

Power for de novo and case/control studies

Hailiang Huang¹, Benjamin Neale¹ and Stephan Sanders²

1 Analytic and Translational Genetics Unit, Massachusetts General Hospital and the Broad Institute of MIT and Harvard

2 Department of Psychiatry, School of Medicine, University of California San Francisco

Power for de novo analysis

Assume the expected number of de novo mutations per individual genome is Q_0 . If we perform the de novo analysis on a proportion of the genome (denoted as $f, f \in [0, 1]$), the expected number of mutations in this selected region of genome is fQ_0 .

For cases, assume a proportion (denoted as $p, p \in [0, 1]$) of the de novo mutations are functional with with relative risk of R . The expected number of de novo mutations in cases is (see proof 1):

$$Q_1 = (1 - p)fQ_0 + RpfQ_0 .$$

Simulation

We simulate the number of de novo mutations for controls using a poisson distribution with the poisson parameter of fQ_0 . We simulate the number of de novo mutations for cases using a mixture of two poisson distributions with poisson parameters of $(1 - p)fQ_0$ and $RpfQ_0$ respectively. We then perform a poisson regression and get the deviance between the poisson model (simulated number of de novo mutations versus case/control status) and the null model. The deviance is chi-square distributed and will be used to calculate the power. This process is repeated several times.

Parametric

When the sample size is large, poisson test can be approximated using the t test. The chi-square distributed test statistic can be calculated as (see proof 2):

$$\chi^2 = \frac{(Q_1 - Q_0)^2}{Q_1/N_1 + Q_0/N_0},$$

in which N_1 and N_0 are the sample size for case and controls respectively.

Power for case/control analysis

Let x_i be the genotype of variant i and y be the phenotype of all individuals. We assume there are s SNPs in a gene and f of them are functional. The allele frequency of the SNPs is sampled from the exponential distribution: $\Pr(AF) = \lambda \exp^{-\lambda AF}$ with $\lambda = 1/\overline{AF}$ (the sampled AF therefore has mean of \overline{AF}). The probability for being a causal variant scales with the allele frequency: $\Pr(\text{causal}|AF) = 1/AF$.

Single variant test

The power for a single variant can be calculated using the chi-square distribution with the non-centrality parameter χ^2 . This parameter is estimated using the 'NCPgen()' function in the source code (basically a 2x2 contingency table test). The power for finding at least one out of the f functional SNPs is

$$1 - \prod_{i \in f} (1 - \text{power}_{\text{snpi}})$$

Burden test

In burden test, assuming SNPs are independent, we have $x = \sum_{i \in f} x_i + \sum_{j \in s-f} x_j$. The coefficients are

$$\begin{aligned} \beta' &= \frac{\text{cov}(y, x)}{\text{var}(x)} \\ &= \frac{\sum_{i \in f} \text{cov}(x_i, y) + \sum_{j \in s-f} \text{cov}(x_j, y)}{\sum_{i \in f} \text{var}(x_i) + \sum_{j \in s-f} \text{var}(x_j)} \\ &= \frac{\sum_{i \in f} \text{cov}(x_i, y) + \sum_{j \in s-f} \text{cov}(x_j, y)}{\sum_{i \in f} \text{var}(x_i) + \sum_{j \in s-f} \text{var}(x_j)} \\ &= \frac{\beta \sum_{i \in f} \text{var}(x_i) + 0 * \sum_{j \in s-f} \text{var}(x_j)}{\sum_{i \in f} \text{var}(x_i) + \sum_{j \in s-f} \text{var}(x_j)} \\ &= \beta \frac{\sum_{i \in f} \text{var}(x_i)}{\sum_{i \in f} \text{var}(x_i) + \sum_{j \in s-f} \text{var}(x_j)} \end{aligned}$$

The power for burden test can be calculated using the chi-square distribution with the non-centrality parameter χ^2 calculated using the new coefficient β' .

Proofs

Proof 1

We prove that under the conditions a) the population prevalence of the disease is small ($\Pr(\text{case}) \ll 1$) and b) the de novo rate is low ($\Pr(A) \equiv \mu \ll 1$),

$$\frac{\Pr(\text{Alcase})}{\Pr(\text{Alcontrol})} \simeq R,$$

in which R is the relative risk. Recall that $R \equiv \frac{\Pr(case|A)}{\Pr(case)}$

$$\begin{aligned}
\frac{\Pr(A|case)}{\Pr(A|control)} &= \frac{\Pr(case|A)}{\Pr(control|A)} \times \frac{\Pr(control)}{\Pr(case)} \\
&= R \times \frac{\Pr(case|A)}{\Pr(control|A)} \times \frac{\Pr(control)}{\Pr(case)} \\
&= R \times \frac{\Pr(case|A)}{\Pr(case)} \times \frac{\Pr(control)}{\Pr(control|A)}.
\end{aligned} \tag{0}$$

We will then prove $\frac{\Pr(case|A)}{\Pr(case)} \simeq 1$ and $\frac{\Pr(control)}{\Pr(control|A)} \simeq 1$.

$$\begin{aligned}
\frac{\Pr(case|A)}{\Pr(case)} &= \frac{\Pr(case|A)}{\Pr(case|A)\Pr(A) + \Pr(case|A)\Pr(A)} \\
&= \frac{\Pr(case|A)}{(\Pr(A) + \frac{\Pr(case|A)}{\Pr(case)}\mu)^{-1}} \\
&= (\Pr(A) + R\mu)^{-1} \\
&= (1 - \mu + R\mu)^{-1} \\
&= (1 + (R - 1)\mu)^{-1} \\
&\simeq 1 - (R - 1)\mu \\
&\simeq 1.
\end{aligned} \tag{1}$$

And

$$\begin{aligned}
\frac{\Pr(control)}{\Pr(control|A)} &= \frac{\Pr(control)}{1 - \Pr(case|A)} \\
&= \frac{\Pr(control)}{1 - R \Pr(case|A)}.
\end{aligned} \tag{2}$$

For $\Pr(case|A)$, we have

$$\begin{aligned}
\Pr(case|A) &= \frac{\Pr(case)}{\Pr(A) + R \Pr(A)} \\
&= \frac{\Pr(case)}{1 + (R - 1)\mu} \\
&\simeq K - K(R - 1)\mu.
\end{aligned} \tag{3}$$

Using equations 2 and 3, we have

$$\begin{aligned}
\frac{\Pr(\text{control})}{\Pr(\text{control}|A)} &\simeq \frac{\Pr(\text{control})}{1 - R(K - K(R - 1)\mu)} \\
&\simeq \frac{1 - K}{1 - RK + R(R - 1)K\mu} \\
&\simeq \frac{(1 - K)(1 + RK - R(R - 1)K\mu)}{1 + (R - 1)K - RK^2} \\
&\simeq \frac{-R(R - 1)K\mu + R(R - 1)K^2\mu}{1}.
\end{aligned} \tag{4}$$

Using equations 0, 1 and 4: $\frac{\Pr(\text{Alcase})}{\Pr(\text{Alcontrol})} \simeq R$.

Proof 2

We prove that the following is chi-square distributed:

$$\chi^2 = \frac{(Q_1 - Q_0)^2}{Q_1/N_1 + Q_0/N_0}$$

. We know the following quantity from the t test follows a normal distribution

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\text{var}(x_1)/N_1 + \text{var}(x_2)/N_2}}$$

. We assume group 1 is the case cohort and group 2 is the control cohort. We know that the poisson parameter is same as the mean and the variance of a poisson distribution. Therefore, $\bar{x}_1 = Q_1$ and $\bar{x}_2 = Q_0$; $\text{var}(x_1) = (1 - p)fQ_0 + RpfQ_0 = Q_1$ (assumes the two poisons in the mixture are independent) and $\text{var}(x_2) = Q_0$. χ^2 is simply z^2 .