



Sequential rank agreement methods for comparison of ranked lists

Claus Thorn Ekstrøm
Biostatistics, University of Copenhagen
ekstrom@sund.ku.dk



Motivation — Colon cancer studies

Rank	Denmark	Australia	Japan
1	228030_at	228030_at	228030_at
2	228915_at	230793_at	236223_s_at
3	243669_s_at	236223_s_at	230921_s_at
4	213385_at	230921_s_at	1559391_s_at
5	230964_at	230621_at	232595_at
6	207607_at	216992_s_at	242700_at
7	1556055_at	207463_x_at	1556055_at
8	243808_at	203008_x_at	242110_at
9	216173_at	231829_at	234207_at
10	230621_at	225802_at	206239_s_at

How many genes to include in subsequent studies?



What we want ...

Question

Can we identify/evaluate an optimal rank until which the lists agree satisfactorily on the items?

Requirements:

- Need a measure of agreement
- Interpretable
- Work on multiple list
- Work on censored/partial ranked lists (handle $n \ll p$ problems)
- Emphasis on top of list



Notation

- L (partially) ranked lists of P items X_1, \dots, X_P .
- $R_l(X_i)$ is rank assigned to item X_i in list l

Rank	List 1	List 2	List 3
1	A	A	B
2	B	C	A
3	C	D	E
4	D	B	C
5	E	E	D

Item	R_1	R_2	R_3
A	1	1	2
B	2	4	1
C	3	2	4
D	4	3	5
E	5	5	3



Agreement

Limits-of-agreement of ranks

Agreement for item X_p is

$$A(X_p) = \sqrt{\frac{\sum_{i=1}^L (R_i(X_p) - \bar{R}(X_p))^2}{L-1}}$$

Sequential rank agreement (pooled SD of items in S_d)

$$\text{sra}(d) = \sqrt{\frac{\sum_{\{p \in S_d\}} (L-1)A(X_p)^2}{(L-1)|S_d|}}$$

Items to consider at depth d

$$S_d = \{R_i^{-1}(r); r \leq d\},$$

Depth	S_d
1	{A, B}
2	{A, B, C}
3	{A, B, C, D, E}
4	{A, B, C, D, E}
5	{A, B, C, D, E}



Golub data

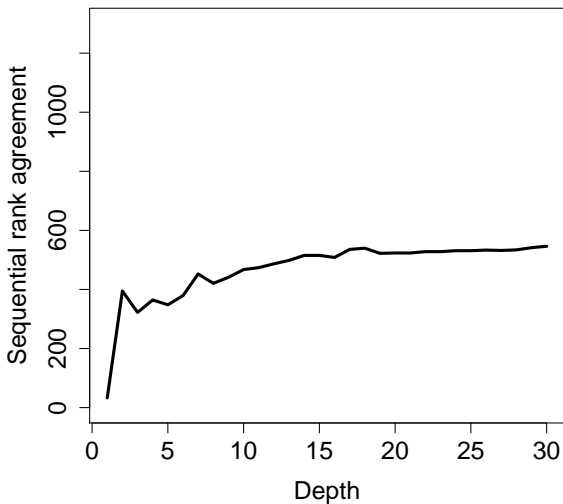
- Classification between leukemia (ALL and AML)
- 3051 gene expression values measured on 38 tumor mRNA samples
- Four methods

Rank	T	logReg	eNet	MIC
1	2124	2124	829	378
2	896	896	2124	829
3	2600	829	2198	896
4	766	394	808	1037
5	829	766	1665	2124
6	2851	2670	1920	808
7	703	2939	1042	108
8	2386	2386	1389	515
9	2645	1834	937	2670
10	2002	378	1767	2600



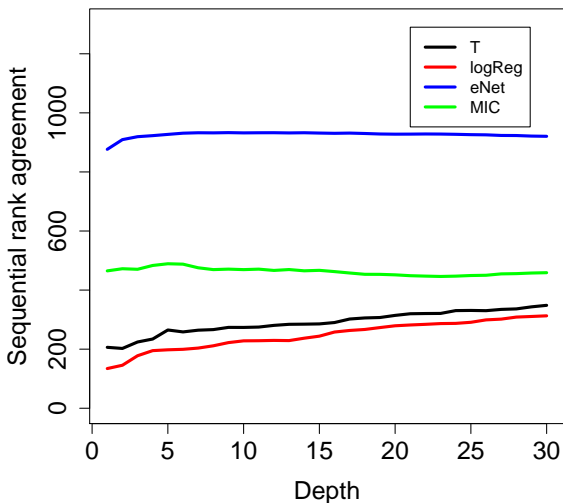
Sequential rank agreement

Predictor agreement



Stability of selections

100 bootstrap samples. Compare predictor ranking for each method.



Evaluating the sra curve

Reference band for the sequential rank agreement

H_0 : The list rankings correspond to complete randomly permuted lists

\tilde{H}_0 : The list rankings are based on data containing no association to the outcome.



Evaluating the sra curve

Reference band for the sequential rank agreement

H_0 : The list rankings correspond to complete randomly permuted lists

\tilde{H}_0 : The list rankings are based on data containing no association to the outcome.

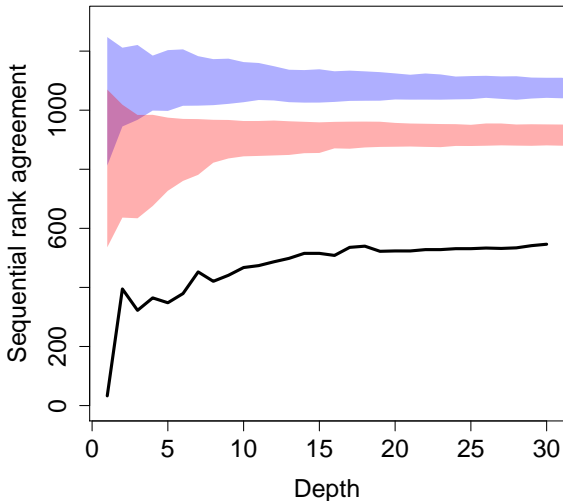
Randomize lists

- Produce completely random lists (H_0)
- Randomize outcomes and compute rankings for same methods (\tilde{H}_0)

Several times — compute pointwise 95% reference bands



Evaluating sequential rank agreement



Partially ranked lists

Partially ranked lists are common:

- Top k lists
- Methods: lasso
- Relevance: significance

Handling partially ranked lists

Impute missing ranks at random for each list B times

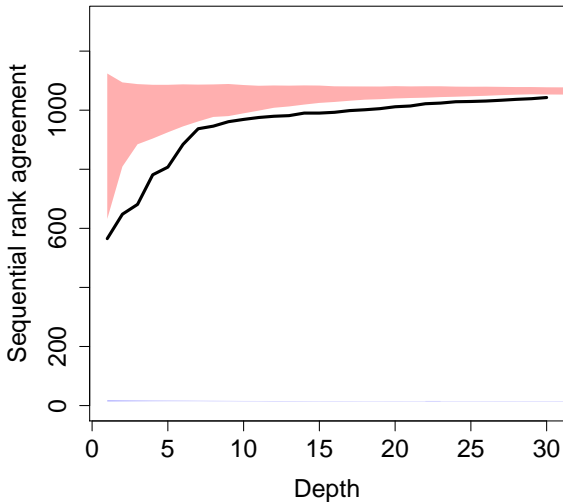
- 1 Compute sra for each fully observed list
- 2 Average over the sequential rank agreement obtained

Note: Assumes censored data are irrelevant.

Note: Cannot just apply mean rank of missing items



Evaluating sra — top 50



Theoretical results

Theorem

Assume that $\{R_l(X)\}_{l=1}^L$ are independent draws from a probability distribution Q on the set of lists Π . Then

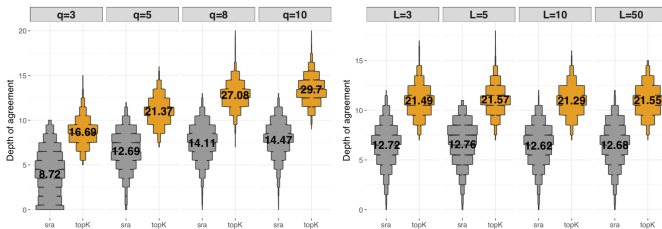
$$\|\widehat{\text{sra}}_L - \text{sra}\| = o_P(1)$$

Corollary

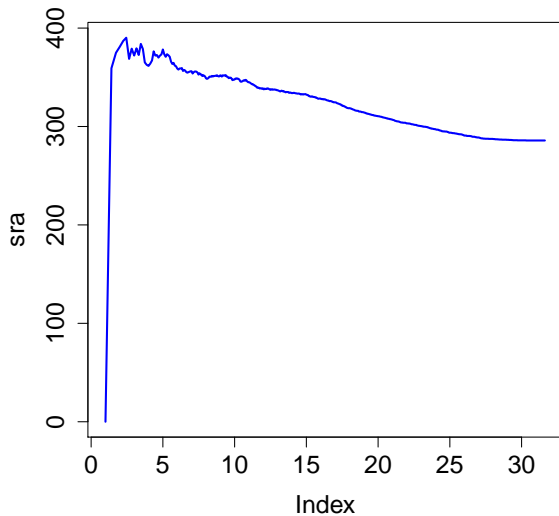
Let \widehat{q}_L be a positive threshold function such that $\|\widehat{q}_L - q\|_\infty = o_P(1)$ for some limiting function q . Then,

$$\widehat{d}_L^*(\widehat{q}_L) \xrightarrow{P} d^*(q) \text{ for } L \rightarrow \infty.$$


Comparing to other methods



Revisiting the colon data



Summary and future ideas

Sequential rank agreement

- Interpretable measure
- Changepoint identification / prior limit
- Versatile
 - Compare ranking from across different samples
 - Compare predictor ranking of methods applied to same data
 - Compare risk predictions across different methods
 - Stability of rankings via bootstrap

Current extensions:

- Cluster methods based on sequential rank agreement
- Use sra as criterion in cross-validation

