Mallows ranking models: maximum likelihood estimate and regeneration

Wenpin Tang
Department of Mathematics, UCLA

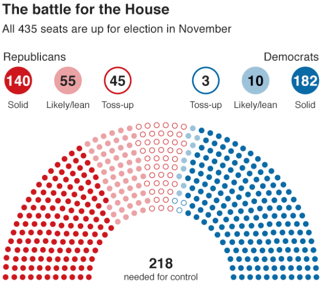
June 14, 2019

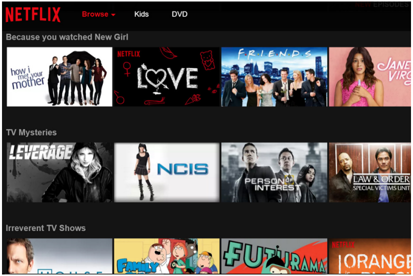
Background

Ranked data appear in many problems of *social choice*, *user recommendation*, *information retrieval*...

Examples :

- ranking candidates by voters in political elections ;
- preference list of competing items collected from consumers ;
- document retrieval by aggregating a ranked list of webpages output by various search algorithms.





Mathematical models

Ranking = Permutation. Given n items, a ranking $\pi \in \mathfrak{S}_n$ is described by

- *word list* : $(\pi(1), \pi(2), \ldots, \pi(n))$,
- *ranked list* : $(\pi^{-1}(1)|\pi^{-1}(2)|\ldots|\pi^{-1}(n))$.

$\pi(i) = j$: the item i has rank j, and $\pi^{-1}(j) = i$: the j^{th} most preferred is item i.

Mallows model :

$$\mathbb{P}_{\theta,\pi_0,d}(\pi) \propto e^{-\theta d(\pi,\pi_0)} \quad \text{for } \pi \in \mathfrak{S}_n,$$

- $\theta > 0$ is the *dispersion parameter*,
- π_0 is the *central ranking*,
- $d(\cdot,\cdot)$ is a discrepancy function which is right invariant :

$$d(\pi,\sigma) = d(\pi \circ \sigma^{-1}, id) \quad \text{for } \pi, \sigma \in \mathfrak{S}_n.$$

Mallows model

Diaconis' list of $d(\cdot,\cdot)$:

- *Mallows' θ model* : $d(\pi,\sigma) = \sum_{i=1}^{n}(\pi(i) - \sigma(i))^2$ is the *Spearman's rho*,
- *Mallows' ϕ model* : $d(\pi,\sigma) = \text{inv}(\pi \circ \sigma^{-1})$ is the *Kendall's tau*...

Mallows' ϕ model is more interesting, since it is an instance of two large models, Fligner and Verducci ('86, '88) :

- *distance-based ranking models*,
- *multistage ranking models*.

Correctness measure : *inversion table* $(s_j(\pi))_{1 \le j \le n-1}$

$$s_j(\pi) := \pi^{-1}(j) - 1 - \sum_{j' < j} 1_{\{\pi^{-1}(j') < \pi^{-1}(j)\}}.$$

$$\mathbb{P}_{\pi_0,\theta} \propto \prod_{j=1}^{n} \exp(-\theta\, s_j(\pi \circ \pi_0^{-1})).$$

MLE $\widehat{\theta}$, $\widehat{\pi}_0$

- MLE $\widehat{\theta}$: easy by convex optimization.
- MLE $\widehat{\pi}_0$: *Kemeny's consensus ranking problem*

$$\widehat{\pi}_0 := \text{argmin}_{\pi_0} \sum_{i=1}^{N} \text{inv}(\pi_i \circ \pi_0^{-1}).$$

This problem is **NP-hard**, with a few heuristic algorithms.

Theoretical properties of $\widehat{\theta}$, $\widehat{\pi}_0$:

1. Are the MLEs $\widehat{\theta}$, $\widehat{\pi}_0$ consistent ?
2. Is the MLE $\widehat{\theta}$ unbiased ?
3. How fast do MLEs $\widehat{\pi}_0$ converge to π_0 ?

Not well studied, only Mukherjee ('16) considered $\widehat{\theta}$.

Properties of MLE

Theorem

Let $\widehat{\theta}$, $\widehat{\pi}_0$ be the MLE of θ, π_0 with N samples.

1. $$\mathbb{E}_{\theta,\pi_0}\widehat{\theta} > \theta.$$

2. $$\sqrt{\frac{2}{\pi N}}\left(\cosh\frac{\theta}{2}\right)^{-N} \leq \mathbb{P}_{\theta,\pi_0}(\widehat{\pi}_0 \neq \pi_0) \leq (n - H_n)n!\left(\cosh\frac{\theta}{2}\right)^{-N}.$$

Hint : For $\pi \sim$ Mallows' ϕ, inv(π) is decomposed as independent *truncated geometric variables*. Then apply LDP bounds.

Infinite Mallows models

Motivation : Tackle the problem of ranking a large number items $\longrightarrow$ infinite ranking/permutation models.

$$\mathbb{P}_{\theta,\pi_0}(\pi) \propto \exp\left(-\theta \sum_{j=1}^{t} s_j(\pi \circ \pi_0^{-1})\right),$$

regarded as a t-marginal of random permutation of $\mathbb{N}_+$.

Theory : Pitman and Tang *Regenerative random permutations of integers*, AoP ('19)

$\longrightarrow$ Infinite Mallows model enjoys the regenerative property : it is a concatenation of i.i.d. indecomposable blocks

$$(\underbrace{2, 3, 4, 1}_{L_1=4}, \underbrace{6, 8, 7, 10, 5, 9}_{L_2=6}, \underbrace{12, 13, 11}_{L_3=3}, \ldots)$$

How to choose the model size t ?

Fact : $\mathbb{E}L = \frac{1}{(e^{-\theta};e^{-\theta})_\infty}$.

```
Algorithm   't' selection algorithm
  procedure t_SEL(T)
    Err ← ∞, t_SEL ← 0                                  ▷ Initialization
    for t in T do
      θ ← MLE(t)                                        ▷ Run MLE heuristic algorithm
      if |t − 1/(e^{−θ}; e^{−θ})_∞| < Err then
        Err ← |t − 1/(e^{−θ}; e^{−θ})_∞|
        t_SEL ← t
      end if
    end for
    return t_SEL                                        ▷ The selected 't' is t_SEL
  end procedure
```

With ‘t’ selected, we fit a *Generalized Mallows model* :

$$\mathbb{P}_{\vec{\theta},\pi_0}(\pi) \propto \exp\left(-\sum_{j=1}^{t} \theta_j\, s_j(\pi \circ \pi_0^{-1})\right).$$

Synthetic data

TABLE: Accuracy of estimated rank & average training time for 50 simulated data with $t_{max} = 10$ (resp. $t_{max} = 20$, $t_{max} = 40$) and $\vec{\theta} = (1, 0.975, \ldots, 0.775, 0, \ldots)$ (resp. $\vec{\theta} = (1, 0.975, \ldots, 0.525, 0, \ldots)$, $\vec{\theta} = (1, 0.975, \ldots, 0.025, 0, \ldots)$) by the IGM model of model size $t = 1$, $t = 10$ and Algorithm.

$t_{max} = 10$	IGM($t = 1$)	IGM($t = 10$)	ALGO
ACC. EST. RANK	**100%**	**100%**	**100%**
AVE. TIME	**1.56 s**	14.45 s	2.80 s
$t_{max} = 20$	IGM($t = 1$)	IGM($t = 10$)	ALGO
ACC. EST. RANK	94%	**100%**	**100%**
AVE. TIME	**5.73 s**	54.45 s	24.42 s
$t_{max} = 40$	IGM($t = 1$)	IGM($t = 10$)	ALGO
ACC. EST. RANK	82%	**100%**	**100%**
AVE. TIME	**70.26 s**	684.65 s	391.20 s

The algorithm is also applied to other data as APA data, university's homepage search data...

Thank you for your attention !