

Simulation is based on 1e+05 simulated haplo-  
types

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	0.004	0.034		29.124	990.758	1.306	4.031
DYS389I	0.545	0.255	125.077	2441.356	1697.606	1.016	1.206
DYS389II	0.049	0.025	1845.990	1861.456	1192.436	0.471	0.215
DYS390			800.304				0.275
DYS391	0.158	0.015	2321.402	1752.595	3779.290	0.073	0.068
DYS392	0.431	0.233	6184.992	6477.672	5626.380	0.114	0.048
DYS393	0.048	0.088	2156.867	2530.527	2280.330	0.135	

Table 1: Deviances for each loci for *berlin* dataset for observed single marginals vs. simulated single marginals for the classification method specified in each row. The missing value is for the locus where the observed marginal distribution has been used.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	1.000	1.000		2.192e-05	0.000	0.934	0.545
DYS389I	0.909	0.968	0.000	0.000e+00	0.000	0.797	0.752
DYS389II	1.000	1.000	0.000	0.000e+00	0.000	1.000	1.000
DYS390			0.000				0.998
DYS391	0.984	0.999	0.000	0.000e+00	0.000	0.995	0.995
DYS392	1.000	1.000	0.000	0.000e+00	0.000	1.000	1.000
DYS393	1.000	1.000	0.000	0.000e+00	0.000	1.000	

Table 2: P-values for each loci for *berlin* dataset for observed single marginals vs. simulated single marginals for the classification method specified in each row. The missing value is for the locus where the observed marginal distribution has been used.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	
DYS19	6146.435	3613.047	2427.304	6056.612	101496.550	11626.910	1133
DYS389I	6800.568	6003.647	47391.265	284175.694	213863.411	15433.152	1522
DYS389II	10305.569	10280.674	957939.025	1025902.574	950110.699	19670.806	1960
DYS390	4819.944	3644.867	183202.984	3174.681	7256.472	7926.205	715
DYS391	7247.513	6387.390	792277.877	608403.021	1071823.289	8984.748	909
DYS392	9781.260	10381.544	1759601.916	1723088.658	1796827.694	27489.492	2860
DYS393	17173.738	18181.792	570667.028	596488.035	588687.332	28473.844	2694

Table 3: Deviances for each loci for *berlin* dataset for simulated single marginals vs. predicted single marginals for the classification method specified in each row.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS389I	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS389II	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS390	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS391	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS392	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS393	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 4: P-values for each loci for *berlin* dataset for simulated single marginals vs. predicted single marginals for the classification method specified in each row.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	0.346	0.298	45.324	2.842	1.719	0.930	0.276
DYS389I	0.001	0.011	429.444	324.124	265.818	0.050	0.111
DYS389II	0.003	0.016		361.196	361.559		0.052
DYS390			99.076		112.374	0.206	1.180
DYS391	0.067	0.059	343.926	368.196	442.527	0.625	0.006
DYS392	0.362	0.280	427.035	452.997		0.119	0.704
DYS393	0.053	0.050	386.076	488.271	573.695	0.131	
DYS437	0.413	0.438	18.896	5.096	9.121	0.695	1.061
DYS438	0.186	0.198	369.918	289.744	369.075	1.844	6.609
DYS439	0.017	0.059	244.164	232.435	236.350	0.024	0.028

Table 5: Deviances for each loci for *dane* dataset for observed single marginals vs. simulated single marginals for the classification method specified in each row. The missing value is for the locus where the observed marginal distribution has been used.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	0.987	0.990	3.404e-09	0.585	0.787	0.920	0.991
DYS389I	1.000	1.000	0.000e+00	0.000	0.000	0.997	0.991
DYS389II	1.000	1.000		0.000	0.000		1.000
DYS390			0.000e+00		0.000	0.999	0.947
DYS391	0.996	0.996	0.000e+00	0.000	0.000	0.891	1.000
DYS392	0.999	1.000	0.000e+00	0.000		1.000	0.994
DYS393	0.997	0.997	0.000e+00	0.000	0.000	0.988	
DYS437	0.813	0.803	7.885e-05	0.078	0.010	0.706	0.588
DYS438	0.996	0.995	0.000e+00	0.000	0.000	0.764	0.158
DYS439	1.000	1.000	0.000e+00	0.000	0.000	1.000	1.000

Table 6: P-values for each loci for *dane* dataset for observed single marginals vs. simulated single marginals for the classification method specified in each row. The missing value is for the locus where the observed marginal distribution has been used.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	
DYS19	36813.152	36823.334	105764.063	91096.629	46878.412	103654.271	10280
DYS389I	43886.085	43596.286	971191.881	612178.551	532762.492	55287.515	5463
DYS389II	36299.390	36063.044	27075.290	425518.928	439524.995	97020.448	9299
DYS390	17848.003	18055.170	118226.858	26290.164	287106.809	54609.489	5068
DYS391	8596.953	8939.870	889934.378	690709.072	927639.552	18947.551	1700
DYS392	31235.424	30891.474	411992.063	561214.883	39642.845	56979.989	6732
DYS393	14572.406	14569.930	441607.189	420809.244	481268.116	32608.814	3078
DYS437	18581.848	18326.046	6824.812	25695.188	22817.184	57769.976	5567
DYS438	31396.407	30680.074	286317.452	397385.617	404370.387	72830.471	8084
DYS439	14483.069	14672.152	173686.417	152605.200	169558.346	24230.403	2142

Table 7: Deviances for each loci for *dane* dataset for simulated single marginals vs. predicted single marginals for the classification method specified in each row.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS389I	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS389II	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS390	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS391	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS392	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS393	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS437	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS438	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DYS439	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 8: P-values for each loci for *dane* dataset for simulated single marginals vs. predicted single marginals for the classification method specified in each row.



	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	3.522e-02	0.072		71.795			
DYS389I	1.424e-02	0.004	1265.321	1276.953	1129.937		
DYS389II			806.328		775.947		
DYS390	3.447e-03	0.014	600.665	351.185	1225.655		
DYS391	1.811e-01	0.251	1046.742	951.723	839.583		
DYS392	1.188e-02	0.033	44.416	86.231	70.017		
DYS393	6.674e-03	0.004	977.487	966.367	510.341		
DYS437	5.596e-04	0.003	1.753	13.765	22.098		
DYS438	1.417e-02	0.036	671.610	1086.726	552.016		
DYS439	4.272e-04	0.007	249.942	380.558	347.514		

Table 9: Deviances for each loci for *somali* dataset for observed single marginals vs. simulated single marginals for the classification method specified in each row. The missing value is for the locus where the observed marginal distribution has been used.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	1.000	1.000		1.750e-13			
DYS389I	1.000	1.000	0.000e+00	0.000e+00	0.000e+00		
DYS389II			0.000e+00		0.000e+00		
DYS390	1.000	1.000	0.000e+00	0.000e+00	0.000e+00		
DYS391	0.913	0.882	0.000e+00	0.000e+00	0.000e+00		
DYS392	1.000	0.998	1.232e-09	0.000e+00	4.219e-15		
DYS393	1.000	1.000	0.000e+00	0.000e+00	0.000e+00		
DYS437	1.000	0.998	4.163e-01	1.026e-03	1.590e-05		
DYS438	1.000	0.998	0.000e+00	0.000e+00	0.000e+00		
DYS439	1.000	1.000	0.000e+00	0.000e+00	0.000e+00		

Table 10: P-values for each loci for *somali* dataset for observed single marginals vs. simulated single marginals for the classification method specified in each row. The missing value is for the locus where the observed marginal distribution has been used.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrm
DYS19	104166.740	104124.363	50226.264	397374.443	129241.982		
DYS389I	20574.035	21183.276	1506258.568	1834625.318	1707477.896		
DYS389II	37155.382	36706.482	1059579.338	49866.327	1297581.520		
DYS390	60015.265	59224.824	779919.822	1024328.017	1721344.347		
DYS391	24347.220	25006.351	1279896.559	1438700.554	1495376.523		
DYS392	97458.496	97443.272	89847.836	122654.291	122184.690		
DYS393	23191.154	23135.886	538993.779	685774.879	579847.748		
DYS437	21818.829	22042.919	28647.136	133725.291	189434.056		
DYS438	78321.312	78321.957	508134.015	842245.659	783630.918		
DYS439	6433.126	6539.921	148604.670	423094.557	335934.868		

Table 11: Deviances for each loci for *somali* dataset for simulated single marginals vs. predicted single marginals for the classification method specified in each row. The missing value is for the locus where the observed marginal distribution has been used.

	rpart	rpart-tun-dev	svm	svm-tun-perf	svm-tun-n	polr	lrn
DYS19	0.000	0.000	0.000	0.000	0.000		
DYS389I	0.000	0.000	0.000	0.000	0.000		
DYS389II	0.000	0.000	0.000	0.000	0.000		
DYS390	0.000	0.000	0.000	0.000	0.000		
DYS391	0.000	0.000	0.000	0.000	0.000		
DYS392	0.000	0.000	0.000	0.000	0.000		
DYS393	0.000	0.000	0.000	0.000	0.000		
DYS437	0.000	0.000	0.000	0.000	0.000		
DYS438	0.000	0.000	0.000	0.000	0.000		
DYS439	0.000	0.000	0.000	0.000	0.000		

Table 12: P-values for each loci for *somali* dataset for simulated single marginals vs. predicted single marginals for the classification method specified in each row. The missing value is for the locus where the observed marginal distribution has been used.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
rpart	1.236	1.449	0.268
rpart-tun-dev	0.650	1.408	0.425
svm	13434.632	2363.861	5664.264
svm-tun-perf	15092.731	2524.901	5185.304
svm-tun-n	15566.800	2372.237	5473.109
polr	3.114	4.623	
lrm	5.842	10.025	

Table 13: Sum of deviances for observed single marginals vs. simulated single marginals for the classification method specified in each row.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
Maximum	6	9	9
rpart	0	0	0
rpart-tun-dev	0	0	0
svm	6	9	8
svm-tun-perf	6	7	9
svm-tun-n	6	8	9
polr	0	0	
lrm	0	0	

Table 14: Number of loci rejected for observed single marginals vs. simulated single marginals for the classification method specified in each row.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
rpart	62275.026	253712.736	473481.558
rpart-tun-dev	58492.960	252617.382	473729.252
svm	4313507.400	3432620.404	5990107.989
svm-tun-perf	4247289.275	3403503.476	6952389.336
svm-tun-n	4730065.449	3351569.139	8362054.546
polr	119605.156	573938.928	
lrm	118027.139	574184.882	

Table 15: Sum of deviances for simulated single marginals vs. predicted single marginals for the classification method specified in each row.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
Maximum	7	10	10
rpart	7	10	10
rpart-tun-dev	7	10	10
svm	7	10	10
svm-tun-perf	7	10	10
svm-tun-n	7	10	10
polr	7	10	
lrm	7	10	

Table 16: Number of loci rejected for simulated single marginals vs. predicted single marginals for the classification method specified in each row.



	<i>berlin</i>	<i>dane</i>	<i>somali</i>
rpart	Inf	768.444	772.936
rpart-tun-dev	458.284	770.766	779.840
svm	Inf	24971.264	53797.852
svm-tun-perf	Inf	Inf	Inf
svm-tun-n	Inf	Inf	Inf
polr	1761.153	Inf	
lrm	1778.344	2126.960	

Table 17: Sum of deviances for observed pairwise marginals vs. simulated pairwise marginals for the classification method specified in each row.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
Maximum	21	45	45
rpart	18	35	34
rpart-tun-dev	17	34	36
svm	21	45	45
svm-tun-perf	21	45	45
svm-tun-n	21	45	45
polr	21	45	
lrm	21	45	

Table 18: Number of loci rejected for observed pairwise marginals vs. simulated pairwise marginals for the classification method specified in each row.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
<code>rpart</code>	Inf	Inf	Inf
<code>rpart-tun-dev</code>	Inf	Inf	Inf
<code>svm</code>	Inf	Inf	Inf
<code>svm-tun-perf</code>	Inf	Inf	Inf
<code>svm-tun-n</code>	Inf	Inf	Inf
<code>polr</code>	Inf	Inf	
<code>lrm</code>	Inf	Inf	

Table 19: Sum of deviances for simulated pairwise marginals vs. predicted pairwise marginals for the classification method specified in each row.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
Maximum	21	45	45
rpart	21	45	45
rpart-tun-dev	21	45	45
svm	21	45	45
svm-tun-perf	21	45	45
svm-tun-n	21	45	45
polr	21	45	
lrm	21	45	

Table 20: Number of loci rejected for simulated pairwise marginals vs. predicted pairwise marginals for the classification method specified in each row.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
Estimate	0.364	0.602	0.277
CI	[0.321; 0.408]	[0.510; 0.694]	[0.212; 0.342]
rpart	0.478	0.71	0.42
rpart-tun-dev	0.451	0.71	0.42
svm	0.526	0.792	0.43
svm-tun-perf	0.633	0.873	0.578
svm-tun-n	0.617	0.847	0.513
polr	0.639	0.886	
lrm	0.637	0.887	

Table 21: Unobserved probabilities using classification methods. The missing values indicate that the method did not succeed for some reason.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
rpart	1730.27	1043.21	662.39
rpart-tun-dev	1616.90	1043.21	662.39
svm	1627.41	883.21	474.45
svm-tun-perf	2284.33	1302.00	802.80
svm-tun-n	2482.69	1235.06	786.74
polr	2589.24	1502.96	
lrm	2592.33	1510.01	

Table 22: Comparing deviance between relative frequencies and predicted probabilities found using classification methods.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
rpart	0.00e+00	1.97e-289	1.70e-178
rpart-tun-dev	0.00e+00	1.97e-289	1.70e-178
svm	0.00e+00	1.09e-254	1.10e-137
svm-tun-perf	0.00e+00	0.00e+00	5.52e-209
svm-tun-n	0.00e+00	0.00e+00	1.69e-205
polr	0.00e+00	0.00e+00	
lrn	0.00e+00	0.00e+00	

Table 23: The probability for the observed counts assuming a multinomial distribution with the predicted probabilities using classification methods as the probabilities parameter.

	<i>berlin</i>	<i>dane</i>	<i>somali</i>
Reciprocal	-4.60e+03	-2.16e+03	-2.11e+03
rpart	-1.24e+03	-6.65e+02	-4.09e+02
rpart-tun-dev	-1.19e+03	-6.65e+02	-4.09e+02
svm	-1.19e+03	-5.85e+02	-3.15e+02
svm-tun-perf	-1.52e+03	-7.94e+02	-4.80e+02
svm-tun-n	-1.62e+03	-7.61e+02	-4.72e+02
polr	-1.67e+03	-8.95e+02	
lrm	-1.67e+03	-8.98e+02	

Table 24: The log probability for the observed counts assuming a multinomial distribution with the predicted probabilities using classification methods as the probabilities parameter. The reciprocal row is the probability of sampling a certain element in the support for the multinomial distribution assuming uniform sampling probabilities.