

Faster computation of BayesR models

A more efficient implementation of BayesR in terms of computing time and memory requirements is provided in the software *bayesRv2* (see Table). The increased performance is largely due to an improved algorithm for updating effects across multiple SNPs in blocks (MPL Calus , *Genetics Selection Evolution*, 2014, **46**:24) and the use of multiple CPUs.

N _{ind} x N _{snp}	<i>bayesR</i>		<i>bayesRv2</i> (4 CPUs)	
	Memory (GB)	Time (h)	Memory (GB)	Time (h)
4,870 x 300,549 ¹⁾	11.5	16.8	3.4	13.3
43,900 x 123,437 ¹⁾	42.4	68.5	11.0	13.9
17,539 x 909,763 ¹⁾	156	not run	32.1	48.5
				10.1 ⁴⁾
106,355 x 500,000 ²⁾	415	~ 240 ³⁾	105	31.5 ⁴⁾

¹⁾ Intel E5-2670, 2.66 GHz; ²⁾ AMD 6378 2.4GHz; ³⁾ extrapolated; ⁴⁾ reduced update 500 SNPs; runtimes are based on 50,000 MCMC iterations.

Details

When updating SNP effects in turn, computing the dot product between residuals and genotypes of individuals and updating the residuals are the most time-consuming steps. Basically, we compute dot products of SNPs within a block at once, and when serially updating the SNP effects account for changes in the residual information from re-estimating previous SNPs within a block.

Currently, *bayesRv2* only accepts ‘best-guess’ genotypes, missing genotypes are allowed and are replaced by the mean genotype value of a given marker. SNPs are allocated to blocks at the start of the program, either in the order as they appear in the PLINK genotype file or in random order when the flag `-permute` is set.

The optimal number of SNPs to include in a block depends on the number of individuals. Our limited experience suggests that a block size of 4 SNPs will be optimal for most situations. Setting up the block structure requires additional overhead and this needs to be taken into account when short chains are used to determine the optimal size of the SNP blocks.

BayesRv2 uses less RAM as the complete scaled and centered genotype matrix is no longer stored in memory.

SNP block options

Additional options in *bayesRv2* are:

argument	type	description	default
<code>-blocksize</code>	[num]	number of SNPs in block	4
<code>-nthreads</code>	[num]	number of threads	4
<code>-shuffle</code>	[flag]	permute order of blocks	false

By default `-nthreads` is set equal to the number of SNPs within a block.

Example

```
bayesRv2 -bfile example/simdata -out simout -msize 500 -blocksize 6 \  
-permute -shuffle
```

The above command runs a ‘reduced update’ model (`-msize 500`) using blocks containing 6 SNPs. SNPs are allocated randomly to blocks (`-permute`) and the order of blocks is randomly permuted in each MCMC cycle (`-shuffle`).

References

MPL Calus. Right-hand-side updating for fast computing of genomic breeding values. *Genetics Selection Evolution* 2014, 46:24.
<http://www.gsejournal.org/content/46/1/24>