# Power of detecting variance effects in MZ twins

*2017-02-08*

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.

**Model 1: The SNP influences the variance**

Each individual has a phenotype $y$ that is a function of an additive genetic value $g$, a noise term $e$ whose magnitude is determined by the genotype at SNP $x$, and a residual variance $c$. The residual variance is estimated jointly for two individuals who compose an MZ pair $i$. They share some of the residual variance, and some of it is specific.

The $g$ is identical for each in a twin pair, and it is sampled with variance $h^2$, the heritability of the trait. The residual variance is sampled from a bivariate normal distribution, such that the variance is $1 - h^2$ and there is a correlation between the two individuals in a pair. Hence, each twin individual has a common and a specific environmental component. The extent to which this is divided is determined by $C$. e.g. The common environmental variance will be $C(1 - h^2)$ and the specific environmental variance will be $(1 - C)(1 - h^2)$.

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

$$
\begin{aligned}
y_{i1} &= g_i + e_{i1} + c_{i1} \\
y_{i2} &= g_i + e_{i2} + c_{i2} \\
g_i &\sim N(0, h^2) \\
e_{i1} &\sim N(0, bx_i) \\
e_{i2} &\sim N(0, bx_i) \\
x_i &\sim Binom(2, p) \\
c_i &\sim MVN(0, \sigma) \\
\sigma &= \begin{bmatrix} 1 - h^2 & C \\ C & 1 - h^2 \end{bmatrix}
\end{aligned}
$$

**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).

$$
\begin{aligned}
y_{i1} &= g_i + bx_i + c_{i1} \\
y_{i2} &= g_i + bx_i + c_{i2} \\
g_i &\sim N(0, h^2) \\
x_i &\sim Binom(2, p) \\
c_i &\sim MVN(0, \sigma) \\
\sigma &= \begin{bmatrix} 1 - h^2 & C \\ C & 1 - h^2 \end{bmatrix}
\end{aligned}
$$

**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.

$$
\begin{aligned}
y_{i1} &= g_i + bx_i + e_{i1} + c_{i1} \\
y_{i2} &= g_i + bx_i + e_{i2} + c_{i2} \\
g_i &\sim N(0, h^2) \\
e_{i1} &\sim N(0, bx_i) \\
e_{i2} &\sim N(0, bx_i) \\
x_i &\sim Binom(2, p) \\
c_i &\sim MVN(0, \sigma) \\
\sigma &= \begin{bmatrix} 1-h^2 & C \\ C & 1-h^2 \end{bmatrix}
\end{aligned}
$$

**Results from power simulations**

1. It appears that the mean and variance influences of a SNP are completely unconfounded. Model 2 gives no indication of detecting associations when the SNP influences only the mean. Model 1 and Model 3 are identical.
2. It is unlikely that SNP effect sizes on the variance exist at high magnitudes. It would be unlikely, for example, to find a SNP that influences 1% of the variance. There are relatively few scenarios where reasonable power would be achieved to detect an effect of that magnitude.

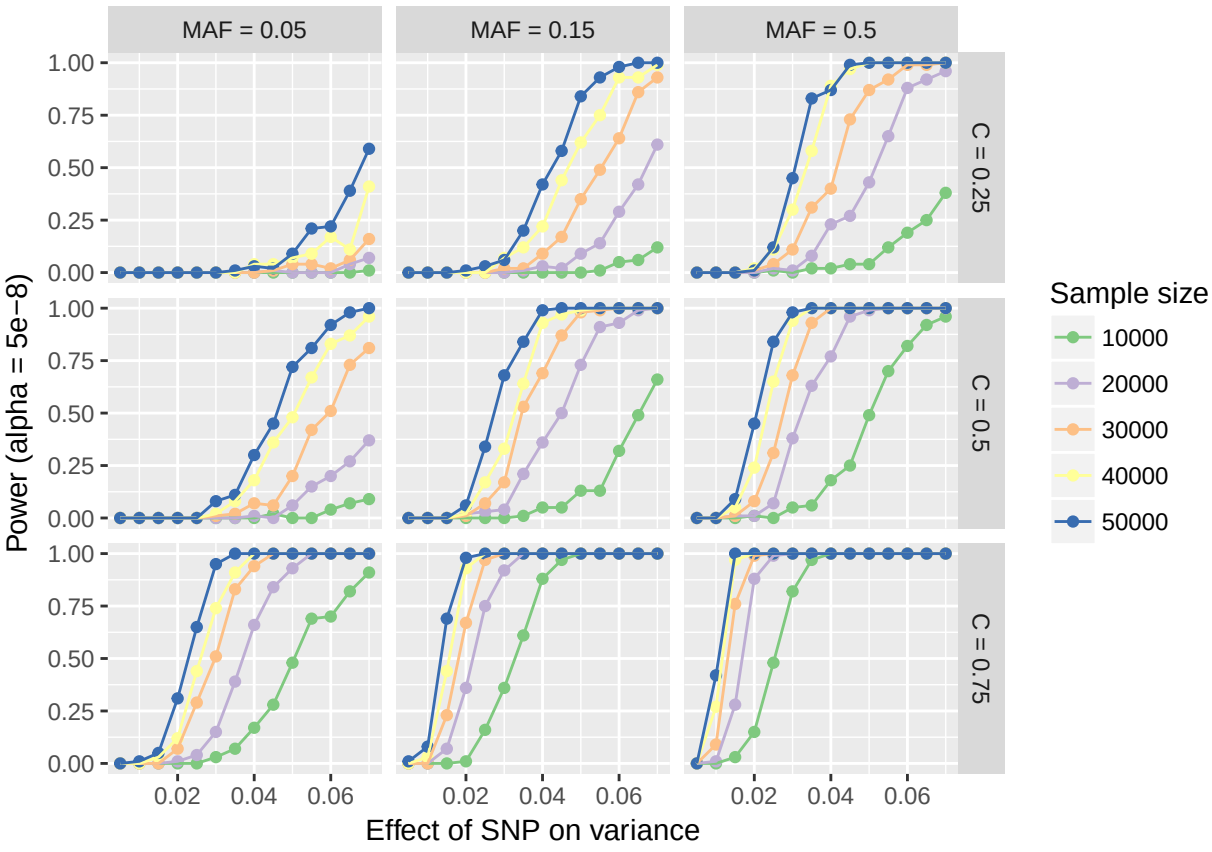


Figure 1: Power estimates for Model 1 across a range of scenarios, where $\alpha = 5 \times 10^{-8}$

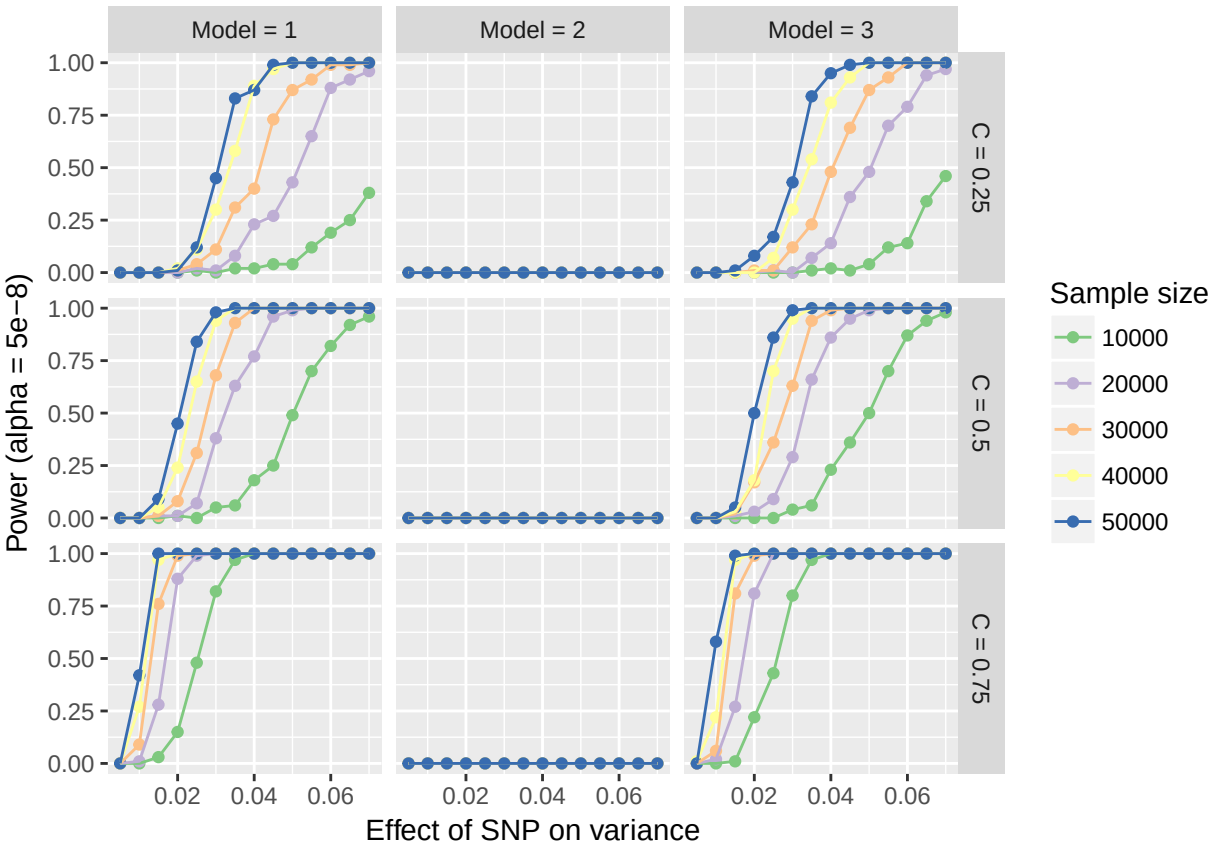


Figure 2: Power estimates for all models for minor allele frequency = 0.5, where $\alpha = 5 \times 10^{-8}$