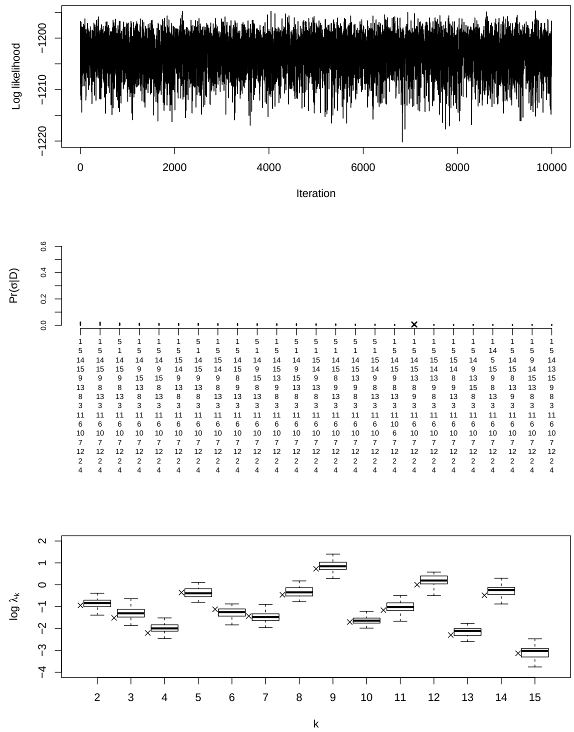


$K = 15$

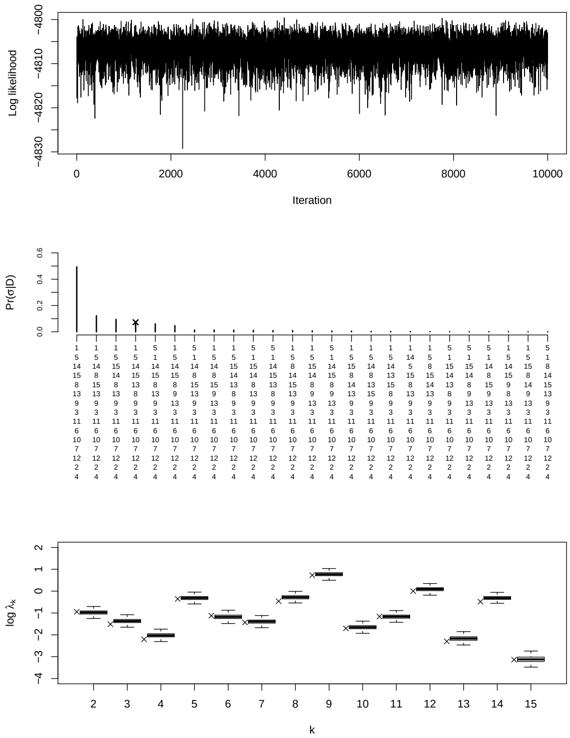
$n = 20$



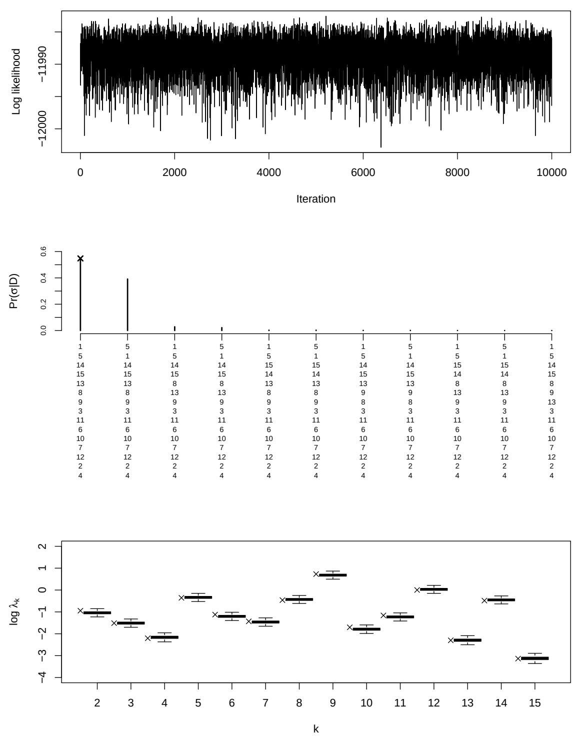
$n = 50$



$n = 200$

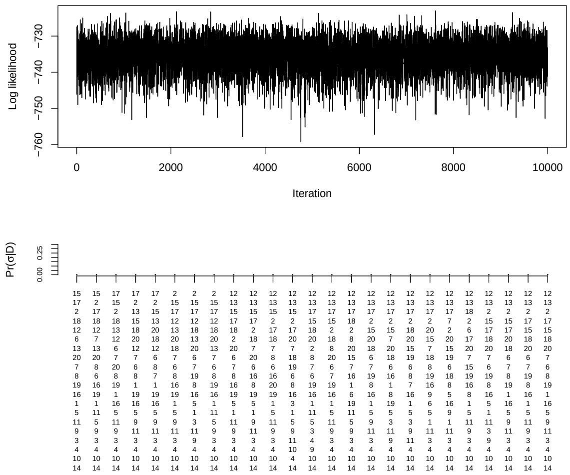


$n = 500$

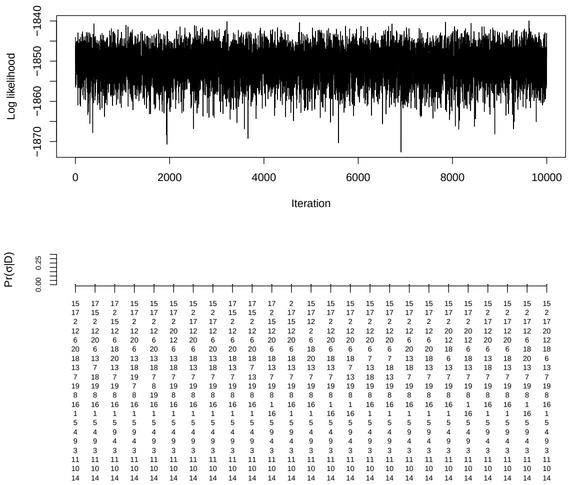


$\boldsymbol{K = 20}$

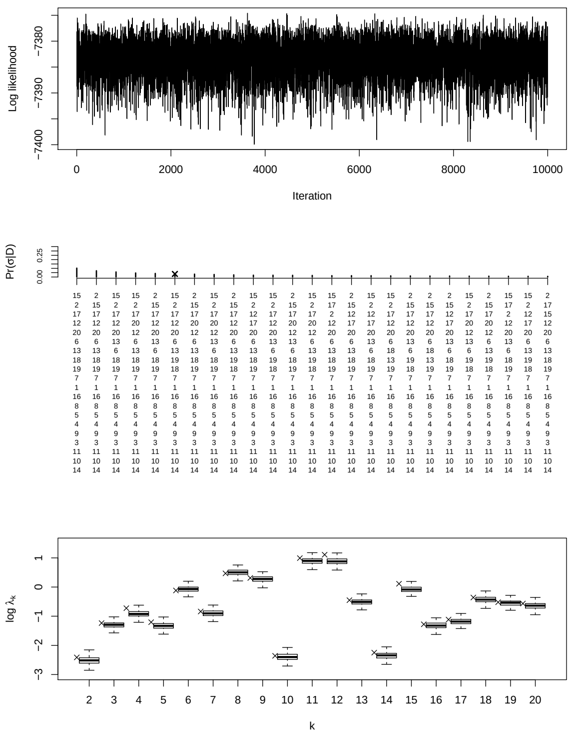
$n = 20$



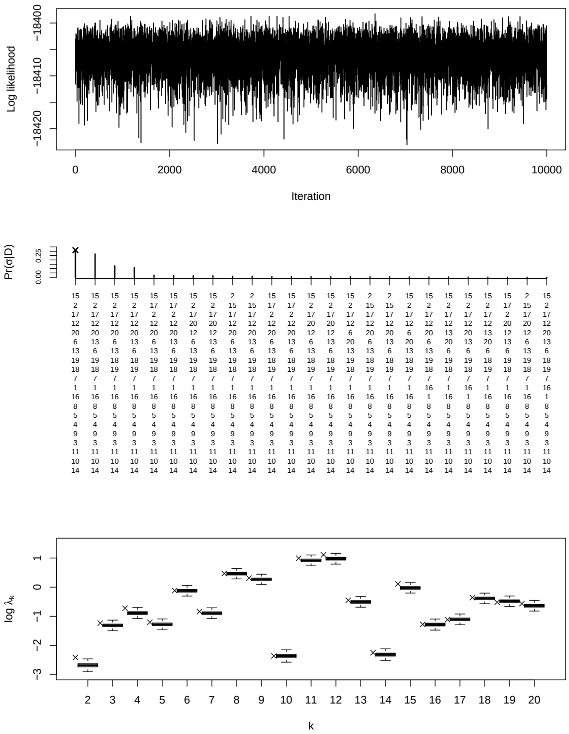
$n = 50$



$n = 200$



$n = 500$



2.2 Marginal posterior distribution for each stage in the ranking process

$K = 15,\ n = 20$

								k							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\Pr(\sigma_1 = k\|\mathcal{D})$	**0.11**	0	0	0	0.33	0	0	0	0.06	0	0	0	0.12	0.26	0.12
$\Pr(\sigma_2 = k\|\mathcal{D})$	0.14	0	0	0	**0.25**	0	0	0.01	0.08	0	0	0	0.14	0.24	0.14
$\Pr(\sigma_3 = k\|\mathcal{D})$	0.17	0	0	0	0.18	0	0	0.02	0.11	0	0	0	0.15	**0.20**	0.17
$\Pr(\sigma_4 = k\|\mathcal{D})$	0.19	0	0	0	0.12	0	0	0.03	0.14	0	0	0	0.17	0.15	**0.20**
$\Pr(\sigma_5 = k\|\mathcal{D})$	0.19	0	0.01	0	0.08	0	0	0.07	0.19	0	0	0	**0.18**	0.09	0.19
$\Pr(\sigma_6 = k\|\mathcal{D})$	0.14	0	0.04	0	0.03	0	0	**0.18**	0.26	0	0.01	0	0.16	0.04	0.14
$\Pr(\sigma_7 = k\|\mathcal{D})$	0.05	0	0.15	0	0	0	0	0.44	**0.13**	0.01	0.09	0	0.07	0.01	0.05
$\Pr(\sigma_8 = k\|\mathcal{D})$	0.01	0	**0.40**	0	0	0.01	0	0.19	0.03	0.04	0.30	0	0.01	0	0.01
$\Pr(\sigma_9 = k\|\mathcal{D})$	0	0	0.31	0	0	0.04	0	0.06	0	0.15	**0.44**	0	0	0	0
$\Pr(\sigma_{10} = k\|\mathcal{D})$	0	0	0.07	0	0	**0.23**	0.02	0.01	0	0.53	0.14	0	0	0	0
$\Pr(\sigma_{11} = k\|\mathcal{D})$	0	0.02	0.01	0	0	0.58	0.1	0	0	**0.25**	0.02	0.02	0	0	0
$\Pr(\sigma_{12} = k\|\mathcal{D})$	0	0.17	0	0	0	0.13	**0.54**	0	0	0.03	0	0.13	0	0	0
$\Pr(\sigma_{13} = k\|\mathcal{D})$	0	0.32	0	0	0	0.01	0.28	0	0	0	0	**0.39**	0	0	0
$\Pr(\sigma_{14} = k\|\mathcal{D})$	0	**0.48**	0	0	0	0	0.06	0	0	0	0	0.46	0	0	0
$\Pr(\sigma_{15} = k\|\mathcal{D})$	0	0	0	**1.00**	0	0	0	0	0	0	0	0	0	0	0

Synthetic dataset: $K = 15$, $n = 20$. Marginal posterior distribution for each stage in the ranking process, that is, $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$. Bold probabilities are $\Pr(\sigma_j = \sigma'_j = k|\mathcal{D})$, that is, the posterior probabilities of each entry in the choice order parameter used to generate the data $\boldsymbol{\sigma}'$.

$\boldsymbol{K = 20,\ n = 20}$

	k																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\Pr(\sigma_1 = k)$	0	0.29	0	0	0	0	0	0	0	0	0	0.05	0.03	0	**0.17**	0	0.40	0.04	0	0.01
$\Pr(\sigma_2 = k)$	0	**0.28**	0	0	0	0	0.01	0	0	0	0	0.08	0.06	0	0.21	0	0.28	0.07	0	0.02
$\Pr(\sigma_3 = k)$	0	0.21	0	0	0	0.01	0.01	0	0	0	0	0.13	0.10	0	0.22	0	**0.17**	0.11	0	0.04
$\Pr(\sigma_4 = k)$	0	0.12	0	0	0	0.02	0.03	0	0	0	0	**0.17**	0.14	0	0.18	0	0.09	0.17	0	0.07
$\Pr(\sigma_5 = k)$	0	0.06	0	0	0	0.04	0.05	0.01	0	0	0	0.18	0.19	0	0.12	0	0.04	0.19	0	**0.12**
$\Pr(\sigma_6 = k)$	0	0.02	0	0	0	**0.08**	0.10	0.02	0	0	0	0.16	0.20	0	0.06	0	0.01	0.17	0	0.18
$\Pr(\sigma_7 = k)$	0	0.01	0	0	0	0.14	0.15	0.04	0	0	0	0.12	**0.16**	0	0.03	0	0	0.13	0.01	0.20
$\Pr(\sigma_8 = k)$	0.01	0	0	0	0	0.22	0.22	0.08	0	0	0	0.08	0.09	0	0.01	0.01	0	0.08	**0.03**	0.18
$\Pr(\sigma_9 = k)$	0.02	0	0	0	0	0.27	0.21	0.18	0	0	0	0.03	0.03	0	0	0.03	0	**0.04**	0.08	0.11
$\Pr(\sigma_{10} = k)$	0.07	0	0	0	0.01	0.16	**0.13**	0.29	0	0	0	0.01	0.01	0	0	0.08	0	0.01	0.19	0.04
$\Pr(\sigma_{11} = k)$	**0.14**	0	0	0	0.02	0.06	0.06	0.22	0	0	0	0	0	0	0	0.19	0	0	0.30	0.01
$\Pr(\sigma_{12} = k)$	0.25	0	0	0	0.06	0.01	0.02	0.11	0.01	0	0	0	0	0	0	**0.30**	0	0	0.24	0
$\Pr(\sigma_{13} = k)$	0.32	0	0.01	0	0.17	0	0	**0.04**	0.02	0	0.03	0	0	0	0	0.28	0	0	0.12	0
$\Pr(\sigma_{14} = k)$	0.15	0	0.04	0	**0.45**	0	0	0.01	0.12	0	0.12	0	0	0	0	0.09	0	0	0.02	0
$\Pr(\sigma_{15} = k)$	0.03	0	0.12	**0**	0.22	0	0	0	0.31	0	0.30	0	0	0	0	0.02	0	0	0	0
$\Pr(\sigma_{16} = k)$	0.01	0	0.22	0.01	0.07	0	0	0	**0.36**	0	0.34	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{17} = k)$	0	0	**0.51**	0.10	0.01	0	0	0	0.17	0	0.20	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{18} = k)$	0	0	0.10	0.86	0	0	0	0	0.01	0.03	**0.01**	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{19} = k)$	0	0	0	0.03	0	0	0	0	0	**0.96**	0	0	0	0.01	0	0	0	0	0	0
$\Pr(\sigma_{20} = k)$	0	0	0	0	0	0	0	0	0	0.01	0	0	0	**0.99**	0	0	0	0	0	0

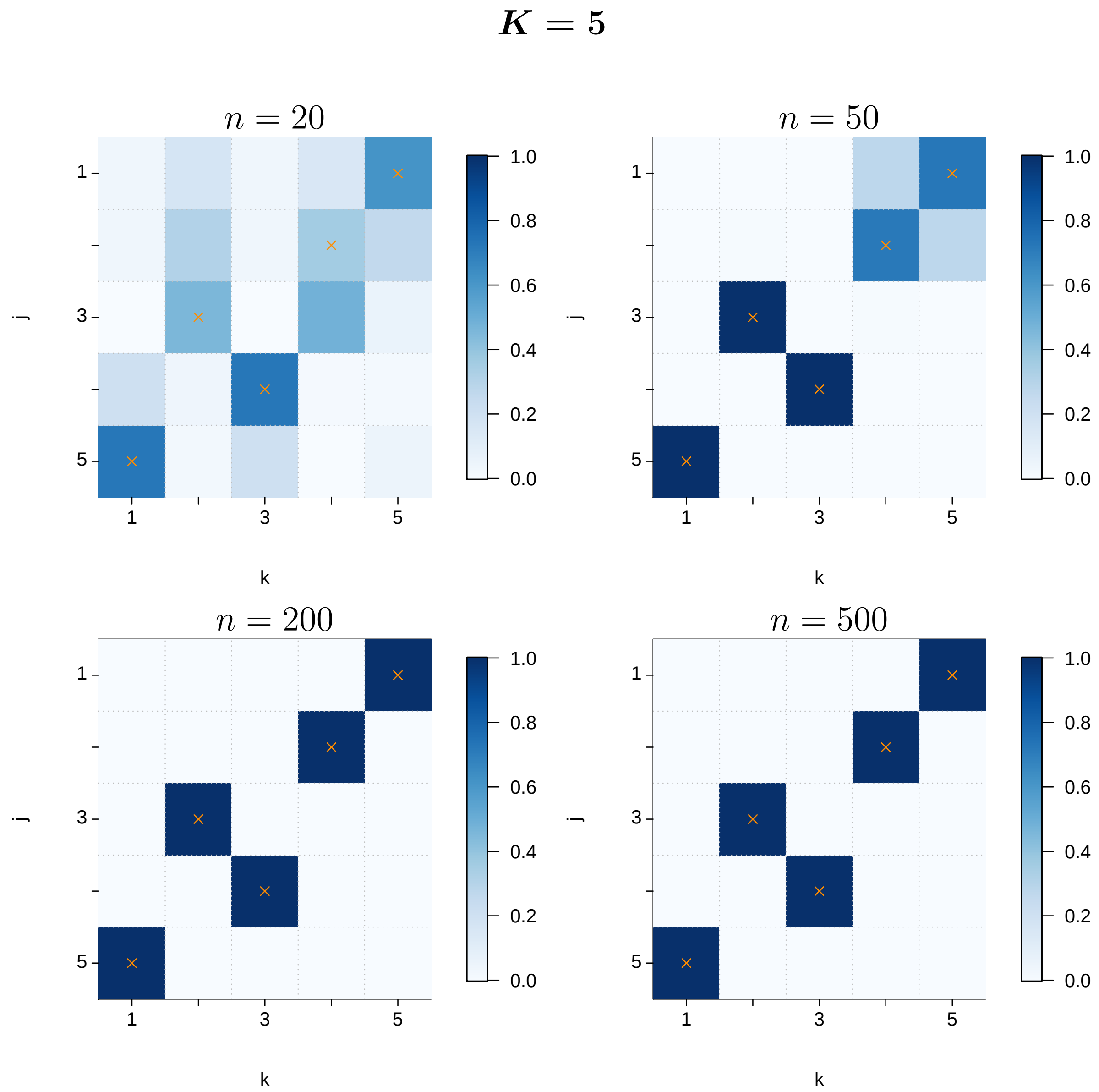
Synthetic dataset: $K = 20$, $n = 20$. Marginal posterior distribution for each stage in the ranking process, that is, $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$. Bold probabilities are $\Pr(\sigma_j = \sigma'_j = k|\mathcal{D})$, that is, the posterior probabilities of each entry in the choice order parameter used to generate the data $\boldsymbol{\sigma}'$.

$\boldsymbol{K = 20,\ n = 50}$

	k																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\Pr(\sigma_1 = k)$	0	0.23	0	0	0	0	0	0	0	0	0	0.04	0	0	**0.34**	0	0.38	0	0	0
$\Pr(\sigma_2 = k)$	0	**0.27**	0	0	0	0.01	0	0	0	0	0	0.07	0	0	0.31	0	0.32	0	0	0.02
$\Pr(\sigma_3 = k)$	0	0.30	0	0	0	0.03	0	0	0	0	0	0.17	0	0	0.23	0	**0.21**	0.01	0	0.06
$\Pr(\sigma_4 = k)$	0	0.15	0	0	0	0.11	0	0	0	0	0	**0.35**	0.01	0	0.10	0	0.07	0.03	0	0.17
$\Pr(\sigma_5 = k)$	0	0.04	0	0	0	0.23	0.01	0	0	0	0	0.24	0.06	0	0.02	0	0.02	0.09	0	**0.29**
$\Pr(\sigma_6 = k)$	0	0.01	0	0	0	**0.28**	0.04	0	0	0	0	0.10	0.14	0	0	0	0	0.19	0	0.24
$\Pr(\sigma_7 = k)$	0	0	0	0	0	0.20	0.10	0	0	0	0	0.03	**0.26**	0	0	0	0	0.25	0.01	0.15
$\Pr(\sigma_8 = k)$	0	0	0	0	0	0.10	0.22	0	0	0	0	0.01	0.31	0	0	0	0	0.25	**0.05**	0.06
$\Pr(\sigma_9 = k)$	0	0	0	0	0	0.03	0.41	0.03	0	0	0	0	0.17	0	0	0	0	**0.15**	0.19	0.02
$\Pr(\sigma_{10} = k)$	0.01	0	0	0	0	0	**0.18**	0.19	0	0	0	0	0.04	0	0	0.01	0	0.03	0.54	0
$\Pr(\sigma_{11} = k)$	**0.06**	0	0	0	0	0	0.03	0.59	0	0	0	0	0	0	0	0.12	0	0	0.19	0
$\Pr(\sigma_{12} = k)$	0.32	0	0	0	0.01	0	0	0.15	0	0	0	0	0	0	0	**0.50**	0	0	0.01	0
$\Pr(\sigma_{13} = k)$	0.56	0	0	0	0.06	0	0	**0.03**	0	0	0	0	0	0	0	0.34	0	0	0	0
$\Pr(\sigma_{14} = k)$	0.05	0	0	0.03	**0.86**	0	0	0	0.03	0	0	0	0	0	0	0.02	0	0	0	0
$\Pr(\sigma_{15} = k)$	0	0	0.01	**0.46**	0.07	0	0	0	0.47	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{16} = k)$	0	0	0.06	0.46	0	0	0	0	**0.46**	0	0.01	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{17} = k)$	0	0	**0.67**	0.05	0	0	0	0	0.04	0	0.25	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{18} = k)$	0	0	0.26	0	0	0	0	0	0	0	**0.74**	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{19} = k)$	0	0	0	0	0	0	0	0	0	**1.00**	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{20} = k)$	0	0	0	0	0	0	0	0	0	0	0	0	0	**1.00**	0	0	0	0	0	0

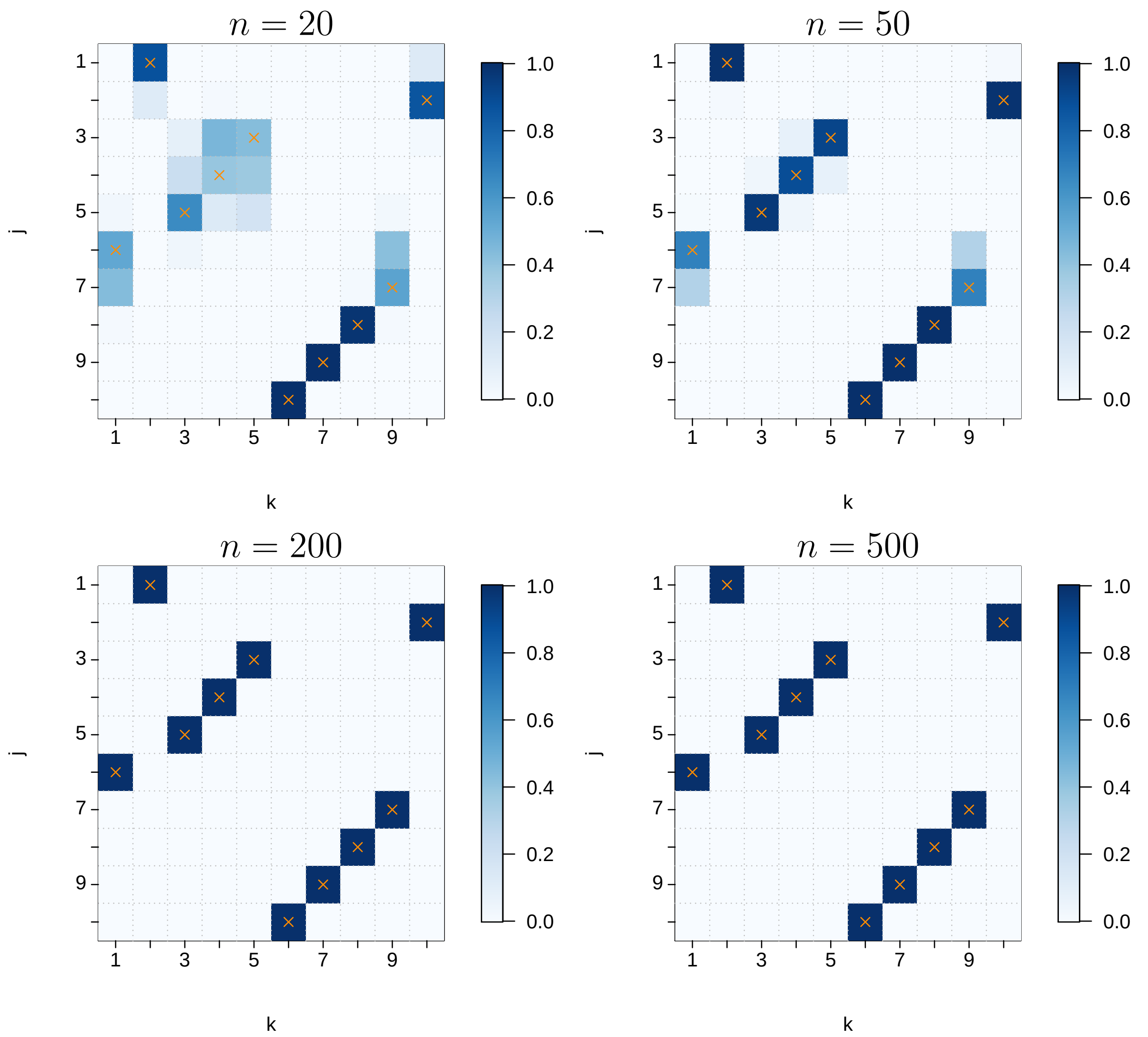
Synthetic dataset: $K = 20$, $n = 50$. Marginal posterior distribution for each stage in the ranking process, that is, $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$. Bold probabilities are $\Pr(\sigma_j = \sigma'_j = k|\mathcal{D})$, that is, the posterior probabilities of each entry in the choice order parameter used to generate the data $\boldsymbol{\sigma}'$.

2.3 Image plots



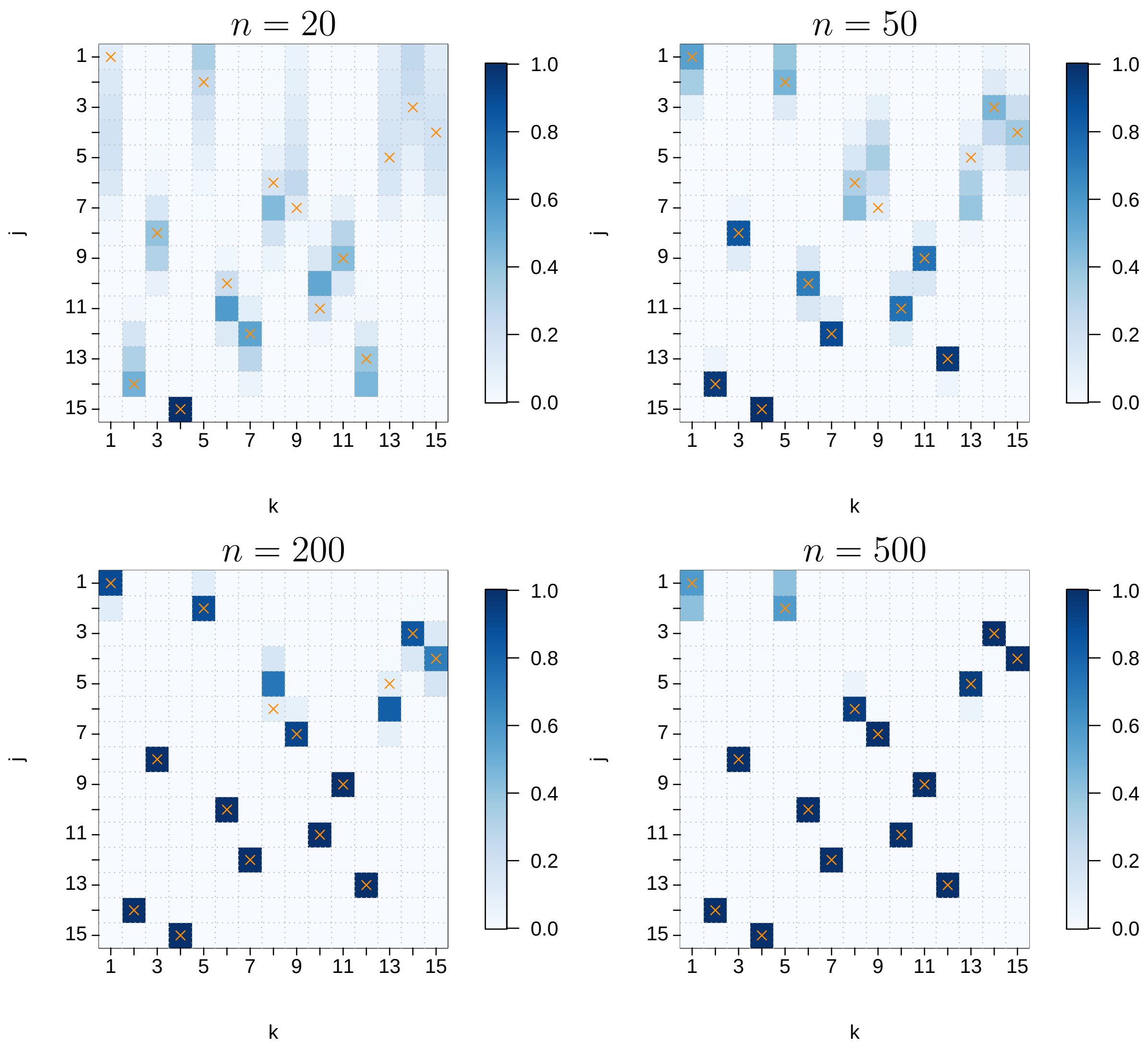
Synthetic data: heat maps of $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$; the crosses highlight $\Pr(\sigma_j = \sigma'_j|\mathcal{D})$ in each case.

$\boldsymbol{K = 10}$



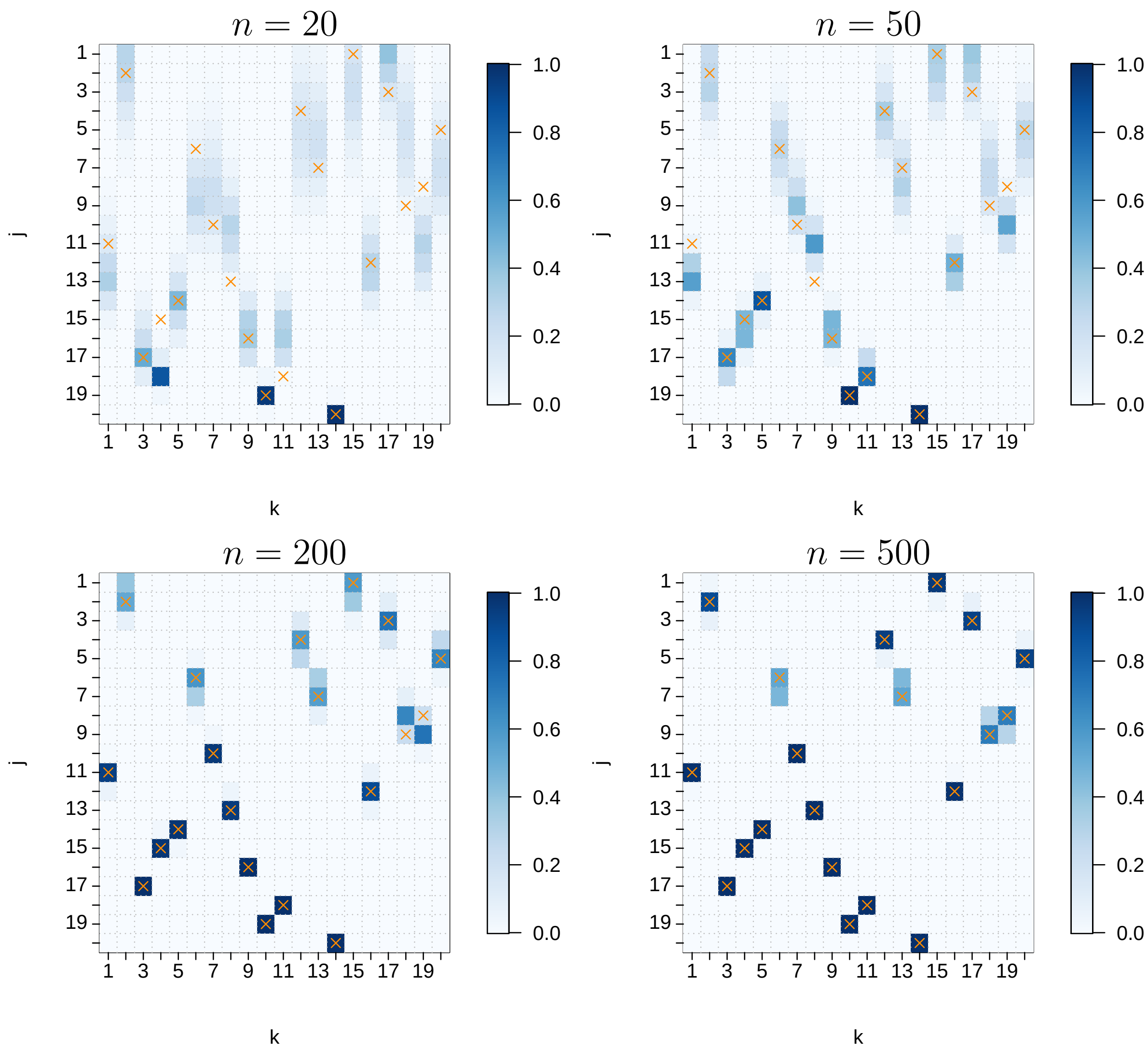
Synthetic data: heat maps of $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \dots, K\}$; the crosses highlight $\Pr(\sigma_j = \sigma'_j|\mathcal{D})$ in each case.

$K = 15$



Synthetic data: heat maps of $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$; the crosses highlight $\Pr(\sigma_j = \sigma'_j|\mathcal{D})$ in each case.

$\boldsymbol{K = 20}$



Synthetic data: heat maps of $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$; the crosses highlight $\Pr(\sigma_j = \sigma'_j|\mathcal{D})$ in each case.

3 Likelihood information about the choice order

To further motivate that the choice order is not only identifiable but also that the likelihood function can also be quite informative (about $\boldsymbol{\sigma}$) we examine the EPL likelihood function evaluated at the maximum a posteriori (MAP) estimate of the skill parameters as a function of the choice order $\boldsymbol{\sigma}$. Conditional on a particular choice order $\boldsymbol{\sigma}$, the (marginal) posterior $\pi(\boldsymbol{\lambda}|\mathcal{D},\boldsymbol{\sigma}) \propto \pi(\mathcal{D}|\boldsymbol{\lambda},\boldsymbol{\sigma})\pi(\boldsymbol{\lambda}|\boldsymbol{\sigma})$ is maximised at the MAP estimate $\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma}_j)$ and this estimate will depend on the choice order. Caron and Doucet (2012) proposed an adaptation of the Minorisation/Maximisation (MM) algorithm of Hunter (2004) that allows the MAP estimate of the (skill) parameters under the standard Plackett-Luce model to be obtained. Recalling that the Extended Plackett-Luce probability is equivalent to the standard Plackett-Luce probability of the permuted rankings $\boldsymbol{x}_i^*$, where $x_{ik}^* = x_{i\sigma_k}$ for $k = 1, \ldots, K$. It follows that, for any $\boldsymbol{\sigma} \in \mathcal{S}_k$, we can let $\mathcal{D} = \{\boldsymbol{x}_i^*\}_{i=1}^n$ be the collection of permuted rankings and use the (adapted) MM algorithm to obtain $\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma})$ so that the marginal posterior under the standard Plackett-Luce model is maximised. Full details of the MM algorithm are provided in Section 3.1.

We generate $n = 5000$ observations (rank orderings) of $K = 5$ entities from the Extended Plackett-Luce model as described in Section 1. These rankings were simulated using the choice order parameter $\boldsymbol{\sigma} = (5, 3, 2, 1, 4)$ and entity parameters $\boldsymbol{\lambda}^{(\boldsymbol{\sigma})} = (5, 4, 3, 2, 1)$. The MAP estimates $\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma}_j)$ are obtained for each of the possible choice orders $j = 1, \ldots, 5!$ using (independent) MM algorithms for each collection of permuted data $\mathcal{D}_j = \{\boldsymbol{x}_i^*\}_{i=1}^n$. We choose $a_k^{(\boldsymbol{\sigma})} = b_k^{(\boldsymbol{\sigma})} = 1$ and so $\lambda_k \overset{indep}{\sim} \text{Ga}(1, 1)$ for $k = 1, \ldots, K$ and all $\boldsymbol{\sigma} \in \mathcal{S}_K$, that is, we provide no information about the preference of the entities *a priori*. Further, given we are enumerating over each of the $K!$ possible choice orders there is no need to specify a prior distribution for $\boldsymbol{\sigma}$ as this parameter is treated as a fixed constant within each analysis. Each MM algorithm is initalised with $\lambda_k = \lambda = 1/K$ and the algorithm proceeds until the L_2-norm of the change in the value of the parameter vector is less than 10^{-9}; this typically takes fewer than 50 iterations. Note that the identifiability issue is resolved by constraining the skill parameters $\boldsymbol{\lambda}$ to the $(K-1)$-dimensional simplex.

Figure 1 shows $\log \pi(\mathcal{D}|\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma}_j), \boldsymbol{\sigma}_j)$, that is, the log Extended Plackett-Luce likelihood evaluated at the MAP estimates of the skill parameters $\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma}_j)$ for each choice order $\boldsymbol{\sigma}_j$. Clearly the log-likelihood is not constant and indeed the assumed choice order can have a large affect on the overall log-likelihood. Table 3 shows (a subset of) the choice order parameter values (permutations) listed according to the value of the log-likelihood under

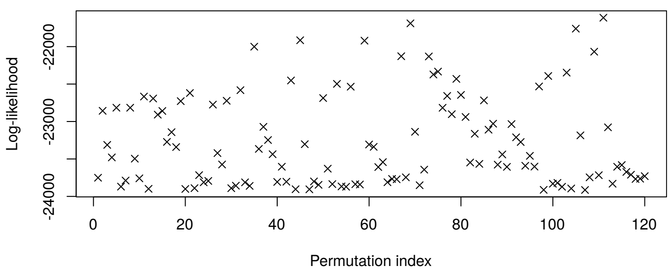


Figure 1: $\log \pi(\mathcal{D}|\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma}_j), \boldsymbol{\sigma}_j)$: log-likelihood evaluated at the MAP estimates of the skill parameters $\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma}_j)$ given each choice order $\boldsymbol{\sigma}_j$ for $j = 1, \ldots, K!$.

Rank	$\log\pi(\mathcal{D}\|\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma}_j),\boldsymbol{\sigma}_j)$	$\boldsymbol{\sigma}_j$
1	-21613.95	(5,3,2,1,4)
2	-21689.21	(3,5,2,1,4)
3	-21760.05	(5,2,3,1,4)
4	-21914.95	(2,5,3,1,4)
5	-21920.27	(3,2,5,1,4)
$\vdots$	$\vdots$	$\vdots$
9	-22131.35	(4,1,2,3,5)
$\vdots$	$\vdots$	$\vdots$
84	-23750.71	(1,2,3,4,5)
$\vdots$	$\vdots$	$\vdots$
116	-23900.67	(1,5,2,4,3)
117	-23904.19	(2,5,1,4,3)
118	-23905.65	(2,5,4,1,3)
119	-23912.46	(5,1,2,4,3)
120	-23914.28	(5,2,4,1,3)

Table 3: A subset of the choice order parameter values (permutations) listed according to the value of the log-likelihood under the corresponding MAP estimates of the skill parameters ($\log\pi(\mathcal{D}|\boldsymbol{\lambda}_{\text{MAP}}(\boldsymbol{\sigma}_j),\boldsymbol{\sigma}_j)$).

the corresponding MAP estimates of the skill parameters. Interestingly each of the top 5 choice orders have $\sigma_4 = 1$ and $\sigma_5 = 4$ and so we have clearly identified that the most preferred entity is chosen 4th and the fourth preferred entity is chosen last. Also, perhaps surprisingly, the reverse choice order ($\boldsymbol{\sigma}_j = (4,1,2,3,5)$) to the one these data were simulated from results in the ninth largest log-likelihood. It follows that, for the Extended Plackett-Luce model, local modes may be separated by large distances within permutation space and this will need to be addressed when constructing a posterior sampling scheme within the Bayesian framework.

3.1 MM Algorithm

Given a particular choice order parameter $\boldsymbol{\sigma}$ the collection of entity parameters $\boldsymbol{\lambda}_{\text{MAP}}$ that maximise the marginal posterior distribution $\pi(\boldsymbol{\lambda}|\mathcal{D},\boldsymbol{\sigma})$ under the Extended Plackett-Luce likelihood can be obtained by instead maximising the standard (forward ranking) Plackett-Luce model given a collection of permuted rankings $X^* = \{\boldsymbol{x}_i^*\}_{i=1}^n$ (where $x_{ij}^* = x_{i\sigma_j}$ for $i = 1,\dots,n$, $j = 1,\dots,K$) as follows.

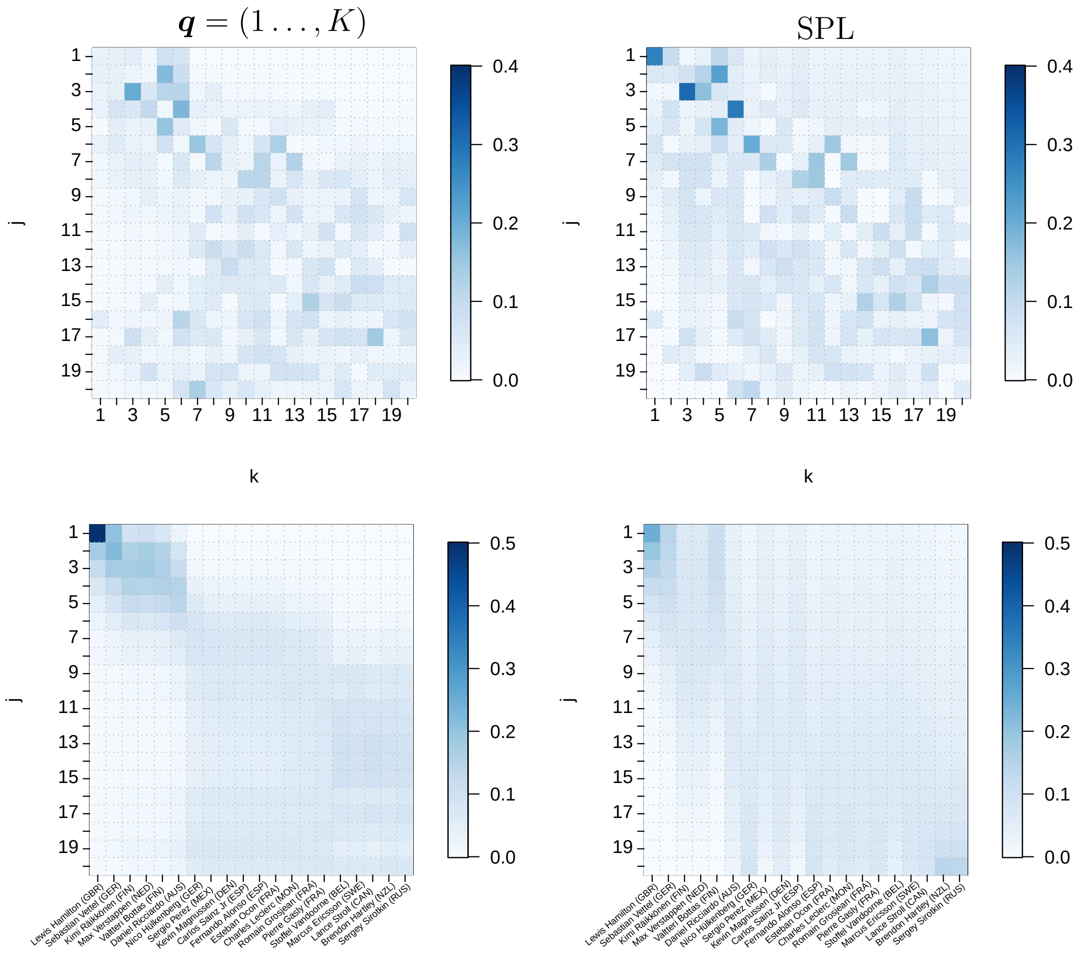
1. Initialise: Let $t = 0$ and $\boldsymbol{\lambda}^{(0)} = (\lambda_1^{(0)}, \lambda_2^{(0)}, \dots, \lambda_K^{(0)})$
2. Let $\lambda_k^{(t)} = \dfrac{a_k - 1 + w_k}{b_k + \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{K}\delta_{ij}(k)\left(\sum\limits_{m=j}^{K}\lambda_{x_{im}^*}^{(t-1)}\right)^{-1}}$ for $k = 1,\dots,K$.
3. Rescale:
   - calculate $\Sigma^{(t)} = \sum\limits_{k=1}^{K}\lambda_k^{(t)}$.
   - let $\lambda_k^{(t)} \to \lambda_k^{(t)}/\Sigma^{(t)}$ for $k = 1,\dots,K$.
4. Set $t = t + 1$ and return to Step 2.

Here the quantity $w_k = \sum_{i=1}^n\sum_{k=1}^K \mathbb{I}(x_{ij}^* = k)$ is the number of times the parameter λ_k^* represents an entity in the collection of permuted rankings and $\delta_{ij}(k) = \mathbb{I}(k \in \{x_{ij}^*,\dots,x_{in_i}^*\})$ is an indicator variable over the event that entity k appears no higher than position j in the permuted ranking i. Note that $w_k = n$ given we only consider complete rankings for the Extended Plackett–Luce model. Following Hunter (2004) we take $\boldsymbol{\lambda}^{(t)}$ to be the MAP estimates when the L_2-norm of the change in the value of the parameter vector is less than 10^{-9}, that is, when

$$||\boldsymbol{\lambda}^{(t)} - \boldsymbol{\lambda}^{(t-1)}||_2 \;=\; \sqrt{\sum_{k=1}^{K}\left(\lambda_k^{(t)} - \lambda_k^{(t-1)}\right)^2} \;<\; 10^{-9}$$

is satisfied. For the interested reader further details are given in Caron and Doucet (2012).

4 F1 analysis



F1 data: heat maps showing the discrepancies $d_{jk} = |\Pr(\tilde{x}_j = k|\mathcal{D}) - \Pr(x_j = k)|$ (top left) and predictive probabilities $\Pr(\tilde{x}_j = k|\mathcal{D})$ (bottom left) for the EPL analysis with $\boldsymbol{q} = (1\ldots,K)$; the corresponding quantities for SPL analysis are shown on the right hand side.

4.1 Prior Sensitivity

$\boldsymbol{q} = (1, \ldots, \boldsymbol{K})$

	k: 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\Pr(\sigma_1 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0	0.02	0.02	0.08	0.08	0.23	0.26	0.20	0.01	0.07	0.01	0	0.02
$\Pr(\sigma_2 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0	0.03	0.03	0.09	0.10	0.21	0.22	0.19	0.02	0.08	0.01	0	0.02
$\Pr(\sigma_3 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0	0.04	0.03	0.11	0.12	0.18	0.17	0.17	0.03	0.09	0.02	0.01	0.03
$\Pr(\sigma_4 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0	0.05	0.04	0.12	0.14	0.15	0.13	0.15	0.04	0.10	0.03	0.01	0.04
$\Pr(\sigma_5 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0	0.07	0.06	0.13	0.14	0.11	0.09	0.11	0.06	0.11	0.05	0.01	0.06
$\Pr(\sigma_6 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0.01	0.09	0.07	0.13	0.13	0.07	0.06	0.08	0.08	0.10	0.07	0.02	0.09
$\Pr(\sigma_7 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0.01	0.10	0.09	0.12	0.10	0.04	0.04	0.05	0.10	0.10	0.10	0.04	0.11
$\Pr(\sigma_8 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0.02	0.12	0.10	0.09	0.08	0.02	0.02	0.03	0.11	0.08	0.14	0.06	0.13
$\Pr(\sigma_9 = k\|\mathcal{D})$	0	0	0	0	0	0	0.01	0.03	0.12	0.11	0.07	0.05	0.01	0.01	0.01	0.13	0.07	0.15	0.09	0.14
$\Pr(\sigma_{10} = k\|\mathcal{D})$	0	0	0	0	0	0	0.01	0.05	0.12	0.12	0.04	0.03	0	0.01	0.01	0.12	0.06	0.16	0.13	0.14
$\Pr(\sigma_{11} = k\|\mathcal{D})$	0	0	0	0	0	0	0.03	0.08	0.11	0.13	0.02	0.02	0	0	0	0.12	0.06	0.14	0.18	0.11
$\Pr(\sigma_{12} = k\|\mathcal{D})$	0	0	0	0	0	0	0.08	0.14	0.09	0.13	0.01	0.01	0	0	0	0.11	0.05	0.09	0.22	0.07
$\Pr(\sigma_{13} = k\|\mathcal{D})$	0	0	0	0	0	0	0.29	0.34	0.03	0.06	0	0	0	0	0	0.05	0.03	0.03	0.15	0.02
$\Pr(\sigma_{14} = k\|\mathcal{D})$	0	0	0	0	0	0.03	0.55	0.31	0.01	0.01	0	0	0	0	0	0.01	0.02	0	0.06	0
$\Pr(\sigma_{15} = k\|\mathcal{D})$	0	0	0	0	0	0.97	0.02	0.01	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{16} = k\|\mathcal{D})$	0	0	0	0.07	0.93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{17} = k\|\mathcal{D})$	0	0	0.04	0.89	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{18} = k\|\mathcal{D})$	0	0.17	0.80	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{19} = k\|\mathcal{D})$	0	0.83	0.16	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{20} = k\|\mathcal{D})$	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

F1 dataset: Marginal posterior distribution for each stage in the ranking process, that is, $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$. Prior distribution specified by $\boldsymbol{q} = (1, \ldots, K)$.

$\boldsymbol{q} = (1, \ldots, 1)$

	k: 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\Pr(\sigma_1 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0	0.04	0.03	0.09	0.1	0.22	0.23	0.2	0.01	0.06	0.01	0	0.01
$\Pr(\sigma_2 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0	0.04	0.04	0.10	0.11	0.20	0.20	0.19	0.02	0.06	0.01	0	0.01
$\Pr(\sigma_3 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0	0.06	0.05	0.11	0.13	0.18	0.17	0.16	0.02	0.07	0.02	0	0.02
$\Pr(\sigma_4 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0.01	0.07	0.06	0.13	0.13	0.15	0.13	0.14	0.04	0.08	0.03	0.01	0.03
$\Pr(\sigma_5 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0.01	0.09	0.07	0.14	0.14	0.11	0.10	0.11	0.05	0.08	0.04	0.01	0.04
$\Pr(\sigma_6 = k\|\mathcal{D})$	0	0	0	0	0	0	0	0.02	0.10	0.09	0.13	0.12	0.07	0.07	0.08	0.07	0.09	0.07	0.02	0.07
$\Pr(\sigma_7 = k\|\mathcal{D})$	0	0	0	0	0	0	0.01	0.02	0.11	0.10	0.11	0.10	0.04	0.04	0.05	0.09	0.09	0.09	0.03	0.09
$\Pr(\sigma_8 = k\|\mathcal{D})$	0	0	0	0	0	0	0.01	0.03	0.12	0.11	0.08	0.07	0.02	0.03	0.03	0.11	0.08	0.13	0.05	0.12
$\Pr(\sigma_9 = k\|\mathcal{D})$	0	0	0	0	0	0	0.02	0.05	0.12	0.10	0.05	0.04	0.01	0.01	0.02	0.12	0.08	0.16	0.08	0.15
$\Pr(\sigma_{10} = k\|\mathcal{D})$	0	0	0	0	0	0	0.03	0.07	0.10	0.11	0.03	0.02	0	0.01	0.01	0.13	0.07	0.16	0.11	0.15
$\Pr(\sigma_{11} = k\|\mathcal{D})$	0	0	0	0	0	0	0.05	0.10	0.08	0.10	0.01	0.01	0	0	0	0.14	0.07	0.14	0.15	0.14
$\Pr(\sigma_{12} = k\|\mathcal{D})$	0	0	0	0	0	0	0.10	0.14	0.05	0.09	0	0.01	0	0	0	0.13	0.06	0.10	0.20	0.11
$\Pr(\sigma_{13} = k\|\mathcal{D})$	0	0	0	0	0	0	0.28	0.26	0.02	0.04	0	0	0	0	0	0.06	0.05	0.04	0.19	0.05
$\Pr(\sigma_{14} = k\|\mathcal{D})$	0	0	0	0	0	0.03	0.47	0.27	0	0.01	0	0	0	0	0	0.02	0.05	0.01	0.13	0.01
$\Pr(\sigma_{15} = k\|\mathcal{D})$	0	0	0	0	0	0.97	0.02	0.01	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{16} = k\|\mathcal{D})$	0	0	0	0.07	0.93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{17} = k\|\mathcal{D})$	0	0	0.04	0.89	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{18} = k\|\mathcal{D})$	0	0.20	0.76	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{19} = k\|\mathcal{D})$	0.01	0.79	0.20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\Pr(\sigma_{20} = k\|\mathcal{D})$	0.99	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

F1 dataset: Marginal posterior distribution for each stage in the ranking process, that is, $\Pr(\sigma_j = k|\mathcal{D})$ for $j, k \in \{1, \ldots, K\}$. Prior distribution specified by $\boldsymbol{q} = \mathbf{1}$.

5 Datasets

Here we provide the datasets considered in Section 6 of the paper. Note that these datasets, along with the synthetic datasets anlaysed within Section 4 of the paper, can also be found at the GitHub repository `https://github.com/srjresearch/ExtendedPL`.

5.1 Song data

Observation					Number of occurrences
3	2	1	4	5	19
2	3	1	4	5	10
1	3	2	4	5	9
4	2	1	3	5	8
1	2	3	4	5	7
3	1	2	4	5	6
3	2	1	5	4	6
5	2	1	3	4	5
2	1	3	4	5	4
2	4	1	3	5	3
4	1	2	3	5	2
4	3	1	2	5	2
5	2	1	4	3	2

Entity number	1	2	3	4	5
Entity name	Score	Instrument	Solo	Benediction	Suit

Table 4: Song dataset

5.2 F1 data

	Rank																			
Observation	1		···		5			···		10			···		15			···		20
1	3	18	4	2	9	16	13	20	1	19	8	15	10	11	14	5	12	7	6	17
2	3	20	18	7	12	13	9	1	6	15	19	8	14	10	5	11	17	4	16	2
3	2	20	4	18	16	13	9	3	19	12	15	8	1	11	17	6	5	7	10	14
4	18	4	8	3	19	10	9	11	1	14	6	7	12	20	5	2	16	13	15	17
5	18	20	16	3	2	12	19	9	8	10	11	14	6	17	1	15	4	5	7	13
6	2	3	18	4	20	15	7	13	16	19	6	8	12	1	5	17	11	14	10	9
7	3	20	16	2	18	4	13	19	15	10	7	5	12	8	6	1	17	9	14	11
8	18	16	4	2	3	12	20	19	13	10	5	1	6	14	17	9	11	8	15	7
9	16	4	3	5	12	15	8	9	10	6	7	19	11	17	1	18	14	2	20	13
10	3	18	4	20	2	13	15	9	12	8	1	11	7	17	16	5	19	6	10	14
11	18	20	4	16	13	5	8	15	6	14	12	19	1	7	10	9	11	3	17	2
12	18	3	4	2	20	7	12	9	19	5	14	13	15	8	6	17	11	1	16	10
13	3	18	16	20	8	15	5	12	7	6	19	17	11	14	1	2	4	10	9	13
14	18	4	20	3	16	15	8	19	11	17	10	1	13	7	6	12	2	9	14	5
15	18	16	3	20	4	2	9	19	10	13	6	1	7	11	5	8	14	12	17	15
16	18	20	3	4	16	2	10	12	15	8	5	13	6	9	11	1	19	17	7	14
17	18	20	16	2	4	3	8	5	15	19	7	6	14	9	1	17	11	10	13	12
18	4	16	18	3	20	13	19	8	14	6	1	7	17	11	10	2	5	9	15	12
19	16	3	4	18	20	13	10	1	6	7	15	11	17	14	12	5	2	8	19	9
20	18	16	4	2	20	3	10	5	12	8	14	19	7	1	15	9	17	11	13	6
21	18	3	16	2	20	19	10	8	5	12	9	14	11	1	17	7	15	6	4	13

Table 5: F1 dataset

Entity number	Driver (Country)
1	Stoffel Vandoorne (BEL)
2	Daniel Ricciardo (AUS)
3	Sebastian Vettel (GER)
4	Kimi Räikkönen (FIN)
5	Romain Grosjean (FRA)
6	Marcus Ericsson (SWE)
7	Pierre Gasly (FRA)
8	Sergio Perez (MEX)
9	Fernando Alonso (ESP)
10	Charles Leclerc (MON)
11	Lance Stroll (CAN)
12	Kevin Magnussen (DEN)
13	Nico Hülkenberg (GER)
14	Brendon Hartley (NZL)
15	Esteban Ocon (FRA)
16	Max Verstappen (NED)
17	Sergey Sirotkin (RUS)
18	Lewis Hamilton (GBR)
19	Carlos Sainz Jr (ESP)
20	Valtteri Bottas (FIN)

Table 6: F1 entity labels

References

Caron, F. and Doucet, A. (2012). Efficient Bayesian inference for generalized Bradley–Terry models. *Journal of Computational and Graphical Statistics*, 21(1):174–196.

Hunter, D. R. (2004). MM algorithms for generalized Bradley-Terry models. *The Annals of Statistics*, 32(1):384–406.