Toy Example

Permutation Closed Testing with Sum-Based Statistics

Data

Data with 5 variables and 10 permutations

G

(1)	(2)	(3)	(4)	(5)
28.42	16.68	9.36	6.12	9.40
0.10	0.06	1.37	0.08	0.56
0.69	3.07	4.33	0.83	0.36
1.07	30.31	1.11	8.55	0.26
0.22	7.45	2.87	0.48	1.02
1.83	0.04	2.85	0.04	0.02
17.68	1.82	6.00	1.52	1.06
1.77	26.12	0.29	0.26	4.07
2.71	0.37	8.47	5.83	4.42
1.14	0.03	24.06	8.84	2.41

We test $S = \{5\}$ with level $\alpha = 0.2$

Analysis

Elements for the Analysis

d_S	D				R			
(5)	(4)	(3)	(2)	(1)				
0.00	0.00	0.00	0.00	0.00	0.00 (1)	0.00 (2)	0.00 (3)	0.00 (4)
-8.84	-6.03	-7.99	-16.62	-28.32	-6.03 (4)	-7.99 (3)	-16.62 (2)	-28.32 (1)
-9.04	-5.29	-5.02	-13.61	-27.72	-5.02 (3)	-5.29 (4)	-13.61 (2)	-27.72 (1)
-9.14	2.43	-8.25	13.63	-27.34	13.63 (2)	2.43 (4)	-8.25 (3)	-27.34 (1)
-8.38	-5.63	-6.49	-9.23	-28.19	-5.63 (4)	-6.49 (3)	-9.23 (2)	-28.19 (1)
-9.38	-6.08	-6.51	-16.64	-26.59	-6.08 (4)	-6.51 (3)	-16.64 (2)	-26.59 (1)
-8.34	-4.59	-3.36	-14.86	-10.74	-3.36 (3)	-4.59 (4)	-10.74 (1)	-14.86 (2)
-5.33	-5.85	-9.07	9.44	-26.65	9.44 (2)	-5.85 (4)	-9.07 (3)	-26.65 (1)
-4.98	-0.28	-0.89	-16.31	-25.71	-0.28 (4)	-0.89 (3)	-16.31 (2)	-25.71 (1)
-6.99	2.72	14.70	-16.65	-27.27	14.70 (3)	2.72 (4)	-16.65 (2)	-27.27 (1)

Analysis

L_v and U_v are the 8-th ordered statistics of

$$\mathbf{d}_{\tilde{V}} = \mathbf{d}_S + \sum_{i=1}^{v} \mathbf{D}_i \qquad\qquad \mathbf{u}_v = \mathbf{d}_S + \sum_{i=1}^{v} \mathbf{R}_i$$

v	0	1	2	3	4
U_v	-5.33	4.11	0.00	-6.22	-33.49
L_v	-5.33	-5.26	-6.16	-6.22	-33.49
rej	T	?	?	T	T

Analysis

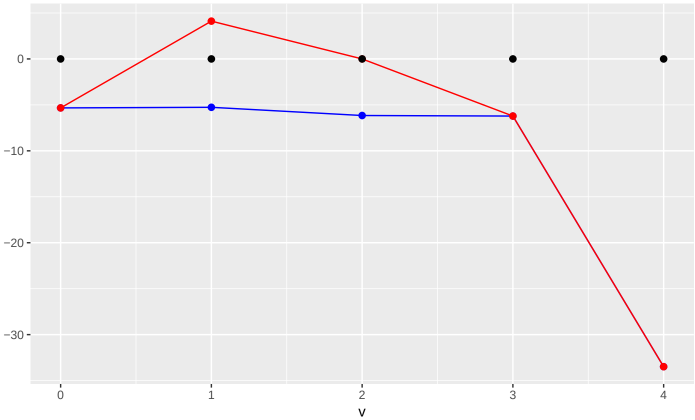


Figure 1: Upper (red) and lower (blue) critical values and observed values (zero, black) by additional superset size v.

Branch and Bound - Lowest Statistic

Branch and Bound - Lowest Statistic

The total space is partitioned according to the inclusion of 4.

In both subspaces, U_v decreases.

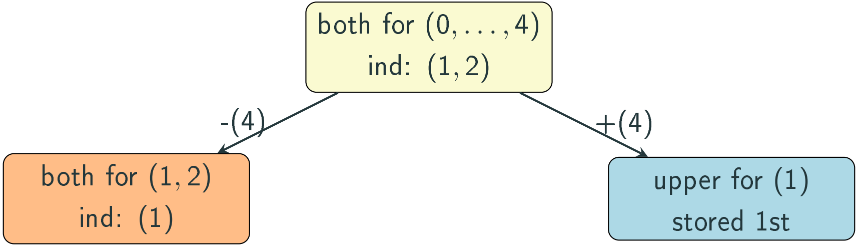
- $\mathbb{S}_{-4}$: L_v may change, hence we examine both bounds
- $\mathbb{S}_{+4}$: L_v does not change, hence we examine U_v

For each node, we save:

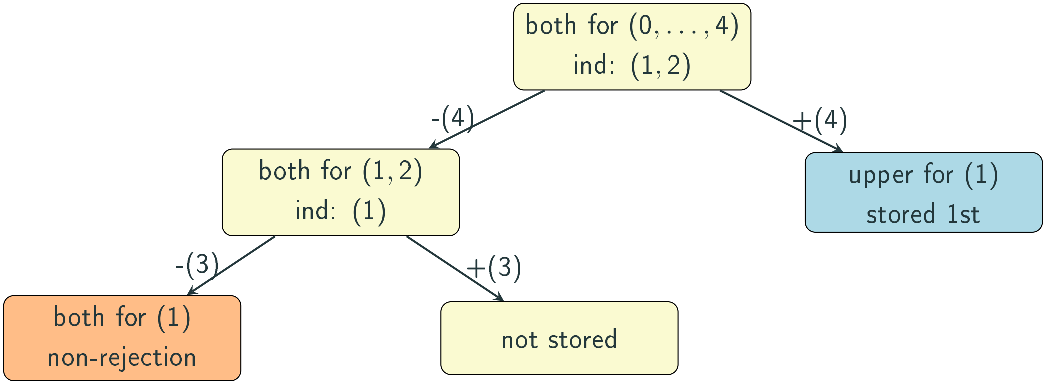
- sizes v to be examined (when keeping an index, v decreases of 1 unit)
- $\mathbf{R}$ and the corresponding indices
- cumulative sums of $\mathbf{d}_S + \mathbf{d}_{\text{kept}}$ with $\mathbf{R}$ and $\mathbf{D}$

Removal First - Step 1

- We enumerate the two subspaces: $\mathbb{S}_{+4}$ is stored, and $\mathbb{S}_{-4}$ is examined (both bounds)
- we keep removing indices until we can close a node
- then we start again from the node that was stored last

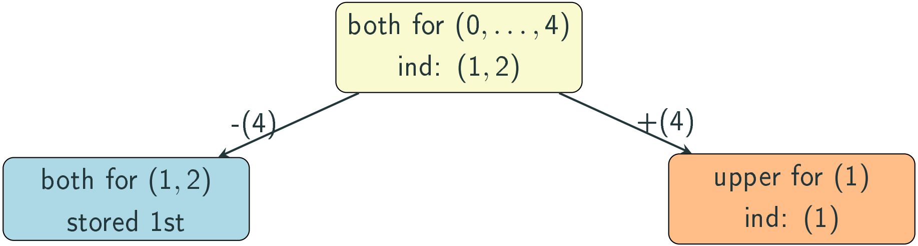


Removal First - Step 2

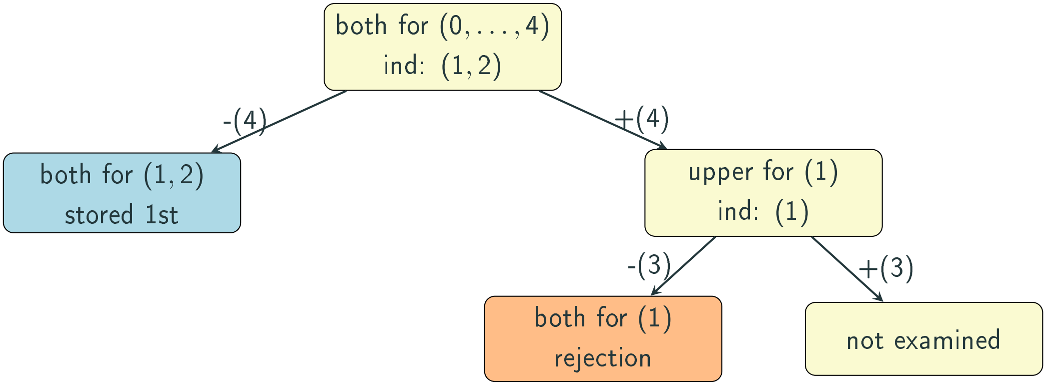


Keeping First - Step 1

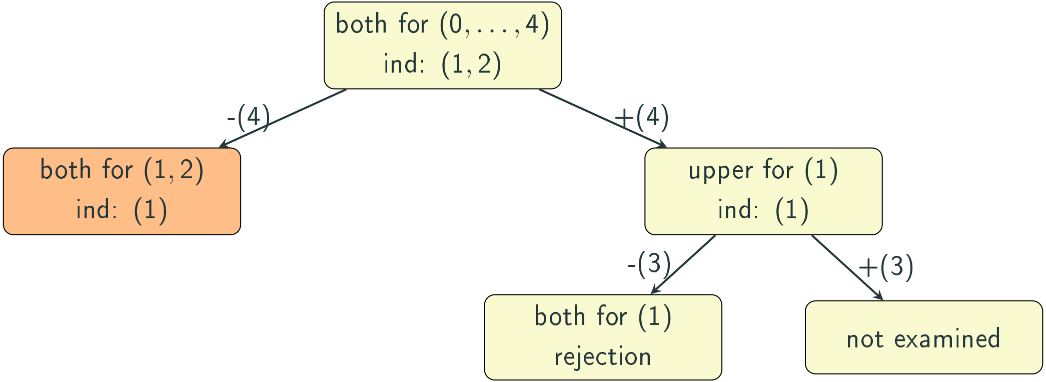
We start by examining U_v in $\mathbb{S}_{+4}$ (hence we cannot find any non-rejection).



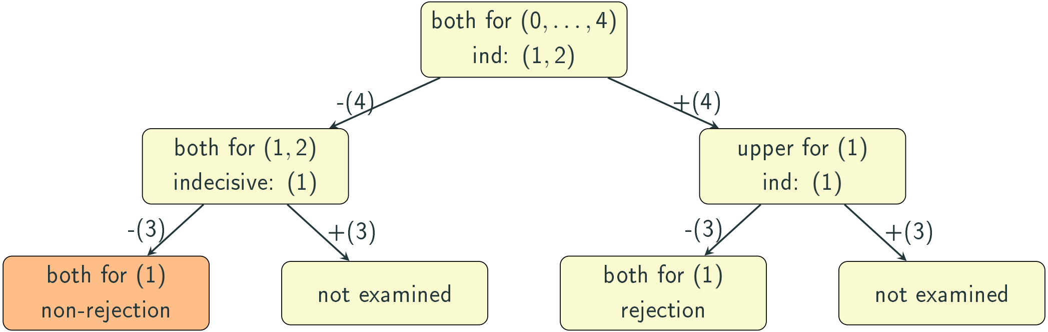
Keeping First - Step 2



Keeping First - Step 3



Keeping First - Step 4



Branch and Bound - Highest Statistic

Branch and Bound - Highest Statistic

The total space may also be partitioned according to the inclusion of 1.

As in the previous case, U_v decreases in both subspaces.

- $\mathbb{S}_{-1}$: L_v does not change, hence we examine U_v
- $\mathbb{S}_{+1}$: L_v may change, hence we examine both bounds

In this case, it takes 3 steps in both cases.

Simulations

Simulations (≈ 5000)

F: f variables, where a percentage f_* is significative.

S: contains a percentage s_{size} of all variables, where a percentage s_* is significative.

- f, $B \in \{10, 50, 100\}$
- $s_{\text{size}} \in \{1, 10, 20, 50, 80, 100\}$ (%)
- f_*, $s_* \in \{0, 1, 10, 20, 50, 80, 100\}$ (%)
- $\alpha \in \{0.05, 0.20\}$
- maximum number of iterations: 10^4

Simulations with $s_* = 0, 1, 10, 20$

s_*	f_*	s_{size}	f	B	α	non rej	RL	KL	RH	KH
1	1	10	100	50	0.05	T	-	**681**	958	-
1	1	10	100	100	0.05	-	-	-	-	-
1	1	20	100	100	0.05	F	-	-	**332**	**332**
1	1	50	50	100	0.20	F	335	335	**14**	**14**
10	20	80	50	50	0.05	T	**4**	6	5	6
20	20	20	50	10	0.20	T	545	18	**9**	93
20	10	50	50	50	0.05	F	16	16	**6**	**6**
20	10	50	50	100	0.05	F	**2**	**2**	4	4

Simulations with $s_* = 50, 80$

s_*	f_*	s_{size}	f	B	α	non rej	RL	KL	RH	KH
50	10	20	50	100	0.05	T	-	1970	**154**	856
50	10	20	100	50	0.20	-	-	-	-	-
50	10	20	100	100	0.20	F	-	-	**2092**	**2092**
50	50	20	10	50	0.20	F	15	15	**8**	**8**
50	50	20	10	100	0.20	F	6	6	**4**	**4**
50	50	50	50	100	0.20	F	1702	1702	**38**	**38**
80	10	10	50	10	0.20	T	-	**91**	309	316
80	20	20	50	50	0.20	F	3162	3162	**102**	**102**
80	20	20	50	100	0.20	F	218	218	**20**	**20**
80	20	20	100	10	0.20	F	-	-	**86**	**86**
80	20	20	100	50	0.20	F	-	-	**410**	**410**
80	20	20	100	100	0.20	F	-	-	**1320**	**1320**

Simulations with $s_* = 100$

s_*	f_*	s_{size}	f	B	α	non rej	RL	KL	RH	KH
100	1	1	100	10	0.20	T	-	**111**	880	-
100	1	1	100	50	0.20	F	-	-	**54**	**54**
100	1	1	100	100	0.20	F	-	-	**9144**	**9144**
100	10	10	50	10	0.20	T	-	**91**	236	972
100	10	10	100	50	0.20	T	-	**345**	-	-
100	10	10	100	100	0.20	-	-	-	-	-
100	20	20	50	50	0.05	T	5413	**35**	38	57
100	20	20	100	100	0.05	F	134	134	**12**	**12**
100	50	50	10	50	0.20	F	**4**	**4**	**4**	**4**
100	80	50	100	10	0.20	F	890	890	**64**	**64**
100	100	50	50	100	0.05	T	625	18	**14**	94
100	100	50	100	10	0.20	F	-	-	**92**	**92**
100	100	50	100	100	0.20	-	-	-	-	-

Simulation Results

When S is rejected, RL=KL and RH=KH.

RH required the smallest number of iterations.

	RL	KL	**RH**	KH
d	45.5	63.6	**84.8**	78.8
p	9.1	24.2	**63.6**	54.5
M	5851	3935	**2012**	2612

- d = percentage of simulations where the BAB leads to a decisive outcome
- p = percentage of simulations where the setting was optimal
- M = mean of iterations (when the number exceeds the maximum, it is approximated to the maximum)

Analysis Time in R

Mean analysis time (in milliseconds) when removing the highest statistic:

	$f = 10$		$f = 50$		$f = 100$	
	$\alpha = 0.05$	$\alpha = 0.20$	$\alpha = 0.05$	$\alpha = 0.20$	$\alpha = 0.05$	$\alpha = 0.20$
$B = 10$	1.2	0.9	1.1	1.4	0.6	16.0
$B = 50$	1.8	1.6	1.9	3.3	12.1	109.7
$B = 100$	1.9	1.3	3.3	3.4	215.9	1129.2
mean	1.6	1.3	2.1	2.7	76.2	418.3