

Confidence ellipses for holistic approaches

Marine Cadoret & François Husson

Agrostat 2012, Paris

29 february 2012

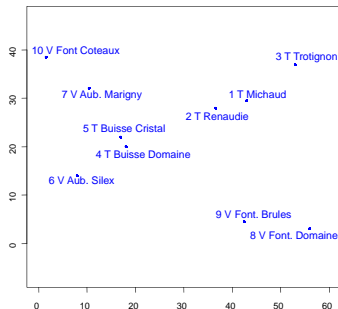


- 1 What do we mean by “holistic approaches” and “confidence ellipse”?
- 2 Confidence ellipses construction
- 3 Why partial bootstrap doesn't work?
- 4 Validity of total bootstrap
- 5 Conclusion and perspectives

- From ὅλος (holos), a Greek word meaning *all, entire, total*
- Products evaluated in their entirety
- Among holistic approaches:
 - Napping
 - Sorting
 - Sorted napping
 - Hierarchical sorting
 - Flash profile and Free choice profiling (between holistic and analytic approaches)

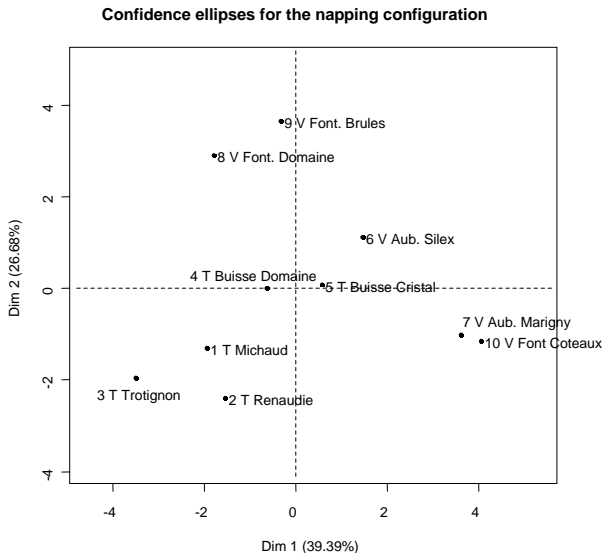
Napping data: an example with 10 wines and 11 judges

Judge 1

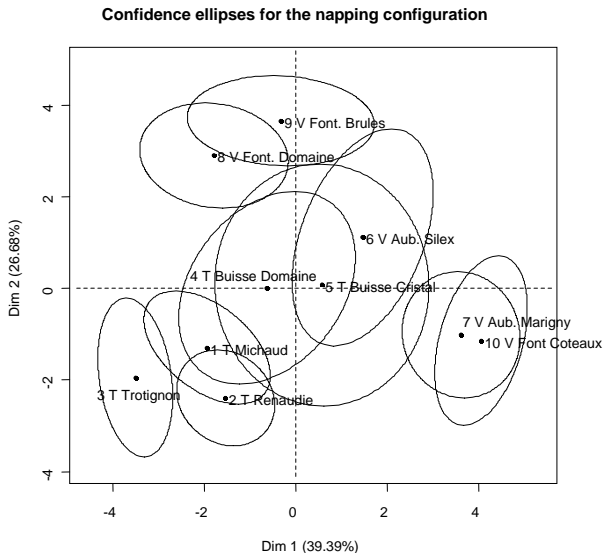


	X1	Y1	..	X11	Y11
1 T Michaud	43	29		48	15
2 T Renaudie	36	28		45	14
3 T Trotignon	53	37		8	23
4 T Buisse Domaine	18	20		31	9
5 T Buisse Cristal	17	22	..	34	31
6 V Aub. Sillex	8	14		20	35
7 V Aub. Marigny	10	32		47	28
8 V Font. Domaine	56	3		4	5
9 V Font. Brules	42	4		8	6
10 V Font Coteaux	1	38		54	36

Napping data: an example with 10 wines and 11 judges



Napping data: an example with 10 wines and 11 judges



- 1 What do we mean by “holistic approaches” and “confidence ellipse”?
- 2 Confidence ellipses construction
- 3 Why partial bootstrap doesn't work?
- 4 Validity of total bootstrap
- 5 Conclusion and perspectives

Bootstrap technique

Real jury

	J1	J2		
	X1,Y1	X2,Y2	...	X11,Y11
P1				
P2				
P3				
⋮				
P10				

Bootstrap technique

Real jury

	J1	J2		
	X1,Y1	X2,Y2	...	X11,Y11
P1	⋮	⋮		⋮
P2	⋮	⋮		⋮
P3	⋮	⋮		⋮
⋮	⋮	⋮		⋮
P10	⋮	⋮		⋮

Virtual jury

			...	
P1	⋮	⋮		⋮
P2	⋮	⋮		⋮
P3	⋮	⋮		⋮
⋮	⋮	⋮		⋮
P10	⋮	⋮		⋮

Bootstrap technique

Real jury

	J1	J2		
	X1,Y1	X2,Y2	...	X11,Y11
P1				
P2				
P3				
⋮				
P10				

Virtual jury

	J1			
	X1,Y1		...	
P1				
P2				
P3				
⋮				
P10				

Bootstrap technique

Real jury

	J1	J2		
	X1,Y1	X2,Y2	...	X11,Y11
P1				
P2				
P3				
⋮				
P10				

Virtual jury

	J1	J1		
	X1,Y1	X1,Y1	...	
P1				
P2				
P3				
⋮				
P10				

Bootstrap technique

Real jury

	J1	J2		
	X1;Y1	X2;Y2	...	X11;Y11
P1	⋮	⋮		⋮
P2	⋮	⋮		⋮
P3	⋮	⋮		⋮
⋮	⋮	⋮		⋮
P10	⋮	⋮		⋮

Virtual jury

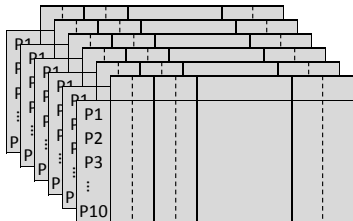
	J1	J1		J10
	X1;Y1	X1;Y1	...	X10;Y10
P1	⋮	⋮		⋮
P2	⋮	⋮		⋮
P3	⋮	⋮		⋮
⋮	⋮	⋮		⋮
P10	⋮	⋮		⋮

Bootstrap technique

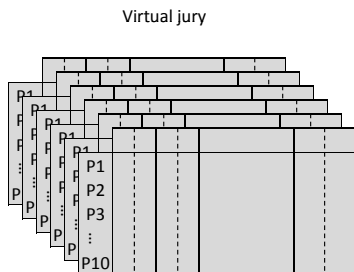
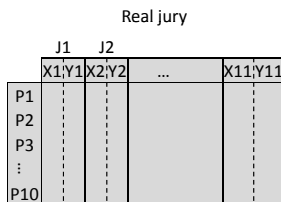
Real jury

	J1	J2		
	X1;Y1	X2;Y2	...	X11;Y11
P1				
P2				
P3				
⋮				
P10				

Virtual jury



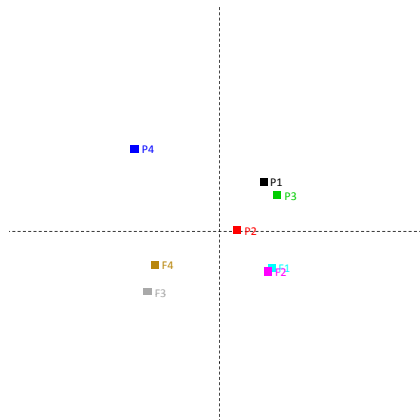
Bootstrap technique



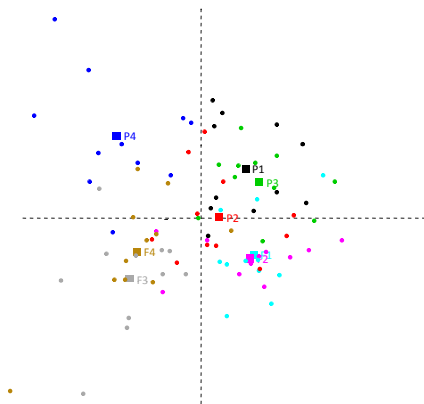
2 ways to use bootstrapped virtual juries:

- by projection (partial bootstrap)
- by procrustean rotation (total bootstrap)

- Multiple Factor Analysis (MFA)

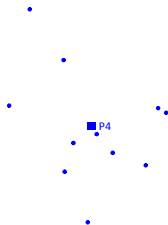


Partial bootstrap

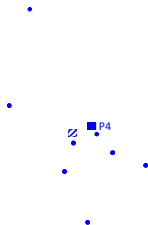


- Multiple Factor Analysis (MFA)
- Projection to get the products according to each judge (of the real jury): partial representation

Partial bootstrap

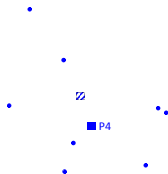


- Multiple Factor Analysis (MFA)
- Projection to get the products according to each judge (of the real jury): partial representation
⇒ barycentric property



- Multiple Factor Analysis (MFA)
- Projection to get the products according to each judge (of the real jury): partial representation \Rightarrow barycentric property
- Creation of virtual jury and calculation of new barycenter

Partial bootstrap

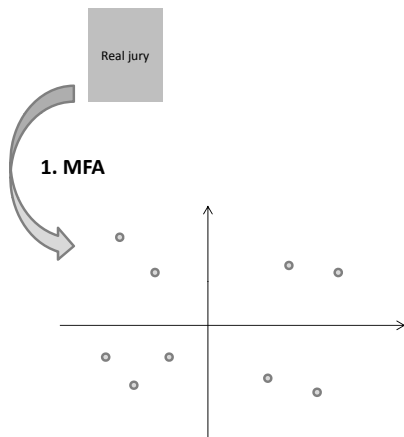


- Multiple Factor Analysis (MFA)
- Projection to get the products according to each judge (of the real jury): partial representation \Rightarrow barycentric property
- Creation of virtual jury and calculation of new barycenter

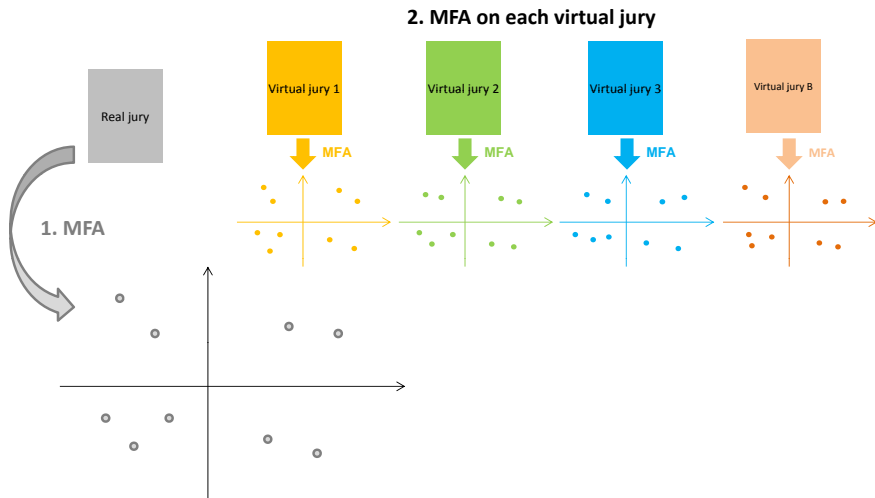


- Multiple Factor Analysis (MFA)
- Projection to get the products according to each judge (of the real jury): partial representation
⇒ barycentric property
- Creation of virtual jury and calculation of new barycenter
- Building confidence ellipses containing 95% of the points

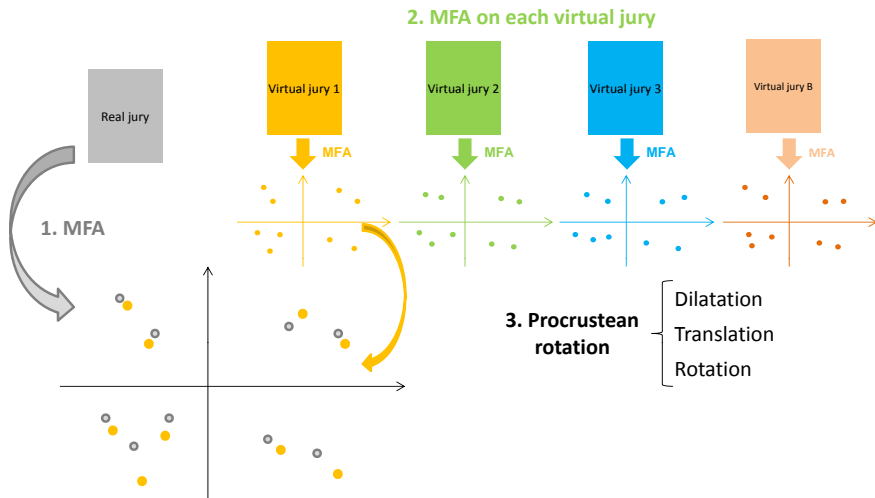
Total bootstrap



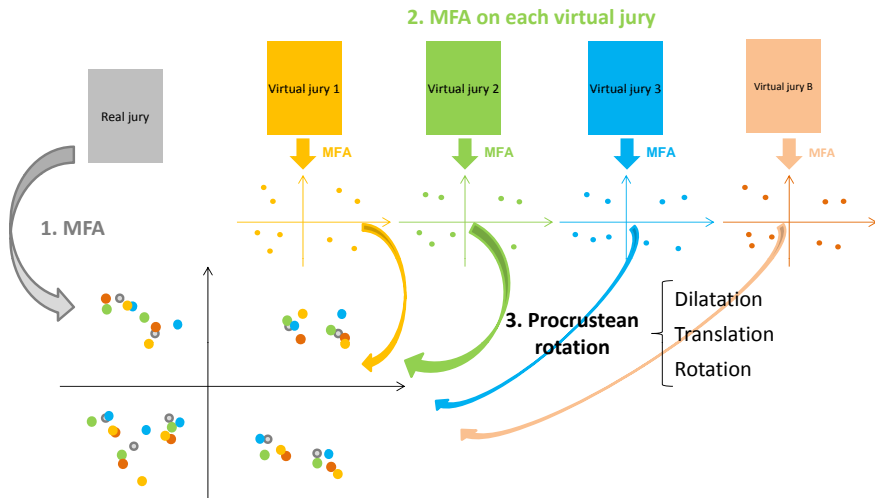
Total bootstrap



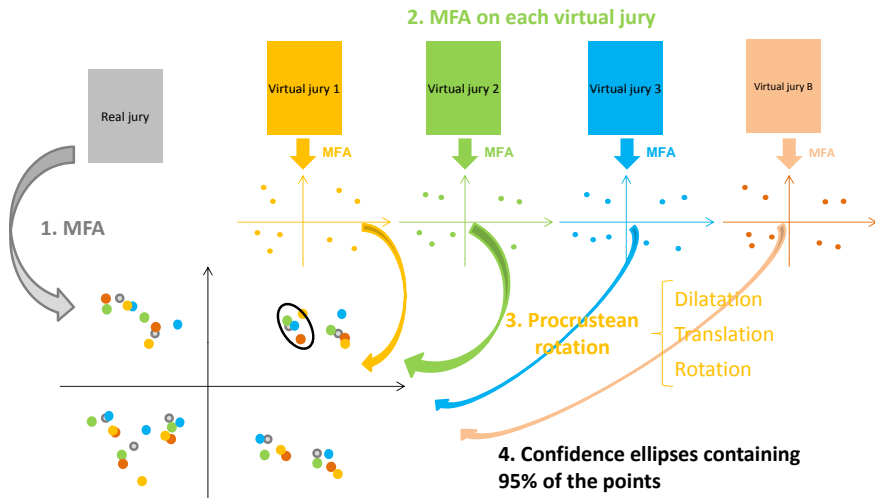
Total bootstrap



Total bootstrap

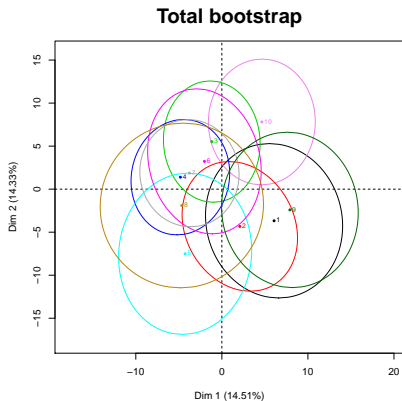
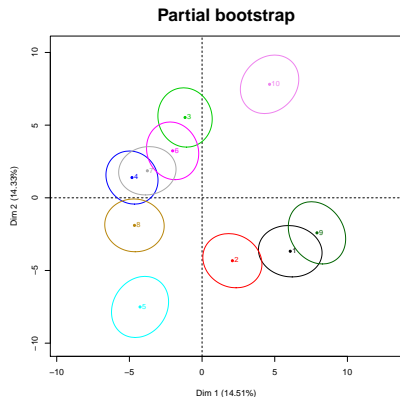


Total bootstrap



Comparison of partial and total bootstrap

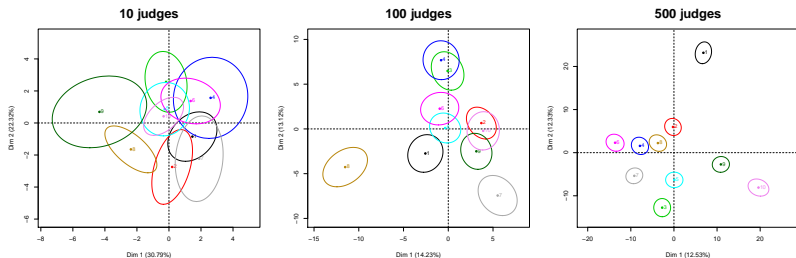
A completely random dataset with 100 judges



- 1 What do we mean by “holistic approaches” and “confidence ellipse”?
- 2 Confidence ellipses construction
- 3 Why partial bootstrap doesn't work?
- 4 Validity of total bootstrap
- 5 Conclusion and perspectives

Partial bootstrap: increased number of judges

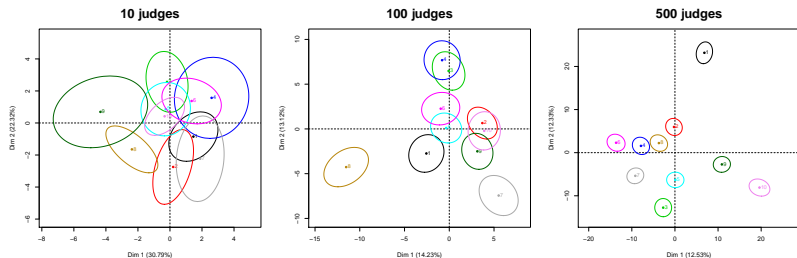
Completely random dataset



- Dimensionality problem (few products in a too large space)?
- Inference problem (barycenter calculated with too many points)?

Partial bootstrap: increased number of judges

Completely random dataset

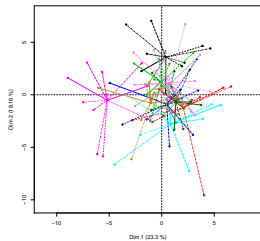


- Dimensionality problem (few products in a too large space)?
- Inference problem (barycenter calculated with too many points)?

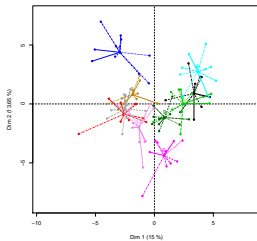
⇒ Modify the dimensionality of the dataset independently to the number of judges

Random datasets: fixed number of judges

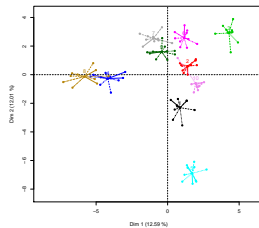
10 judges – 2 descriptors per judge



10 judges – 20 descriptors per judge

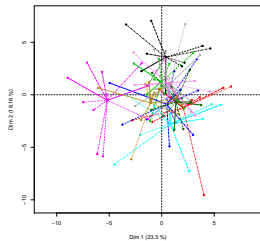


10 judges – 200 descriptors per judge

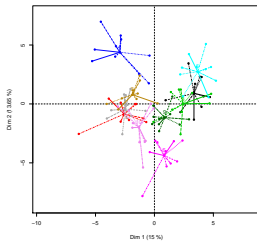


Random datasets: fixed number of judges

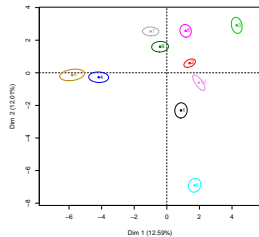
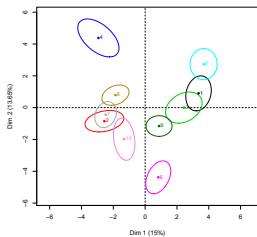
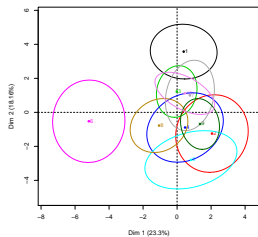
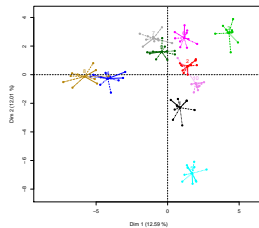
10 judges – 2 descriptors per judge



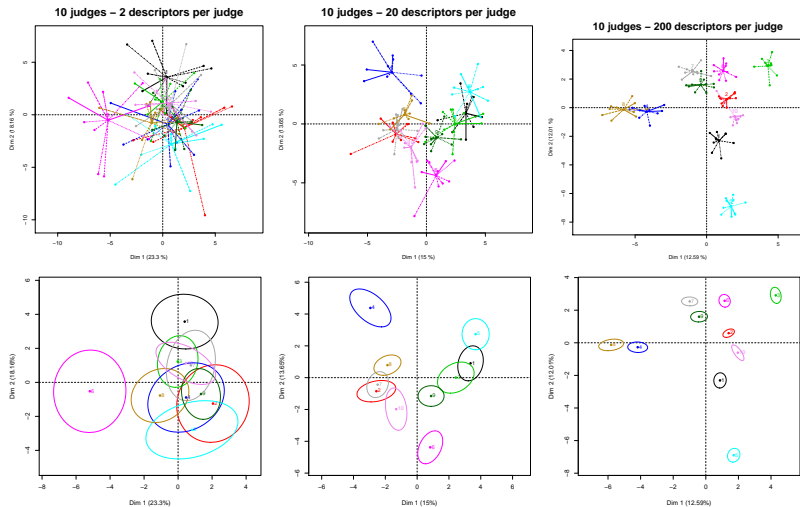
10 judges – 20 descriptors per judge



10 judges – 200 descriptors per judge

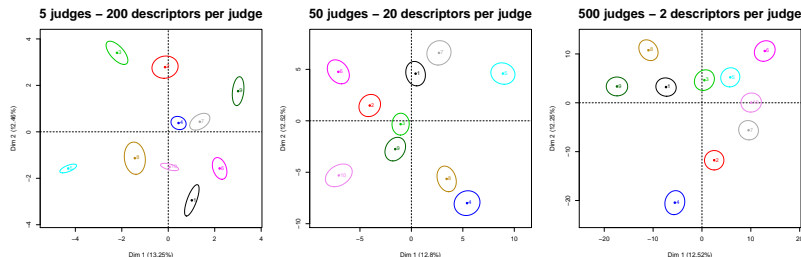


Random datasets: fixed number of judges



Products are better separated when the number of dimensions increases (same problem with GPA)

Random datasets: fixed size of dataset



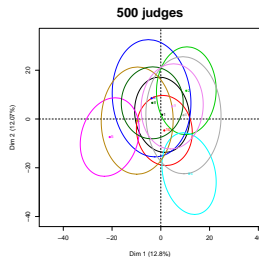
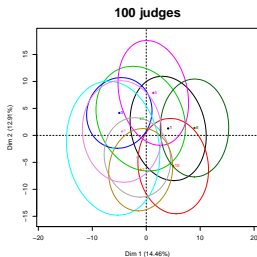
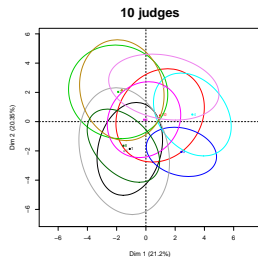
The sizes of the ellipses don't depend on the number of judges but only on the dimensionality of the dataset

Outline

- 1 What do we mean by “holistic approaches” and “confidence ellipse”?
- 2 Confidence ellipses construction
- 3 Why partial bootstrap doesn't work?
- 4 **Validity of total bootstrap**
- 5 Conclusion and perspectives

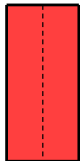
The case of completely random data

Dimensionality problem with completely random data?

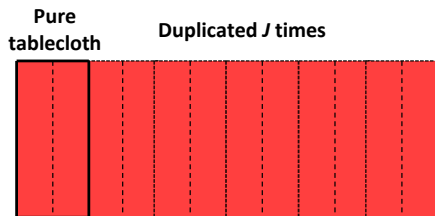


Data simulation procedure

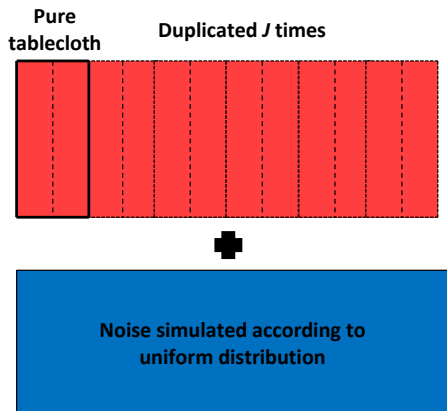
**Pure
tablecloth**



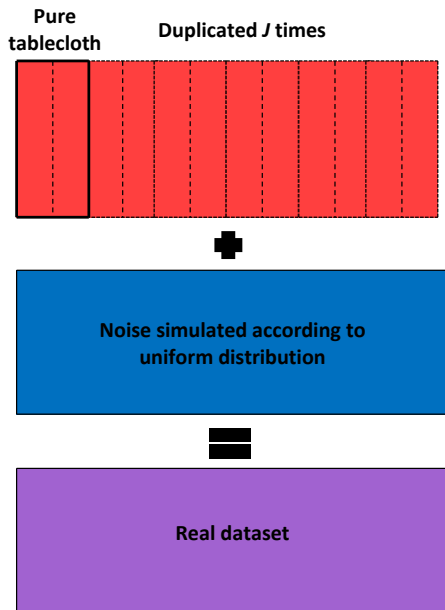
Data simulation procedure



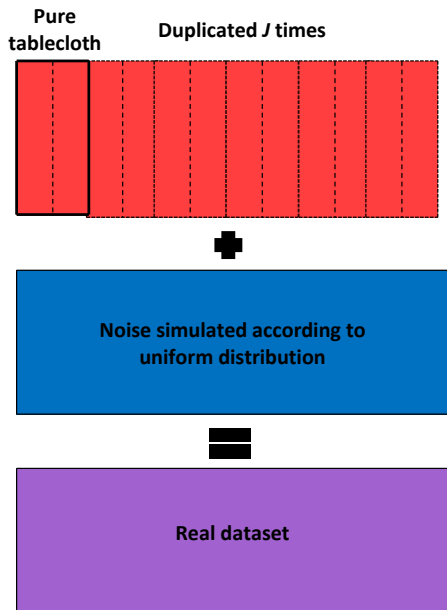
Data simulation procedure



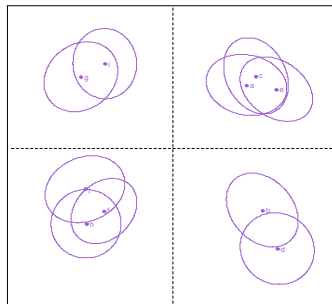
Data simulation procedure



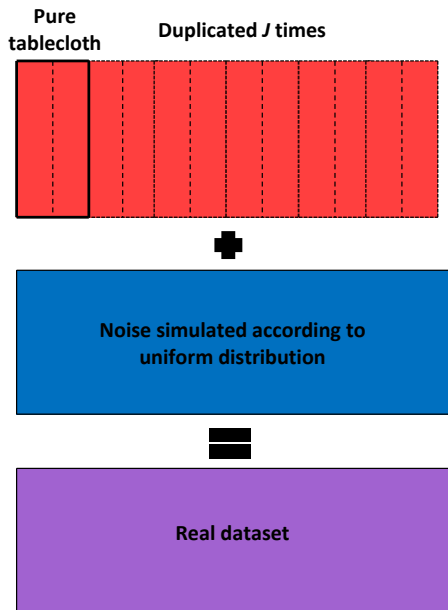
Data simulation procedure



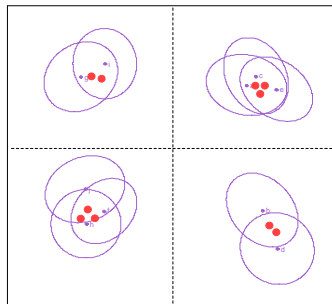
Ellipses according to real data



Data simulation procedure



Ellipses according to real data



Results of the simulations

30 judges

Noise/Signal	Frequency
10%	91.12%
20%	91.58%
40%	91.83%
100%	91.17%
200%	91%
400%	91.08%

Noise/Signal = 20%

Nb judges	Frequency
30	91.58%
50	92.87%
100	93.37%
200	93.37%
500	93.42%

Results of the simulations

30 judges

Noise/Signal	Frequency
10%	91.12%
20%	91.58%
40%	91.83%
100%	91.17%
200%	91%
400%	91.08%

Noise/Signal = 20%

Nb judges	Frequency
30	91.58%
50	92.87%
100	93.37%
200	93.37%
500	93.42%

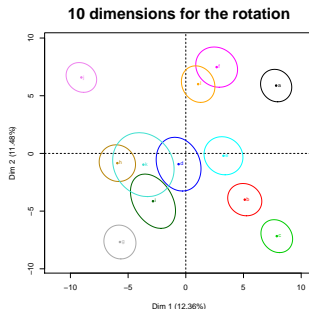
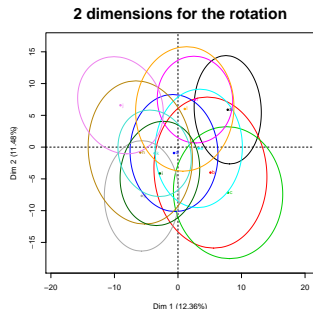
⇒ Small underestimation of the confidence level

- 1 What do we mean by “holistic approaches” and “confidence ellipse”?
- 2 Confidence ellipses construction
- 3 Why partial bootstrap doesn't work?
- 4 Validity of total bootstrap
- 5 Conclusion and perspectives

- Dimensionality problem highlighted: Confidence ellipses are essential (but may be built according to total bootstrap)
- Total bootstrap can be applied to all holistic approaches: napping, sorting, sorted napping, hierarchical sorting, free choice profiling
- Available into the R package `SensoMineR` through the `boot` function
- One parameter must be chosen: the number of dimensions for the Procrustean rotations

Conclusion and perspectives

Choice of the number of dimensions for the rotation



When the number of dimensions used for the Procrustean rotation increases:

- The size of the ellipses decreases
- The confidence level decreases