Modelling preference data with the Wallenius distribution

Brunero Liseo

MEMOTEF, Sapienza Università di Roma
brunero.liseo@uniroma1.it

with Clara Grazian & Fabrizio Leisen, University of Kent

Padova, Workshop PRIN 2015

18th February 2019
We naturally tend to rank objects in everyday life. 
Rankings arise when

- users express preferences about products and services,
- voters cast ballots in elections,
- research projects are evaluated based on their merits,
- genes are ordered based on their expression levels under various experimental conditions.

A ranking represents a statement about the relative quality or relevance of the items being ranked.
Modern technologies (smartphones, web, etc.) made available a huge amount of ranked data.

They can provide information about social and psychological behavior, marketing strategies and political preferences.
Introduction

There exists a huge literature on the statistical analysis of ranked data. There are of different kinds:

- complete ranking (List of all items in a specific order)
- partial ranking (Only a subset of the items have been ranked)
- preference data (chose few elements from a list)
- pairwise comparison (sport competitions)

Different models are needed for different types of data.
Permutations vs. Ranking

Let $I = \{1, 2, \ldots, n\}$ be a set of items to be ranked. A ranking $\pi$ is a bijection from $I$ onto itself. A ranking has a one-to-one correspondence with a permutation, say

$$\sigma = \pi^{-1}, \quad \sigma \in \mathcal{P}_n$$
I will use $\pi(i)$ to denote the position given to item $i$.
I will use $\sigma(j) = \pi^{-1}(j)$ to denote the item assigned to position $j$.
$\pi$ and $\sigma = \pi^{-1}$ are vectors whose $i$-th component is $\pi(i)$ or $\sigma(i) = \pi^{-1}(i)$.

For example, a ranking $B > C > A > D$ produce

\[
\begin{align*}
\pi & \to 3 \quad 1 \quad 2 \quad 4 \\
\sigma & \to 2 \quad 3 \quad 1 \quad 4.
\end{align*}
\]
Literature review

There are several models available in the literature to model ranked or preference data.

- **Thurstone model** (or law of comparative judgement): a ranking process where the ranking $\pi_j$ of $n$ objects given by a random sample of $k$ independent judges ($j = 1, \ldots, k$) is determined by the relative ordering of $n$ random utilities ($X_{1j}, X_{2j}, \ldots, X_{nj}$).
  
  **Disadvantages:** It works well with few items

- **Plackett-Luce model:** assumes that there is a latent score $w_i$ ($i = 1, 2, \ldots, n$) associated to each item $i$.

- **Bradley and Terry model:** particularly useful when data refers to pairwise comparisons, like in sport tournaments.

- **Mallows model:** It is a model on permutations, with a “mode” and a scale parameter.
Our goal is to suggest a new perspective on ranking literature. We notice that, in some occasions, ranked items can be further classified into categories of different importance.

We will make use of an extension of the Hypergeometric distribution, namely the **Wallenius distribution** [Wallenius, 1963](#). The Wallenius distribution arises quite naturally in situations where

- sampling is performed without replacement
- units in the population have different probabilities to be drawn
Wallenius distribution

Consider a urn with $N$ balls of $c$ different colors.

- For $i = 1, \ldots, c$ there are $m_i$ balls of color $i$, such that $\sum_{i=1}^{c} m_i = N$.
- In addition, color $i$ has a priority $\omega_i > 0$ which specifies its relative importance with respect to the other colors ($i = 1, \ldots, c$).

Suppose we have drawn $n$ balls without replacement from the urn and let $X_n = (X_{1n}, X_{2n}, \ldots, X_{cn})$ denote the frequencies of balls of different colors in the sample. Let $Z_{n+1}$ be the colour of the ball drawn at time $n+1$. Then

$$P(Z_{n+1} = i | X_n) = \frac{(m_i - X_{in}) \omega_i}{\sum_{j=1}^{c} (m_j - X_{jn}) \omega_j}.$$
Wallenius probability distribution function

For a given integer \( n \) (number of sampled balls), parameters \( \mathbf{m} = (m_1, \ldots, m_c) \) (numbers of initial balls for each color), \( \omega = (\omega_1, \ldots, \omega_c) \) (priorities of colors) the probability of observing a vector of color frequencies \( \mathbf{x} = (x_1, \ldots, x_c) \) is

\[
P(\mathbf{x}; n, \mathbf{m}, \omega) = \prod_{j=1}^{c} \binom{m_j}{x_j} \int_0^1 \prod_{j=1}^{c} \left(1 - t^{\omega_j/d}\right)^{x_j} dt,
\]

where \( d = \sum_{j=1}^{c} \omega_j(m_j - x_j) \) [Chesson, 1976]. Wallenius [1963] only provided the above expression for \( c = 2 \).
Wallenius distribution

For a given integer $n$ (number of sampled balls), parameters $\mathbf{m} = (m_1, \ldots, m_c)$ (numbers of initial balls for each color), $\omega = (\omega_1, \ldots, \omega_c)$ (priorities of colors) the probability of observing a vector of color frequencies $\mathbf{x} = (x_1, \ldots, x_c)$ is

$$P(\mathbf{x}; n, \mathbf{m}, \omega) = \prod_{j=1}^{c} \binom{m_j}{x_j} \int_0^1 \prod_{j=1}^{c} \left(1 - t^{\omega_j/d}\right)^{x_j} dt,$$

where $d = \sum_{j=1}^{c} \omega_j(m_j - x_j)$ [Chesson, 1976]
Some uses of Wallenius distribution

- **Audit problems**: monetary unit sampling
- **Wada** (World Anti Doping Agency) Control Test are distributed among sports and disciplines according to their corresponding number of athletes and priority (risk assessment).
- Bootstrap calculations in finite population sampling (Ranalli, 2018).
Bayesian inference for the Wallenius distribution

For $h = 1, \ldots, k$ (the sample size), let $\mathbf{x}_h = (x_{h1}, \ldots, x_{hc})$ be a draw of $n_h$ balls from the Wallenius urn and $\sum_{j=1}^{c} x_{hj} = n_h$.

$\omega$ is the ‘unknown’ parameter vector.

For a given prior distribution $\pi(\omega)$ the posterior distribution is

$$
\pi(\omega|\mathbf{x}_1, \ldots, \mathbf{x}_k) \propto \pi(\omega) \prod_{h=1}^{k} \left[ \int_{0}^{1} \prod_{j=1}^{c} \left(1 - t_h^{\omega_j/d_h}\right)^{x_{hj}} \, dt_h \right],
$$

with $d_h = \sum_{j=1}^{c} \omega_j (m_j - x_{hj})$.

The above posterior is impractical, due to the presence of integrals in the likelihood.

This has represented a hurdle for the popularization of the Wallenius model.
Approximate Bayesian computation

When the likelihood is not available in closed form or it is too costly to evaluate, use a likelihood-free rejection technique

Foundations

Assume we have data $x \sim f(x|\theta)$ and $\theta \sim \pi(\theta)$. If one keeps jointly simulating

$$\theta^* \sim \pi(\theta) \text{ and } z \sim f(z|\theta^*)$$

until the auxiliary data set $z = x$, then the selected

$$\theta^* \sim \pi(\theta|x)$$

[Rubin, 1984; Tavaré et al. 1997]
When $x$ has a continuous distribution, strict equality $x = z$ is impossible and it is replaced by a tolerance zone

$$\rho(x, z) \leq \epsilon.$$ 

Then, the output is distributed proportional to

$$\pi(\theta | \rho(x, z) \leq \epsilon)$$

[Sisson et al., 2018]
Approximate Bayesian computation

**ABC Rejection algorithm [Marin et al., 2012]**

For $l = 1, \cdots, N$ do

Repeat

- Generate $\theta^*$ from the prior distribution $\pi(\cdot)$
- Generate $z$ from the likelihood $f(\cdot \mid \theta^*)$

Until $\rho(\eta(z), \eta(x)) < \varepsilon$

Set $\theta_l = \theta^*$

End For

where $\eta(x)$ defines a set of (almost never sufficient . . . ) statistics

Fog [2008] provided methods and algorithm to sample from the Wallenius distribution (see the $\mathbb{R}$ package BiasedUrn).
Approximate Bayesian computation

The ABC algorithm samples from the marginal in $\theta$ of

$$
\pi_\varepsilon(\theta, z|x) = \frac{\pi(\theta) f(z|\theta) \mathbb{I}_{A_{\varepsilon,x}}(z)}{\int_{A_{\varepsilon,x} \times \Theta} \pi(\theta) f(z|\theta) dz d\theta}
$$

$A_{\varepsilon,x} = \{z | \rho(z, x) \leq \varepsilon\}$.

The general idea behind ABC methodology is that the summary statistics $\eta(x)$ coupled with a small tolerance level $\varepsilon$ could provide a good approximation of the posterior distribution

$$
\pi_\varepsilon(\theta|x) = \int \pi_\varepsilon(\theta, z|x) dz \approx \pi(\theta|x)
$$
Summary statistics $k$ assessors, $c$ classes of objects

\[
\begin{array}{cccc}
   x_{11} & x_{12} & \cdots & x_{1c} \\
   x_{21} & x_{22} & \cdots & x_{2c} \\
   \cdots & \cdots & \cdots & \cdots \\
   x_{k1} & x_{k2} & \cdots & x_{kc} \\
\end{array}
\]

\[
\eta(x^{(\ell)}) = \hat{p}^{(\ell)} = \frac{1}{k} \sum_{h=1}^{k} p_h^{(\ell)},
\]

\[
p_h^{(\ell)} = \left( \frac{x_{h1}^{(\ell)}}{n_h}, \ldots, \frac{x_{hc}^{(\ell)}}{n_h} \right)
\]

Distance $\rho \rightarrow Distance in variation$ [Bremaud, 1998]

\[
\rho(\hat{p}^{(\ell)}, \hat{p}^{(t)}) = \frac{1}{2} \sum_{j=1}^{c} |\hat{p}_j^{(\ell)} - \hat{p}_j|
\]

Tolerance level $\varepsilon \rightarrow$ [Allingham et al., 2009]
The vector of parameters $\omega = (\omega_1, \ldots, \omega_c)$ takes values on $\mathbb{R}_+^c$ and different priors can be considered.

However, one must take into account that the priority parameters $\omega_j$ must be interpreted in a relative way.

In fact, the quantity $d$ in the p.m.f. of the Wallenius distribution depends on the priority parameters $\omega$.

In particular,

$$d = \sum_{j=1}^c \omega_j (m_j - x_j)$$
An Identifiability issue II

If we consider two different vectors $\omega'$ and $\omega$ such that $\omega' = \kappa \omega$ for $\kappa > 0$, we have that

$$\frac{\omega'_j}{d'} = \frac{\kappa \omega_j}{\sum_{j=1}^{c} \kappa \omega_j (m_j - x_j)} = \frac{\omega_j}{d} = \frac{\omega'_j}{d'}$$

(1)

where $d'$ and $d$ are computed respectively with $\omega'$ and $\omega$.

Equation (1) implies that the p.m.f. of the Wallenius distribution does not change if we consider the vector of priorities $\omega'$ instead of $\omega$.

This induces an identifiability issue, which is resolved through a normalization step.

In this perspective, the most natural way to follow is to assume that $\sum_{j=1}^{c} \omega_j = 1$, and to assume a Dirichlet prior on the normalized vector.
Hereafter we will assume that the Dirichlet prior we adopt in simulations and real data examples are symmetric (i.e., all the hyperparameters are equal).

Our default choice will be to set them all equal to 1, making the prior uniform on its support.

An alternative default choice, especially useful when $c$ is large, is given by $\alpha = 1/c$, as explained in Berger et al. (2015).
The R package BiasedUrn allows the approximate numerical evaluation of the probability mass function of the Wallenius distribution.

In a classical setting, this makes feasible the computation of the MLE.

In a Bayesian setting this enables the implementation of standard MCMC algorithms, such as the Metropolis-Hastings sampler.

Nonetheless, we deem more appropriate to use the ABC approach illustrated in this section for several reasons.
The output of the Bayesian approach is far richer than the one available in a classical setting. For instance, in the next real data examples we are able to easily compute important summaries of the posterior distribution, i.e. the probability $p_{ij} = \Pr(\omega_i > \omega_j)$.

standard MCMC methods require repeated evaluations of the likelihood function. This could lead to an unsustainable computational burden compared to ABC.

Last but not least, a simulation study regarding the behaviour of the maximum likelihood estimator of $\omega$ shows that it typically tends to produce unstable estimates, especially when the 'true' $\omega$ is close to the boundary of the simplex and/or when the number of categories is large.
Simulation study with Dirichlet prior I

Simulation performed with different values of $c$, ranging between 2 and 20, and using 3 configurations for both $m$ and $\omega$

- same number of balls for each colour, i.e. $m_j = m$, $j = 1, \ldots, c$; uniform importance weights, i.e. $\omega_j = \omega$ (HypGeo case);
- increasing values of $m_j$'s ([$1 : c$]) and $\omega$'s (again [$1 : c$] and then normalized);
- increasing values for $m_j$'s ([$1 : c$]) and decreasing values of $\omega$'s ([c : 1] and then normalized);

We have used 3 different sample sizes, ($k = 5, 50, 1000$).

Values of $n_h$'s have been taken to be half of the number of balls in the urn.
**Table:** Simulation study: 20 replications of the experiment with uniform true values for $\omega$ and $m$ for several numbers of categories ($c = 2, \ldots, 20$).

<table>
<thead>
<tr>
<th>$c$</th>
<th>$k=5$ RMSE</th>
<th>acc. rate</th>
<th>$k=50$ RMSE</th>
<th>acc. rate</th>
<th>$k=1000$ RMSE</th>
<th>acc. rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.3084</td>
<td>0.0018</td>
<td>0.3071</td>
<td>0.0017</td>
<td>0.3071</td>
<td>0.0016</td>
</tr>
<tr>
<td>3</td>
<td>0.2922</td>
<td>0.0057</td>
<td>0.2887</td>
<td>0.0057</td>
<td>0.2886</td>
<td>0.0057</td>
</tr>
<tr>
<td>4</td>
<td>0.1714</td>
<td>0.0082</td>
<td>0.1667</td>
<td>0.0080</td>
<td>0.1667</td>
<td>0.0080</td>
</tr>
<tr>
<td>5</td>
<td>0.1118</td>
<td>0.0096</td>
<td>0.1119</td>
<td>0.0095</td>
<td>0.1119</td>
<td>0.0094</td>
</tr>
<tr>
<td>6</td>
<td>0.0912</td>
<td>0.0104</td>
<td>0.0819</td>
<td>0.0102</td>
<td>0.0818</td>
<td>0.0102</td>
</tr>
<tr>
<td>7</td>
<td>0.0811</td>
<td>0.0108</td>
<td>0.0634</td>
<td>0.0110</td>
<td>0.0632</td>
<td>0.0109</td>
</tr>
<tr>
<td>8</td>
<td>0.0662</td>
<td>0.0115</td>
<td>0.0511</td>
<td>0.0113</td>
<td>0.0508</td>
<td>0.0114</td>
</tr>
<tr>
<td>9</td>
<td>0.0576</td>
<td>0.0119</td>
<td>0.0423</td>
<td>0.0117</td>
<td>0.0420</td>
<td>0.0117</td>
</tr>
<tr>
<td>10</td>
<td>0.0534</td>
<td>0.0121</td>
<td>0.0356</td>
<td>0.0121</td>
<td>0.0357</td>
<td>0.0121</td>
</tr>
<tr>
<td>15</td>
<td>0.1326</td>
<td>0.0132</td>
<td>0.1292</td>
<td>0.0131</td>
<td>0.1292</td>
<td>0.0131</td>
</tr>
<tr>
<td>20</td>
<td>0.1845</td>
<td>0.0138</td>
<td>0.1830</td>
<td>0.0136</td>
<td>0.1829</td>
<td>0.0136</td>
</tr>
</tbody>
</table>
## Results II

**Table:** Simulation study; 20 replications of the experiment with increasing values for $\omega$ and $m$ for several numbers of categories ($c = 2, \ldots, 20$).

<table>
<thead>
<tr>
<th>$c$</th>
<th>$K=5$</th>
<th>$K=50$</th>
<th>$K=1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>acc. rate</td>
<td>RMSE</td>
</tr>
<tr>
<td>2</td>
<td>0.4792</td>
<td>0.0014</td>
<td>0.4590</td>
</tr>
<tr>
<td>3</td>
<td>0.4471</td>
<td>0.0048</td>
<td>0.6627</td>
</tr>
<tr>
<td>4</td>
<td>0.4547</td>
<td>0.0093</td>
<td>0.5150</td>
</tr>
<tr>
<td>5</td>
<td>0.4102</td>
<td>0.0115</td>
<td>0.4339</td>
</tr>
<tr>
<td>6</td>
<td>0.3461</td>
<td>0.0112</td>
<td>0.3866</td>
</tr>
<tr>
<td>7</td>
<td>0.3472</td>
<td>0.0124</td>
<td>0.3538</td>
</tr>
<tr>
<td>8</td>
<td>0.3061</td>
<td>0.0137</td>
<td>0.3255</td>
</tr>
<tr>
<td>9</td>
<td>0.2734</td>
<td>0.0144</td>
<td>0.2982</td>
</tr>
<tr>
<td>10</td>
<td>0.2590</td>
<td>0.0172</td>
<td>0.2806</td>
</tr>
<tr>
<td>15</td>
<td>0.1971</td>
<td>0.0189</td>
<td>0.2153</td>
</tr>
<tr>
<td>20</td>
<td>0.1628</td>
<td>0.0198</td>
<td>0.1803</td>
</tr>
</tbody>
</table>
### Results III

**Table:** Simulation study; 20 replications of the experiment with increasing true values for $m$ and decreasing values for $\omega$ for several numbers of categories ($c = 2, \ldots, 20$).

<table>
<thead>
<tr>
<th>$c$</th>
<th>$K=5$</th>
<th>$K=50$</th>
<th>$K=1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>acc. rate</td>
<td>RMSE</td>
</tr>
<tr>
<td>2</td>
<td>0.0117</td>
<td>0.0014</td>
<td>0.0013</td>
</tr>
<tr>
<td>3</td>
<td>0.1464</td>
<td>0.0052</td>
<td>0.2428</td>
</tr>
<tr>
<td>4</td>
<td>0.0888</td>
<td>0.0092</td>
<td>0.0975</td>
</tr>
<tr>
<td>5</td>
<td>0.0633</td>
<td>0.0116</td>
<td>0.0579</td>
</tr>
<tr>
<td>6</td>
<td>0.0890</td>
<td>0.0128</td>
<td>0.0741</td>
</tr>
<tr>
<td>7</td>
<td>0.0882</td>
<td>0.0138</td>
<td>0.0724</td>
</tr>
<tr>
<td>8</td>
<td>0.0961</td>
<td>0.0144</td>
<td>0.0693</td>
</tr>
<tr>
<td>9</td>
<td>0.0907</td>
<td>0.0148</td>
<td>0.0715</td>
</tr>
<tr>
<td>10</td>
<td>0.0875</td>
<td>0.0154</td>
<td>0.0709</td>
</tr>
<tr>
<td>15</td>
<td>0.0940</td>
<td>0.0172</td>
<td>0.0753</td>
</tr>
<tr>
<td>20</td>
<td>0.0891</td>
<td>0.0182</td>
<td>0.0732</td>
</tr>
</tbody>
</table>
Movies Dataset I

Dataset describes ratings from MovieLens, a movie recommendation service (http://grouplens.org/datasets/movielens/). [The dataset is likely to change over time.

- 105,339 ratings across 10,329 movies
- 668 users between April 03, 1996 and January 09, 2016
- users were randomly selected by MovieLens, with no demographic information;
- each of them has rated at least 20 movies;
- the movies in the dataset were described by genre, following the Imdb information (18 genres)
- 5-star rating (with half-star increments)
<table>
<thead>
<tr>
<th>Category</th>
<th>$\omega$</th>
<th>Category</th>
<th>$\omega$</th>
<th>Category</th>
<th>$\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>War</td>
<td>1.969</td>
<td>Western</td>
<td>1.616</td>
<td>Adventure</td>
<td>1.566</td>
</tr>
<tr>
<td></td>
<td>(1.620)</td>
<td></td>
<td>(1.862)</td>
<td></td>
<td>(1.467)</td>
</tr>
<tr>
<td>Crime</td>
<td>1.762</td>
<td>Animation</td>
<td>1.601</td>
<td>Romance</td>
<td>1.496</td>
</tr>
<tr>
<td></td>
<td>(1.078)</td>
<td></td>
<td>(1.150)</td>
<td></td>
<td>(0.917)</td>
</tr>
<tr>
<td>Children</td>
<td>1.667</td>
<td>Drama</td>
<td>1.596</td>
<td>Horror</td>
<td>1.420</td>
</tr>
<tr>
<td></td>
<td>(2.021)</td>
<td></td>
<td>(1.082)</td>
<td></td>
<td>(1.478)</td>
</tr>
<tr>
<td>Film-Noir</td>
<td>1.661</td>
<td>Documentary</td>
<td>1.594</td>
<td>Comedy</td>
<td>1.277</td>
</tr>
<tr>
<td></td>
<td>(1.803)</td>
<td></td>
<td>(1.759)</td>
<td></td>
<td>(0.926)</td>
</tr>
<tr>
<td>Thriller</td>
<td>1.652</td>
<td>Fantasy</td>
<td>1.591</td>
<td>Sci-Fi</td>
<td>1.264</td>
</tr>
<tr>
<td></td>
<td>(1.556)</td>
<td></td>
<td>(1.342)</td>
<td></td>
<td>(0.731)</td>
</tr>
<tr>
<td>Musical</td>
<td>1.632</td>
<td>Mystery</td>
<td>1.576</td>
<td>Action</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(1.591)</td>
<td></td>
<td>(1.158)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Brunero Liseo (Sapienza, Roma)  
ABC for Wallenius  
18th February 2019  
29 / 59
Movies Dataset III

Movies - Posterior distributions

Brunero Liseo (Sapienza, Roma)
We have asked to Italian Academics in Statistics (Sector SECS-S/01) their preferences on scientific journals grouped in the 2015 “Statistics and Probability” list of (124) journals of the Institute for Scientific Information (ISI).

Note that

- Academics in Probability and Mathematical Statistics, Medical, Economic and Social Statistics are not included.
- Survey considers only faculty people with both teaching and research contracts.
- preferences (not the rank...), **between a minimum of ten and a maximum of twenty**
- Survey was conducted between the 25th of October 2016 and the 4th of November 2016.
We got 174 (out of 422) responses, distributed, according their role:

- 49 Full professors (Professori Ordinari)
- 72 Associate Professors (Professori Associati)
- 53 Assistant Professors (RTI, RTDa, RTDb)

We have then grouped journals by category (cluster analysis? )

- Methodology (45)
- Probability (23)
- Applied Statistics (34)
- Computational Statistics (9)
- Econometrics and Finance (13)
<table>
<thead>
<tr>
<th>Methodology</th>
<th>Probability</th>
<th>Applied</th>
<th>Computational</th>
<th>Econometrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>ω</td>
<td>0.335</td>
<td>0.070</td>
<td>0.228</td>
<td>0.244</td>
</tr>
<tr>
<td>ε = 0.130</td>
<td>(0.070)</td>
<td>(0.047)</td>
<td>(0.065)</td>
<td>(0.130)</td>
</tr>
<tr>
<td>ω</td>
<td>0.315</td>
<td>0.051</td>
<td>0.213</td>
<td>0.320</td>
</tr>
<tr>
<td>ε = 0.085</td>
<td>(0.044)</td>
<td>(0.031)</td>
<td>(0.042)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>ω</td>
<td>0.310</td>
<td>0.048</td>
<td>0.207</td>
<td>0.339</td>
</tr>
<tr>
<td>ε = 0.070</td>
<td>(0.037)</td>
<td>(0.027)</td>
<td>(0.033)</td>
<td>(0.073)</td>
</tr>
</tbody>
</table>
Figure: Approximations of the posterior distributions of the weights $\omega$ for each category included in the Journals dataset. Solid lines represent the approximations for $\epsilon = 0.13$, dashed lines for $\epsilon = 0.85$, and dotted lines for $\epsilon = 0.07$. 
Table: 1. Each entry $p_{ij}$ represents the ABC approximation of $\Pr(\omega_i > \omega_j)$.

<table>
<thead>
<tr>
<th></th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
<th>$\omega_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>-</td>
<td>1.000</td>
<td>0.999</td>
<td>0.394</td>
<td>1.000</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>-</td>
<td>-</td>
<td>0.000</td>
<td>0.000</td>
<td>0.226</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.104</td>
<td>0.951</td>
</tr>
<tr>
<td>$\omega_4$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.992</td>
</tr>
</tbody>
</table>
Violin Plots for the journals categories preferences

Brunero Liseo (Sapienza, Roma)
Conclusions

- We have proposed a novel model for preference data based on the Wallenius distribution.
- So far, the Wallenius model has been definitely under-employed, due to the intractability of the probability mass function.
- ABC represents a fast and reliable approach to estimate $\omega$
- Ease of implementation

Further Research:

- Nested Models (ranking within categories)
- Nonparametric extensions (how?)
- Application to other areas of interest

It is available at [https://arxiv.org/abs/1701.08142](https://arxiv.org/abs/1701.08142)

Email addresses: brunero.liseo@uniroma1.it

Website: [https://sites.google.com/a/uniroma1.it/bruneroliseo/](https://sites.google.com/a/uniroma1.it/bruneroliseo/)
References I


REFERENCES II


Thurstone Model: some details

An example: ranking 5 objects

\[ \begin{align*}
B1 & \rightarrow 5.6 \\
B2 & \rightarrow 4.9 \\
B3 & \rightarrow 9.3 \\
B4 & \rightarrow 2.1 \\
B5 & \rightarrow 3.4
\end{align*} \]
An example: ranking 5 objects

\[ B_1 \rightarrow 5.6 \]
\[ B_2 \rightarrow 4.9 \]
\[ B_3 \rightarrow 9.3 (\cdot, \cdot, \cdot, \cdot) \]
\[ B_4 \rightarrow 2.1 (\cdot, \cdot, \cdot, \cdot) \]
\[ B_5 \rightarrow 3.4 \]
An example: ranking 5 objects

\[
\begin{align*}
B_1 & \rightarrow 5.6 \\
B_2 & \rightarrow 4.9 \\
B_3 & \rightarrow 9.3 \\
B_4 & \rightarrow 2.1 \\
B_5 & \rightarrow 3.4
\end{align*}
\]
An example: ranking 5 objects

\[
\begin{align*}
B1 & \rightarrow 5.6 \\
B2 & \rightarrow 4.9 \\
B3 & \rightarrow 9.3 \\
B4 & \rightarrow 2.1 \\
B5 & \rightarrow 3.4 \\
\end{align*}
\]
### An example: ranking 5 objects

<table>
<thead>
<tr>
<th>Object</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_1$</td>
<td>5.6</td>
</tr>
<tr>
<td>$B_2$</td>
<td>4.9</td>
</tr>
<tr>
<td>$B_3$</td>
<td>9.3</td>
</tr>
<tr>
<td>$B_4$</td>
<td>2.1</td>
</tr>
<tr>
<td>$B_5$</td>
<td>3.4</td>
</tr>
</tbody>
</table>
An example: ranking 5 objects

\[
\begin{align*}
B1 & \rightarrow 5.6 \sim X_1 \\
B2 & \rightarrow 4.9 \sim X_2 \\
B3 & \rightarrow 9.3 \sim X_3 \quad (4,3,5,1,2) \\
B4 & \rightarrow 2.1 \sim X_4 \\
B5 & \rightarrow 3.4 \sim X_5
\end{align*}
\]
Thurstone Model: some details

**Basics**

- Each item is associated with a ‘true‘ real value: loudness of a sound, sweetness of a cake, etc.
- A judge assesses the cakes or the sounds and classify them
- Measurement error due to lack of precision of the sensorial apparatus
- The output is a ranking, that is a permutation $\sigma$ of the first $n$ integer
Thurstone Model: some details

### The model

Thurstone (1927) proposed a ranking process where the ranking $\pi_j$ of $n$ objects given by a random sample of independent judges ($j = 1, \ldots, n$) is determined by the relative ordering of $n$ random utilities $X_{1j}, X_{2j}, \ldots, X_{nj}$.

The probability of observing a given ranking is

$$P(\pi) = P\left(X_{\sigma^{-1}(1)} < X_{\sigma^{-1}(2)} < \cdots < X_{\sigma^{-1}(n)}\right)$$

Typically, some specific parametric model is assumed for the $X_i$'s.
Thurstone Model: some details

Learning

\[ p(\sigma) = PP \left( X_{\sigma^{-1}(1)} < X_{\sigma^{-1}(2)} < \cdots < X_{\sigma^{-1}(n)} \right) = \int_{\Omega} F(x_1, x_2, \ldots, x_n) \, dx_1 \cdots dx_n \]

where

\[ \Omega = \left\{ (x_1, \ldots, x_n) \mid x_{\sigma^{-1}(1)} < \cdots < x_{\sigma^{-1}(n)}; x_i \in \mathbb{R} \right\} \]

Limitations: It works well with few items (Gibbs sampling efficient with \( n \leq 10 \) )
An example: ranking 5 objects
What are permutations?

A permutation $\sigma$ is a bijection from the set $\{1, 2, \ldots, n\}$ onto itself:

$$\sigma : \{1, 2, \ldots, n\} \rightarrow \{1, 2, \ldots, n\}$$

- $\sigma(i)$ is the rank associated to the $i$-th element of a list
- $\pi = \sigma^{-1}(i)$ is the $i$-th ranked item

$S_n$ is the set of permutations of $n$ elements; $(S_n, \cdot)$ forms the symmetric group, i.e.

$$\sigma_1 \cdot \sigma_2(i) = \sigma_1(\sigma_2(i))$$
Permutations

Notation

- \( e \) denotes the **identity** permutation
- each permutation can be represented by a **permutation matrix** \( M \) with

\[
M_{ij} = \begin{cases} 
1 & \sigma(i) = j \\
0 & \text{otherwise}
\end{cases}
\]
A problem

- How many parameters are needed to evaluate

\[ p(1, 2, 3, \ldots, n), p(3, n, \ldots, 1), p(n - 1, 2, n, \ldots, 4), \ldots \text{and so on?} \]

\[ n! - 1 \]

\ldots Too many!
We really need a statistical model!
Hypergeometric distribution

Imagine a urn with marbles of two colors, red ones and green ones.
Hypergeometric distribution

Imagine a urn with marbles of two colors, red ones and green ones.

Let $N$ be the number of marbles in the urn and $m$ be the number of red marbles, $(N - m)$ is the number of green marbles. Draw $n$ balls from the urn without replacement and let $X$ be the random variable which describes the number of red marbles actually drawn.

Then

$$P(X = k) = \binom{m}{k} \binom{N - m}{n - k} / \binom{N}{n},$$

with some constraints on the support of $X$. 
Multivariate Hypergeometric distribution

The previous model can be easily extended to more than two colors
Multivariate Hypergeometric distribution

The model of an urn with red and green marbles can be extended to the case where there are more than two colors of marbles.

Assume there are $m_i$ marbles of color $i$, $i = 1, \ldots, c$, and you draw $n$ marbles without replacement. The result of the experiment is now a vector $(x_1, x_2, \ldots, x_c)$ and its distribution is the Multivariate Hypergeometric one.

$$P(X_1 = k_1, X_2 = k_2, \ldots, X_c = k_c) = \frac{\binom{m_1}{k_1} \binom{m_2}{k_2} \cdots \binom{m_c}{k_c}}{\binom{N}{n}}.$$
Now assume that different colors have different importance or size and, consequently, a different probability to be selected.

Here we need the Wallenius distribution!
Why does it work

Trivial proof, based on *Accept-Reject* ideas.
Let $f$ be the density of the selected values of $\theta$.

\[
f(\theta_i) \propto \sum_{z \in X} \pi(\theta_i) f(z | \theta_i) I_y(z)
\]

\[
\propto \pi(\theta_i) f(y | \theta_i)
\]

\[
\propto \pi(\theta_i | y)
\]