

# Sequential rank agreement methods for comparison of ranked lists

February 25, 2015

## Abstract

## 1 Introduction

Fra Kaspers PhD. Ikke rettet til endnu:

Comparison of ranked lists, though not a recently emerged subject has been receiving increased attention with the use of internet search engines, but lists of this kind are found in many areas. The advances in biotechnology have renewed this interest since many genetic studies are producing vast amounts of data that need to be sorted and evaluated. For example, studies of gene expression often rank genes according to their difference in expression across samples, while search engine results are presented with the best match first. Common for these types of lists is that it is up to the investigator to decide which  $k$  is relevant for a given problem when comparing top  $k$  lists. Typically, the items in the lists will be more similar at the top of the lists, since strong effects ensure that an item are more likely to be identified rather than further down where items consisting purely or mostly of noise are arranged more randomly.

## 2 Methods

Let there be given  $L$  lists,  $L \geq 2$  of length  $P$  that each rank the same set of  $P$  items (variables) such that the first element of a list  $l$  is the item with rank 1, the second element of  $l$  is the item with rank 2 etc. Without any loss of generality we can assume that the predictors are numbered from 1 to  $P$  so that an ordered list,  $l$ , essentially is a permutation of the numbers from 1 to

a)								c)	
List				List				Depth	
Rank	1	2	3	Item	1	2	3	$d$	$S_d$
1	A	A	B	A	1	1	2	1	{A, B}
2	B	C	A	B	2	4	1	2	{A, B, C}
3	C	D	E	C	3	2	4	3	{A, B, C, D, E}
4	D	B	C	D	4	3	5	4	{A, B, C, D, E}
5	E	E	D	E	5	5	3	5	{A, B, C, D, E}

Figure 1: Example set of ranked lists. a) presents the ranked list of predictors/items for each of three lists while b) presents the ranks obtained by each predictor/item in each of the three lists.

$P$ . This gives rise to a bijective permutation function,  $\pi_l(r) \in \{1, \dots, P\}$  which returns the predictor that was found at rank  $r$  in list  $l$ . Conversely, we can define the rank function

$$R_l(p) = \pi_l^{-1}(p) \quad (1)$$

which returns the rank given to item  $p$  in list  $l$  (see Figure 1 for a schematic example of these two functions).

For a rank  $d$  we define the set of items that have been ranked less than or equal to  $d$  in any of the  $L$  lists as

$$S_d = \{\pi_l(r); r \leq d, l = 1, \dots, L\}. \quad (2)$$

$S_d$  contains the set of items found in the  $L$  top  $d$  lists. Finally, we define  $A(p)$  as the measure of agreement of the rankings given to item  $p$  in the  $L$  lists,

$$A(p) = f(R_1(p), \dots, R_L(p)), \quad (3)$$

for a given pre-specified function  $f$ . Throughout this paper we will use the sample standard error as our function  $f$ ,

$$f(R_1(p), \dots, R_L(p)) = \sqrt{\frac{\sum_{i=1}^L (R_i(p) - \bar{R}(p))^2}{L-1}},$$

but other choices could be made (see the discussion). The sample standard error has a nice interpretation as the average distance the individual rankings for item  $p$  is from the average ranking of item  $p$ . Thus the investigator can base his or her assessment ... on a ranking

... define the *sequential rank agreement* as the pooled standard deviation of the items found in the set  $S_d$ :

$$\text{SRA}(d) = \sqrt{\frac{\sum_{\{p \in S_d\}} (L-1)A(p)^2}{(L-1) \cdot |S_d|}}. \quad (4)$$

## 2.1 All lists fully observed

We shall start by the simplest case where all  $L$  lists are fully observed.

## 2.2 Analysis of top $k$ lists

Not uncommon for lists to be censored

Let  $\Lambda_l, l = 1, \dots, L$  be the set of items found in list  $l$  so  $\Lambda_l$  is the top  $k_l$  list of items from list  $l$  where  $k_l = |\Lambda_l|$ . Note that we observe the top  $k$  items for each of the  $L$  lists if  $k_1 = \dots = k_L = k$ . For censored lists the rank function becomes

$$R_l(p) \text{ is } \begin{cases} \pi_l^{-1}(p) & \text{for } p \in \Lambda_l \\ > k_l & \text{for } p \notin \Lambda_l \end{cases} \quad (5)$$

where we only know that the rank for the unobserved items in list  $l$  must be larger than the largest rank observed in that list.

In the case of censored lists it only makes sense to look at depths where we have corresponding observations so the largest rank we should consider will be

$$d \leq \max(k_1, \dots, k_L). \quad (6)$$

We cannot directly compute  $A(p)$  for all predictors because we only observe a censored version of  $R$  for some of the lists. Instead we assume that the rank assigned to predictor  $p$  in list  $l$  is uniformly distributed among the ranks that have been unassigned for list  $l$ . The ranks are clearly not independent since each of the lists essentially contains full set of ranks

$$A(p) = \frac{\sum_{r_1; r_1 \in R_1(p)} \dots \sum_{r_L; r_L \in R_L(p)} f(r_1, \dots, r_L)}{\prod_l |R_l(p)|} \quad (7)$$

Notationen er ikke helt skarp da  $R$  jo returnerer delvist et tal og delvist en mngde, men det giver nppe anledning til misforstelser

## **3 Applications**

### **3.1 Evaluating results from top- $k$ lists**

### **3.2 Comparing results across different method**

### **3.3 Stability of results**

Bootstrap across a single method and compare results. Discuss collinearity

## **4 Discussion**

Mention/discuss different measures.