Hyperparameter experiment - effect of # samples on the KME and Wishart test

In this experiment I test how much the sample size per variable influence the scores of the KME and Wishart test. This is an interesting experiment because the amount of samples is one of the few settings that can impact the Wishart test. For this experiment 10 test datasets where used per hyperparameter setting combination. It was assumed that giving the KME the same samples size during training and testing would give the best results.

The Wishart test clearly performs better when given more samples per variable. Interesting to see is that this is caused by the Wishart test getting better results on cases that are not t-separated. Meaning that on low samples sizes, the Wishart test too easily classifies a tetrad as t-separated.

Overview of the Wishart test results for the linear case

	b	d	nsamples	linear	accuracy	acc_std	trueneg	falseneg	truepos	falsepos
0	0	0	200	True	0.57037	0.0317539	252.1	5.1	594.9	632.9
1	0	0	500	True	0.741886	0.0356175	506.8	5.1	594.9	378.2
2	0	0	1000	True	0.832795	0.015716	640.7	4	596	244.3
3	0	0	2000	True	0.898047	0.00941798	736.9	3.3	596.7	148.1
4	0	0	10000	True	0.942761	0.0336506	804	4	596	81
5	0.05	0	200	True	0.577037	0.0430888	260.9	4	596	624.1
6	0.05	0	500	True	0.721347	0.0302136	477.4	6.2	593.8	407.6
7	0.05	0	1000	True	0.839327	0.019063	649.8	3.4	596.6	235.2
8	0.05	0	2000	True	0.90532	0.0147985	747.6	3.2	596.8	137.4
9	0.05	0	10000	True	0.951246	0.0118955	817.8	5.2	594.8	67.2
10	0.1	0	200	True	0.571178	0.0494783	255	6.8	593.2	630
11	0.1	0	500	True	0.727811	0.0326918	484.5	3.7	596.3	400.5
12	0.1	0	1000	True	0.833199	0.0132456	639.8	2.5	597.5	245.2
13	0.1	0	2000	True	0.896229	0.0124956	736	5.1	594.9	149
14	0.1	0	10000	True	0.955152	0.0192315	821.7	3.3	596.7	63.3

Overview of the Wishart test results for the nonlinear case

	b	d	nsamples	linear	accuracy	acc_std	trueneg	falseneg	truepos	falsepos
0	0.01	0.01	200	False	0.563165	0.026451	240	3.7	596.3	645
1	0.01	0.01	500	False	0.73899	0.0267751	502.4	5	595	382.6
2	0.01	0.01	1000	False	0.824983	0.0236458	631.8	6.7	593.3	253.2
3	0.01	0.01	2000	False	0.895623	0.0116377	734.4	4.4	595.6	150.6
4	0.01	0.01	10000	False	0.9567	0.0199461	825.1	4.4	595.6	59.9
5	0.01	0.05	200	False	0.587677	0.0407839	279.8	7.1	592.9	605.2
6	0.01	0.05	500	False	0.73138	0.0228724	493	6.9	593.1	392
7	0.01	0.05	1000	False	0.828215	0.0213809	636.6	6.7	593.3	248.4
8	0.01	0.05	2000	False	0.897037	0.0132851	735.3	3.2	596.8	149.7
9	0.01	0.05	10000	False	0.95138	0.0227384	814.9	2.1	597.9	70.1
10	0.05	0.01	200	False	0.58835	0.0152441	277.5	3.8	596.2	607.5
11	0.05	0.01	500	False	0.731246	0.0345831	488.6	2.7	597.3	396.4
12	0.05	0.01	1000	False	0.824377	0.0122486	627.6	3.4	596.6	257.4
13	0.05	0.01	2000	False	0.893872	0.0124582	729.2	1.8	598.2	155.8
14	0.05	0.01	10000	False	0.962155	0.0240731	832	3.2	596.8	53
15	0.05	0.05	200	False	0.584175	0.0386725	271	3.5	596.5	614
16	0.05	0.05	500	False	0.746734	0.0348039	515.4	6.5	593.5	369.6
17	0.05	0.05	1000	False	0.83367	0.0188766	644.6	6.6	593.4	240.4
18	0.05	0.05	2000	False	0.901145	0.0100525	743.2	5	595	141.8
19	0.05	0.05	10000	False	0.942963	0.0289398	803.6	3.3	596.7	81.4

Some results for the KME given different sample size

The kme was only tested on nonlinar data in this case.

n_samples	b	d	KME	E	K	n_distributions	best_score	mean_score	stdev_score
200	0.05	0.05	4	1000	400	4000	81.6162	74.8687	4.12293
500	0.05	0.05	4	1000	400	4000	82.6936	80.3098	1.38242
1000	0.05	0.05	4	1000	400	4000	81.4141	79.0707	1.2619
2000	0.05	0.05	4	1000	400	4000	84.3771	79.9529	3.67877

A single experiment with the KME to display the distribution of the confusion matrix.

	train_lin	b	d	KME	E	K	n_samples	n_distributions	score	trueneg	falseneg	truepos	falsepos
0	[False]	0.05	0.05	4	500	400	2000	4000	0.816835	843	230	370	42
1	[False]	0.05	0.05	4	500	400	2000	4000	0.823569	848	225	375	37
2	[False]	0.05	0.05	4	500	400	2000	4000	0.783165	809	246	354	76
3	[False]	0.05	0.05	4	500	400	2000	4000	0.847138	870	212	388	15
4	[False]	0.05	0.05	4	500	400	2000	4000	0.826263	847	220	380	38
5	[False]	0.05	0.05	4	500	400	2000	4000	0.806734	801	203	397	84
6	[False]	0.05	0.05	4	500	400	2000	4000	0.79798	809	224	376	76
7	[False]	0.05	0.05	4	500	400	2000	4000	0.788552	778	207	393	107
8	[False]	0.05	0.05	4	500	400	2000	4000	0.8	839	251	349	46
9	[False]	0.05	0.05	4	500	400	2000	4000	0.793266	804	226	374	81

Hyperparameter experiment - effect of # weights and trees on score KME.

This is an experiment to see if the # of trees E and the # of weights K influence the score of the KME given a set amount of samples and fixed nonlinearity parameters. The reason for this experiment is that the Wishart test performs increasingly well when given a larger amount of samples per variable. I wanted to see if a larger amount of samples per variable could also have a positive effect on the KME score. Each mean score represent 10 test cases.

	train_lin	b	d	KME	E	K	n_samples	n_distributions	best_score	mean_score	std_score
0	[False]	0.05	0.05	4	1000	400	4000	4000	87.2054	84.1953	1.61982
1	[False]	0.05	0.05	4	1000	800	4000	4000	85.3199	80.7407	2.02648
2	[False]	0.05	0.05	4	1000	1500	4000	4000	84.1751	82.5589	1.17218
3	[False]	0.05	0.05	4	2000	400	4000	4000	86.936	84.0067	1.57704
4	[False]	0.05	0.05	4	2000	800	4000	4000	87.0034	82.4916	2.0845
5	[False]	0.05	0.05	4	2000	1500	4000	4000	85.1852	82.6061	1.38761
6	[False]	0.05	0.05	4	4000	400	4000	4000	87.0034	84.0606	1.54737
7	[False]	0.05	0.05	4	4000	800	4000	4000	87.4747	82.3704	2.31943
8	[False]	0.05	0.05	4	4000	1500	4000	4000	85.0505	82.1414	1.42653

Questions answered

How does the amount of samples per variable influence accuracy of the methods?

For the KME there is no strong correlation between the amount of samples per variable and the accuracy.

The Wishart test clearly improves given more samples per variable.

Does the amount of weights and trees used influence the results of the KME?

There seems to be no clear improvement by increasing these parameters. Strange enough, it looks like having a low amount of weights improves the scores.
