Power of detecting variance effects in MZ twins

2018-04-09

To estimate statistical power, we need to simulate the phenotype as a function of SNPs and random noise. Then we see how often a significant association at some alpha level is found using the model for a specific set of parameters.

For reference, the FTO variant reported to have an influence on the variance of BMI (Yang et al 2012) had an effect of 0.035 and an allele frequency f 0.4. This effect estimate is of course not necessarily entirely a variance effect as the method to estimate variance was not orthogonal to the additive effect.

Model 1: The SNP influences the variance

Each individual has a phenotype y that is a function of an additive genetic value a, a noise term z whose magnitude is determined by the genotype at SNP x, a common environment effect c and a residual variance e. The residual variance is specific to each individual $i1$ and $i2$.

The g is identical for each individual in a twin pair, and it is sampled with variance h^2, the heritability of the trait. This is also the case for the common environment effect, which is sampled with variance s^2. The residual variance is sampled from a normal distribution with variance $1 - h^2 - s^2$. Hence, each twin individual has a common and a specific environmental component.

Finally, the SNP x is in Hardy-Weinberg equilibrium with allele frequency p. It has a per allele effect on the variance of b.

$$y_{i1} = a_i + z_{i1} + c_i + e_{i1}$$
$$y_{i2} = a_i + z_{i2} + c_i + e_{i2}$$
$$a_i \sim N(0, h^2)$$
$$z_{i1} \sim N(0, bx_i)$$
$$z_{i2} \sim N(0, bx_i)$$
$$x_i \sim Binom(2, p)$$
$$c_i \sim N(0, s^2)$$

Model 2: The SNP influences the mean

This is identical to model 1 except for the following. The SNP no longer has an influence on the variance of the trait, it simply has an influence on the mean (as in standard GWAS).

$$y_{i1} = a_i + bx_i + c_i + e_{i1}$$
$$y_{i2} = a_i + bx_i + c_i + e_{i2}$$
$$a_i \sim N(0, h^2)$$
$$x_i \sim Binom(2, p)$$
$$c_i \sim N(0, s^2)$$

Model 3: The SNP influences the mean and the variance

This is a combination of model 1 and model 2 - the SNP influences both the mean and the variance.

$$y_{i1} = a_i + bx_i + z_{i1} + c_i + e_{i1}$$
$$y_{i2} = a_i + bx_i + z_{i2} + c_i + e_{i2}$$
$$a_i \sim N(0, h^2)$$
$$z_{i1} \sim N(0, bx_i)$$
$$z_{i2} \sim N(0, bx_i)$$
$$x_i \sim Binom(2, p)$$
$$c_i \sim N(0, s^2)$$

Tests

In the following the phenotypes y_1 and y_2 are standardised to have mean 0 and variance 1.

1. Influence of the SNP x on the squared difference between individuals in a twin pair. e.g. For a twin pair i

$$(y_{i1} - y_{i2})^2 = bx_i + e_i$$

2. Influence of the SNP x on the squared sum between individuals in a twin pair. e.g. For a twin pair i

$$(y_{i1} + y_{i2})^2 = bx_i + e_i$$

3. Influence of the SNP x on the raw phenotype of one individual from each pair. e.g. For a twin pair i

$$y_{i1} = bx_i + e_i$$

4. The discovery model used in Yang et al - influence of the SNP on the square of the phenotype. e.g. For a twin pair i

$$y_{i1}^2 = bx_i + e_i$$

5. Combining (1) and (2) in a meta analysis, as these are in principal orthogonal.

Results from power simulations

All simulations were performed for $h^2 = 0.2$.

1. The D^2 test appears to be unconfounded, returning only the variance effect estimate uncontaminated by the mean.
2. The S^2 and y^2 methods both improve power slightly when there is a mean and a variance effect.
3. The power of D^2 is much higher than y^2, though of course there are fewer readily available MZ twin pairs than unrelated samples.
4. Higher s^2 (common environment) improves power to detect variance effect.
5. More intermediate allele frequencies have higher power.
6. There is sufficient power to detect much smaller effects on the variance than were discovered in Yang et al at sample sizes of e.g. 20k MZ twin pairs, which is probably a reasonable target.
7. Winner's curse should help from doing multiple tests in GWAS.

A potential strategy could be to use the y^2 approach as a discovery and then follow up with D^2 to obtain orthogonal estimates of variance effects.
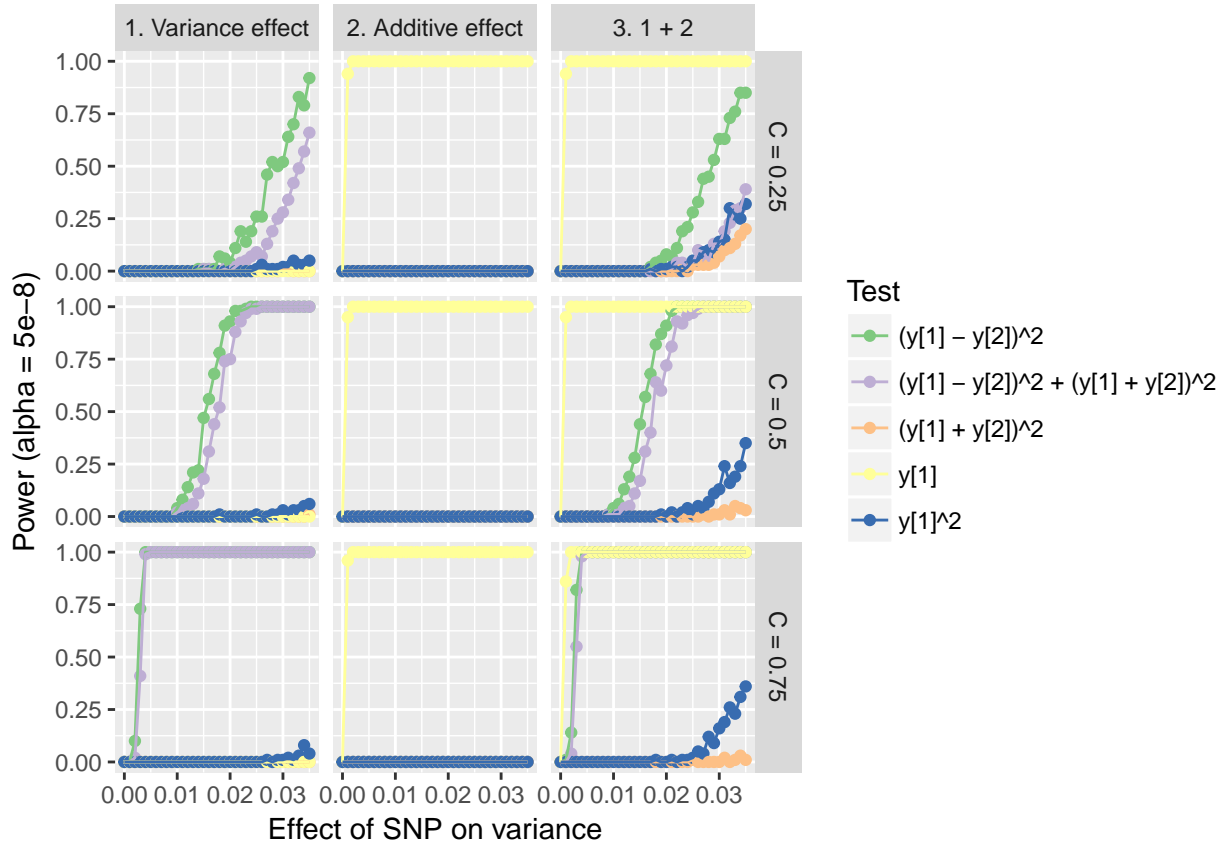
Figure 1: Power estimates for SNP MAF $= 0.4$ and 50000 MZs, across a range of scenarios and models, where $\alpha = 5 \times 10^{-8}$
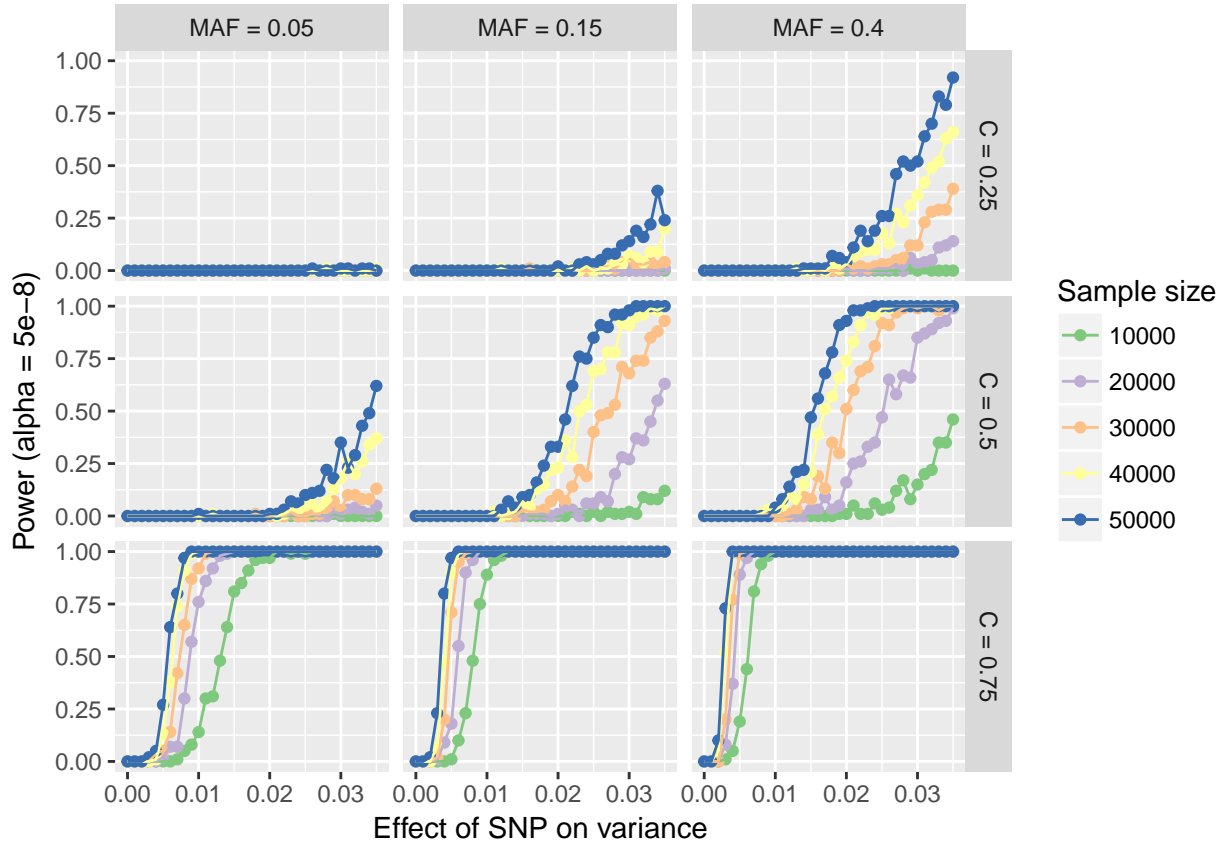
Figure 2: Power estimates for Model 1 using the D^2 method, where $\alpha = 5 \times 10^{-8}$