Continuous Buhmbox - Update 11/14

Alexandre Lamy - `all2187@columbia.edu`

November 13, 2017

Vectorization and speedup

I wrote vectorized versions of the SNP generation, population generation, and buhmbox code. Together these allowed for an $\approx 20x$ speedup over the previous unvectorized individuals.

More specifically 100 runs of generating 3 populations of 100000 individuals (independent, pleiotropic, heterogeneous) and running buhmbox and continuous buhmbox on each used to take about 2 hours and now takes about 6-7 minutes.

Added noise to reduce heritability

I added random noise to phenotypes ($N(0, \sigma^2)$ with $\sigma^2 = \frac{1-h}{h} VarPhen$) to allow for a custom, user defined heritability.

Results

We used 100 runs, 100000 individuals, 100 snps per phenotype, and ran with $h \in \{1, .5, .3, .2, .15, .1, .05\}$. The results were the following:

Total Heritability	Regular Buhmbox Mean and Std	Continuous Buhmbox Mean and Std
1.00	-0.08, 1.00	0.23, 0.75
0.50	0.11, 1.04	0.24, 0.78
0.30	-0.08, 0.96	0.13, 0.67
0.20	0.08, 1.06	0.21, 0.74
0.15	0.12, 0.93	0.22, 0.74
0.10	-0.24, 0.96	-0.04, 0.76
0.05	0.09, 1.03	0.06, 0.69

Table 1: Independent Population

Since we are looking at only phen1 snps but phen2 cases, h is technically always 0. (Is this correct?)

Total Heritability	Regular Buhmbox Mean and Std	Continuous Buhmbox Mean and Std
1.00	-17.46, 0.62	-17.61, 0.85
0.50	-8.36, 0.76	-8.55, 0.72
0.30	-5.18, 0.96	-5.17, 0.79
0.20	-3.11, 0.96	-3.26, 0.70
0.15	-2.57, 0.90	-2.52, 0.69
0.10	-1.68, 0.98	-1.53, 0.77
0.05	-0.66, 1.02	-0.62, 0.73

Table 2: Pleiotropic Population

Since we are looking at only phen1 snps but phen2 cases, h is technically 0.5 of the total h. (Is this correct?)

Total Heritability	Regular Buhmbox Mean and Std	Continuous Buhmbox Mean and Std
1.00	39.92, 1.36	34.32, 1.13
0.50	19.41, 1.33	17.03, 0.89
0.30	11.29, 1.07	10.25, 0.79
0.20	7.55, 1.20	6.82, 0.82
0.15	5.24, 1.08	4.94, 0.78
0.10	3.73, 1.15	3.51, 0.81
0.05	1.47, 0.89	1.66, 0.69

Table 3: Heterogeneous Population

Since we are looking at only phen1 snps but phen2 cases, h is technically 0.375 of the total h. (Is this correct?)

There was a previous mistake where regular buhmbox was not using full data, hence the better scores this time. That said the standard deviation for independent population for the continuous buhmbox is significantly lower than 1 (≈ 0.7) which would allow us to multiply by an additional factor of $\frac{1}{0.7}^2 \approx 2$ which would make the results much more conclusive.

Tasks

- Find mathematical backing for the extra 2x scaling

- Try different weight functions?

- Other?