Bayesian Nonparametric Models for Ranking Data

François Caron[1], Yee Whye Teh[1] and Brendan Murphy[2]

[1]Dept of Statistics, University of Oxford, UK
[2]School of Mathematical Sciences, University College Dublin, Ireland

BNPSki, January 2014

Ranked Data

"Rank these 5 movies by preference"

"Rank your 5 favourite movies among these 30"

"Rank your 5 favourite movies"

Aim

- Bayesian nonparametric model to model top-m partial rankings of a potentially infinite number of items.
- Each item is modelled using a positive rating parameter that is inferred from partial rankings.
- Develop efficient computational procedure for posterior simulation.
- Mixture generalizations.
- Application to preferences in college programmes

Preferences for college degree programmes

- Application for college degree programmes in Ireland in 2000
- 53 757 applicants provided top-10 rankings
- 533 degree programmes available for selection
- Heterogenity in the data

Table: Two samples from the preference data.

Rank	CAO code	College	Degree Programme
1	DN002	University College Dublin	Medicine
2	GY501	NUI - Galway	Medicine
3	CK701	University College Cork	Medicine
4	DN006	University College Dublin	Physiotherapy
5	TR053	Trinity College Dublin	Physiotherapy
6	DN004	University College Dublin	Radiotherapy
7	TR007	Trinity College Dublin	Clinical Speech
8	FT223	Dublin IT	Human Nutrition
9	TR084	Trinity College Dublin	Social Work
10	DN007	University College Dublin	Social Science

Rank	CAO code	College	Degree Programme
1	MI005	Mary Immaculate Limerick	Education - Primary Teaching
2	CK301	University College Cork	Law
3	CK105	University College Cork	European Studies
4	CK107	University College Cork	Language - French
5	CK101	University College Cork	Arts

Parametric Plackett-Luce Model

- Population of M items $X_1, \ldots, X_M$.
- To each item X_k assign a rating parameter $w_k > 0$.

Stage-wise interpretation:

- Pick first item with probabilities proportional to w_k's.
- Remove first item.
- Pick second item with probabilities proportional to w_k's.
- Remove second item.
- ...

The probability of a ranking $(X_{\rho_1}, \ldots, X_{\rho_M})$, with $\rho = (\rho_1, \ldots, \rho_M)$ a permutation, is

$$P(\rho|w) = \prod_{i=1}^{M} \frac{w_{\rho_i}}{\sum_{k=1}^{M} w_k - \sum_{j=1}^{i-1} w_{\rho_j}}$$

[Luce, 1959, Plackett, 1975], [Gormley and Murphy, 2008, Gormley and Murphy, 2009]

Nonparametric Plackett-Luce Model

- Population of items $X_1, X_2, \ldots$ of infinite size.
- Each item X_k assigned a rating parameter $w_k > 0$.

Stage-wise interpretation:

- Pick first item with probabilities proportional to w_k's.
- Remove first item.
- Pick second item with probabilities proportional to w_k's.
- Remove second item.
- ...

The probability of a (finite, partial) ranking $X_{\rho_1}, \ldots, X_{\rho_m}$ is

$$P(\rho|w) = \prod_{i=1}^{m} \frac{w_{\rho_i}}{\sum_{k=1}^{\infty} w_k - \sum_{j=1}^{i-1} w_{\rho_j}}$$

Thurstonian Interpretation

- For each item X_k define an exponential random variate:

$$z_k \sim \mathrm{Exp}(w_k)$$

- Induces a partial permutation ρ such that $z_{\rho_1} < \cdots < z_{\rho_m} < \cdots$

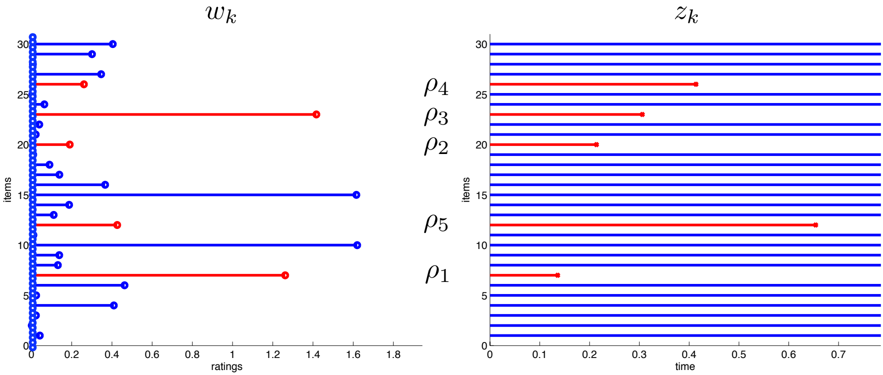
A race among items:

- z_k is the time that item X_k finished the race.
- ρ is the order of champion, runner-up, second runner-up...

The probability of a partial permutation is:

$$\begin{aligned} P(\rho|w) =& P(z_{\rho_1} < z_{\rho_2} < \cdots < z_{\rho_m} < \text{everything else}) \\ =& \prod_{i=1}^{m} \frac{w_{\rho_i}}{\sum_{k=1}^{\infty} w_k - \sum_{j=1}^{i-1} w_{\rho_j}} \end{aligned}$$

Thurstonian Interpretation



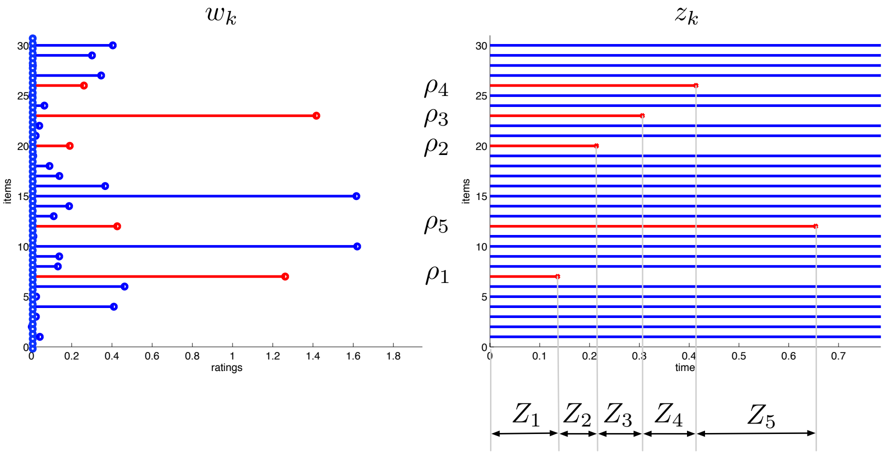
$$P(\rho|w) = P(z_{\rho_1} < z_{\rho_2} < \cdots < z_{\rho_m} < \text{everything else})$$

$$= \prod_{i=1}^{m} \frac{w_{\rho_i}}{\sum_{k=1}^{\infty} w_k - \sum_{j=1}^{i-1} w_{\rho_j}}$$

Data Augmentation

- Reparametrize in terms of inter-arrival durations:

$$Z_1 = z_{\rho_1}, Z_2 = z_{\rho_2} - z_{\rho_1}, \dots Z_m = z_{\rho_m} - z_{\rho_{m-1}}$$



Data Augmentation

- Augmented system with auxiliary variables $Z_1, \dots, Z_m$:

$$Z_i | \rho, w, X \sim \mathrm{Exp}\left(\sum_{k=1}^{\infty} w_k - \sum_{j=1}^{i-1} w_{\rho_j} \right)$$

Joint probability is:

$$P(\rho, Z | w, X) = \prod_{i=1}^{m} w_{\rho_i} \exp\left[-\left(\sum_{k=1}^{\infty} w_k - \sum_{j=1}^{i-1} w_{\rho_j} \right) Z_i \right]$$

What prior to use for w, X?

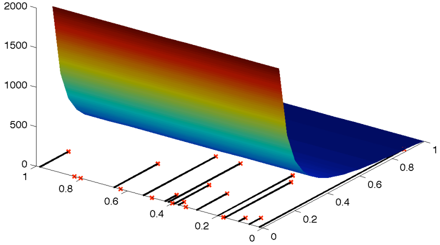
- Independent Gamma's for w_k's do not work.
- Gamma process.
- Better: completely random measures.

Completely Random Measures

- Lévy intensity $\lambda(w)$.
- Base distribution H with density $h(x)$.
- Random atomic measure

$$G = \sum_{k=1}^{\infty} w_k \delta_{X_k}$$

Construction: two-dimensional Poisson process $N = \{w_k, X_k\}$ with intensity $\lambda(w)h(x)$:



[Kingman, 1967, Lijoi and Prünster, 2010]

Completely Random Measures

Conditions on Lévy intensity:

$$\int_0^\infty \lambda(w)dw = \infty \qquad \Rightarrow \text{ infinitely many items.}$$

$$\int_0^\infty (1 - e^{-w})\lambda(w)dw < \infty \qquad \Rightarrow \text{ finite total } \sum_{k=1}^\infty w_k.$$

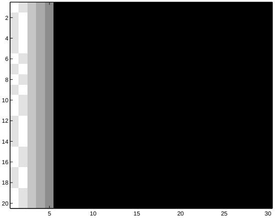
Plackett-Luce sampling without replacement:

- Pick first item with probabilities proportional to w_k's; remove.
- Pick second item with probabilities proportional to w_k's; remove;
- ...

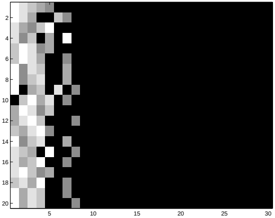
$\Rightarrow$ Size-biased sampling of the atoms in G.

Prior Draws

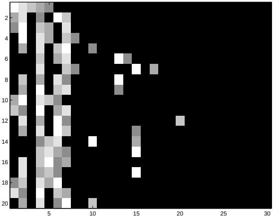
Generalized Gamma process with $\boldsymbol{\lambda(w) = \frac{\alpha}{\Gamma(1-\sigma)} w^{-\sigma-1} e^{-\tau w}}$, $\boldsymbol{\tau = 1}$.



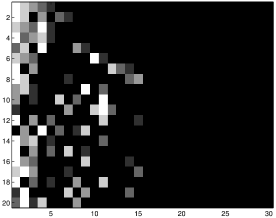
(a) $\boldsymbol{\alpha = 0.1, \sigma = 0}$



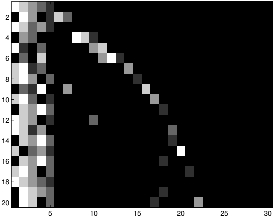
(b) $\boldsymbol{\alpha = 1, \sigma = 0}$



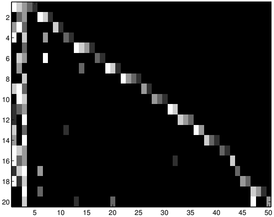
(c) $\boldsymbol{\alpha = 3, \sigma = 0}$



(d) $\boldsymbol{\alpha = 1, \sigma = 0.1}$



(e) $\boldsymbol{\alpha = 1, \sigma = 0.5}$



(f) $\boldsymbol{\alpha = 1, \sigma = 0.9}$

Posterior Characterization

- Observe L partial rankings $Y_\ell = (Y_{\ell 1}, \ldots, Y_{\ell m_\ell})$ for $\ell = 1, \ldots, L$.
- $X_1^*, \ldots, X_K^*$ are the K unique items among observations
- Associated auxiliary variables $Z_{\ell 1}, \ldots, Z_{\ell m_\ell}$.

Theorem

The posterior distribution given partial rankings Y and auxiliary variables Z is a CRM with fixed atoms:

$$G|Y, Z = G^* + \sum_{k=1}^{K} w_k^* \delta_{X_k^*}$$

where G^* and $w_1^*, \ldots, w_K^*$ are mutually independent. The law of G^* is still CRM with Lévy intensity $\lambda^*(w) = \lambda(w) e^{-w(\sum_{\ell i} Z_{\ell i})}$ while the masses have distributions,
$P(w_k^*|Y, Z) \propto (w_k^*)^{n_k} e^{-w_k^*(\sum_{\ell i} \delta_{\ell i k} Z_{\ell i})} \lambda(w_k^*)$.

- Characterization similar to that for normalized random measures.

[Prünster, 2002, James, 2002, James et al., 2009]

Bayesian Inference via Gibbs Sampling

- Rating parameters w_k^* of observed items.
- Latent variables $Z_{\ell i}$.
- CRM G^* containing unobserved items.
- Marginalize out G^*, keeping only its total mass w_*^*.

Easy Gibbs sampler for generalized gamma process class of CRM

$$Z_{\ell i} | \text{rest} \sim \text{Exponential}$$
$$w_k^* | \text{rest} \sim \text{Gamma}$$
$$w_*^* | \text{rest} \sim \text{Exponentially tilted stable}$$

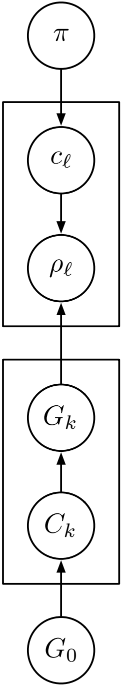
Nonparametric Plackett-Luce Mixture Model

Partial rankings reflecting the preferences of a heterogenous population.

$$\begin{aligned}
\pi &\sim \mathrm{GEM}(\theta) \\
c_\ell | \pi &\sim \mathrm{Discrete}(\pi) \\
Y_\ell | c_\ell, G_{c_\ell} &\sim \mathrm{PL}(G_{c_\ell})
\end{aligned}$$

$$\begin{aligned}
G_k | C_k &\sim \mathrm{Gamma}(\alpha H + C_k, \tau + \phi) \\
C_k | G_0 &\sim \mathrm{Poisson}(\phi G_0) \\
G_0 &\sim \mathrm{Gamma}(\alpha H, \tau)
\end{aligned}$$

- $G_k \sim \mathrm{Gamma}(\alpha H, \tau)$ does not work.



Irish University Programme Applications

- 53757 Irish university applicants.
- Each applicant ranks their top-10 desired university programmes.
- Point estimate of the partition

$$\widehat{c} = \arg\min_{c^{(i)} \in \{c^{(1)},\dots,c^{(N)}\}} \sum_k \sum_\ell (\delta_{c_k^{(i)} c_\ell^{(i)}} - \zeta_{k\ell})^2$$

where the coclustering matrix ζ is obtained with

$$\zeta_{k\ell} = \frac{1}{N} \sum_{i=1}^{N} \delta_{c_k^{(i)} c_\ell^{(i)}}$$

$c^{(i)}, i = 1, \dots, N$ are the Monte Carlo samples and $\delta_{k\ell} = 1$ if $k = \ell$, 0 otherwise.

Irish University Programme Applications

Table: Description of the different clusters. The size of the clusters, the entropy and a cluster description are provided.

Cluster	Size	Entropy	Description	Cluster	Size	Entropy	Description
1	3325	0.72	Social Science/Tourism	14	1918	0.71	Engineering
2	3214	0.71	Science	15	1835	0.48	Teaching/Arts
3	3183	0.64	Business/Commerce	16	1835	0.68	Art/Music - Dublin
4	2994	0.58	Arts	17	1740	0.71	Engineering - Dublin
5	2910	0.63	Business/Marketing - Dublin	18	1701	0.55	Medicine
6	2879	0.68	Construction	19	1675	0.70	Arts/Religion/Theology
7	2803	0.66	CS - outside Dublin	20	1631	0.76	Arts/History - Dublin
8	2225	0.67	CS - Dublin	21	1627	0.66	Galway
9	2303	0.67	Arts/Social - outside Dublin	22	1392	0.70	Limerick
10	2263	0.63	Business/Finance - Dublin	23	1273	0.65	Law
11	2198	0.65	Arts/Psychology - Dublin	24	1269	0.72	Business - Dublin
12	2086	0.63	Cork	25	1225	0.79	Arts/Bus. - Dublin
13	2029	0.64	Comm./Journalism - Dublin	26	47	0.96	Mixed

Irish University Programme Applications

Table: Cluster 7: Computer Science - outside Dublin

Rank	Aver. Norm. Weight	College	Degree Programme
1	0.081	Cork IT	Computer Applications
2	0.075	Limerick IT	Software Development
3	0.072	University of Limerick	Computer Systems
4	0.064	Waterford IT	Applied Computing
5	0.061	Cork IT	Software Dev & Comp Net
6	0.046	IT Carlow	Computer Networking
7	0.038	Athlone IT	Computer and Software Engineering
8	0.036	University College Cork	Computer Science
9	0.033	Dublin City University	Computer Applications
10	0.033	University of Limerick	Information Technology

Table: Cluster 8: Computer Science - Dublin

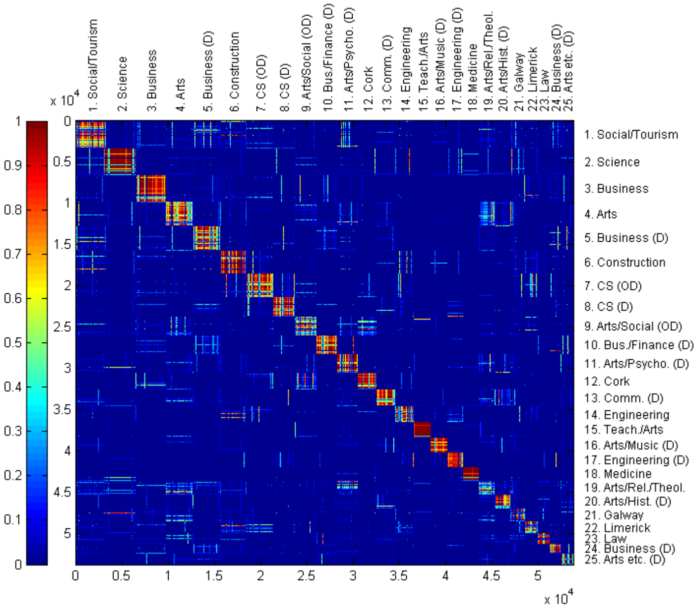
Rank	Aver. Norm. Weight	College	Degree Programme
1	0.141	Dublin City University	Computer Applications
2	0.054	University College Dublin	Computer Science
3	0.049	NUI - Maynooth	Computer Science
4	0.043	Dublin IT	Computer Science
5	0.040	National College of Ireland	Software Systems
6	0.038	Dublin IT	Business Info. Systems Dev.
7	0.036	Trinity College Dublin	Computer Science
8	0.035	Dublin IT	Applied Sciences/Computing
9	0.030	Trinity College Dublin	Information & Comm. Tech.
10	0.029	University College Dublin	B.A. (Computer Science)

Irish University Programme Applications

Table: Cluster 12: Cork

Rank	Aver. Norm. Weight	College	Degree Programme
1	0.105	University College Cork	Arts
2	0.072	University College Cork	Computer Science
3	0.072	University College Cork	Commerce
4	0.067	University College Cork	Business Information Systems
5	0.057	Cork IT	Computer Applications
6	0.049	Cork IT	Software Dev & Comp Net
7	0.035	University College Cork	Finance
8	0.031	University College Cork	Law
9	0.031	University College Cork	Accounting
10	0.026	University College Cork	Biological and Chemical Sciences

Irish University Programme Applications



Summary

- A Bayesian nonparametric Plackett-Luce model for partial rankings.
- Completely random measures and posterior characterization.
- Easy Gibbs sampling for posterior simulation.
- Mixture generalization.
- Future:
  - Dependent general CRM models

Bibliography I

Caron, F., Teh, Y. W., and Murphy, B. (2013).
Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes.
To appear in Annals of Applied Statistics.

Gormley, I. C. and Murphy, T. B. (2008).
Exploring voting blocs with the Irish electorate: a mixture modeling approach.
Journal of the American Statistical Association, 103(483):1014–1027.

Gormley, I. C. and Murphy, T. B. (2009).
A grade of membership model for rank data.
Bayesian Analysis, 4(2):265–296.

James, L., Lijoi, A., and Prünster, I. (2009).
Posterior analysis for normalized random measures with independent increments.
Scandinavian Journal of Statistics, 36(1):76–97.

James, L. F. (2002).
Poisson process partition calculus with applications to exchangeable models and bayesian nonparametrics.
arXiv preprint math/0205093.

Bibliography II

Kingman, J. F. C. (1967).
Completely random measures.
Pacific Journal of Mathematics, 21(1):59–78.

Lijoi, A. and Prünster, I. (2010).
Models beyond the Dirichlet process.
In Hjort, N. L., C. Holmes, P. M., and Walker, S. G., editors, *Bayesian Nonparametrics*.
Cambridge University Press.

Luce, R. D. (1959).
Individual choice behavior: A theoretical analysis.
Wiley.

Plackett, R. (1975).
The analysis of permutations.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 24(2):193–202.

Prünster, I. (2002).
Random probability measures derived from increasing additive processes and their application to Bayesian statistics.
PhD thesis, University of Pavia.