

Text S1: the Wald test

The full logistic model can be written as

$$\text{logit}(\pi_{gh}) = \mu + \alpha_1 I(g=1) + \alpha_2 I(g=2) + \beta_1 I(h=1) + \beta_2 I(h=2) + \\ \lambda_{11} I(g=1)I(h=1) + \lambda_{12} I(g=1)I(h=2) + \lambda_{21} I(g=2)I(h=1) + \lambda_{22} I(g=2)I(h=2),$$

where π_{gh} is the probability of disease given genotype g at SNP1 and h at SNP2.

The log-likelihood is $l = \sum_{g=0}^2 \sum_{h=0}^2 [n_{gh1} \log(\pi_{gh}) + n_{gh0} \log(1 - \pi_{gh})]$. It is easy to verify that the

MLEs of π_{gh} are n_{gh1} / n_{gh+} , where $n_{gh+} = n_{gh0} + n_{gh1}$. Therefore, the MLEs of parameters

$\mu, \alpha_g, \beta_h, \lambda_{gh}$ are respectively

$$\hat{\mu} = \text{logit}(\hat{\pi}_{00}) = \log(n_{001} / n_{000})$$

$$\hat{\alpha}_g = \text{logit}(\hat{\pi}_{g0}) - \text{logit}(\hat{\pi}_{00}) = \log\left(\frac{n_{g01} / n_{g00}}{n_{001} / n_{000}}\right), g = 1, 2$$

$$\hat{\beta}_h = \text{logit}(\hat{\pi}_{0h}) - \text{logit}(\hat{\pi}_{00}) = \log\left(\frac{n_{0h1} / n_{0h0}}{n_{001} / n_{000}}\right), h = 1, 2$$

$$\hat{\lambda}_{gh} = \text{logit}(\hat{\pi}_{gh}) + \text{logit}(\hat{\pi}_{00}) - \text{logit}(\hat{\pi}_{g0}) - \text{logit}(\hat{\pi}_{0h}) \\ = \log\left(\frac{n_{gh1}n_{001} / n_{gh0}n_{000}}{n_{g01}n_{0h1} / n_{g00}n_{0h0}}\right), g = 1, 2, h = 1, 2$$

Let $a_{gh} = \frac{n_{gh0}n_{gh1}}{n_{gh0} + n_{gh1}}$, $a_{g+} = \sum_{h=0}^2 a_{gh}$, $a_{+h} = \sum_{g=0}^2 a_{gh}$, $a_{++} = \sum_{g=0}^2 \sum_{h=0}^2 a_{gh}$, where

$$g = 0, 1, 2; h = 0, 1, 2.$$

The observed Fisher Information matrix is

$$I = \begin{pmatrix} a_{++} & a_{1+} & a_{2+} & a_{+1} & a_{+2} & a_{11} & a_{12} & a_{21} & a_{22} \\ a_{1+} & a_{1+} & 0 & a_{11} & a_{12} & a_{11} & a_{12} & 0 & 0 \\ a_{2+} & 0 & a_{2+} & a_{21} & a_{22} & 0 & 0 & a_{21} & a_{22} \\ a_{+1} & a_{11} & a_{21} & a_{+1} & 0 & a_{11} & 0 & a_{21} & 0 \\ a_{+2} & a_{12} & a_{22} & 0 & a_{+2} & 0 & a_{12} & 0 & a_{22} \\ a_{11} & a_{11} & 0 & a_{11} & 0 & a_{11} & 0 & 0 & 0 \\ a_{12} & a_{12} & 0 & 0 & a_{12} & 0 & a_{12} & 0 & 0 \\ a_{21} & 0 & a_{21} & a_{21} & 0 & 0 & 0 & a_{21} & 0 \\ a_{22} & 0 & a_{22} & 0 & a_{22} & 0 & 0 & 0 & a_{22} \end{pmatrix}$$

By the asymptotic properties of MLEs, the distribution of the MLE vector $\hat{\theta}$ is approximately multivariate normal with variance-covariance matrix $(I)^{-1}$. Therefore, the joint distribution of the MLEs for the interaction parameters $\hat{\theta}_\lambda = (\hat{\lambda}_{11}, \hat{\lambda}_{12}, \hat{\lambda}_{21}, \hat{\lambda}_{22})^T$ is approximately multivariate normal with variance-covariance matrix being the elements in the last four rows and columns of $(I)^{-1}$. Using symbolic matrix inverse in Mathematica, we found that the covariance-covariance matrix is

$$\widehat{\text{cov}}(\hat{\theta}_\lambda) = \begin{pmatrix} \frac{1}{a_{00}} + \frac{1}{a_{01}} + \frac{1}{a_{10}} + \frac{1}{a_{11}} & \frac{1}{a_{00}} + \frac{1}{a_{10}} & \frac{1}{a_{00}} + \frac{1}{a_{01}} & \frac{1}{a_{00}} \\ \frac{1}{a_{00}} + \frac{1}{a_{10}} & \frac{1}{a_{00}} + \frac{1}{a_{10}} + \frac{1}{a_{02}} + \frac{1}{a_{12}} & \frac{1}{a_{00}} & \frac{1}{a_{00}} + \frac{1}{a_{02}} \\ \frac{1}{a_{00}} + \frac{1}{a_{01}} & \frac{1}{a_{00}} & \frac{1}{a_{00}} + \frac{1}{a_{01}} + \frac{1}{a_{20}} + \frac{1}{a_{21}} & \frac{1}{a_{00}} + \frac{1}{a_{20}} \\ \frac{1}{a_{00}} & \frac{1}{a_{00}} + \frac{1}{a_{02}} & \frac{1}{a_{00}} + \frac{1}{a_{20}} & \frac{1}{a_{00}} + \frac{1}{a_{02}} + \frac{1}{a_{20}} + \frac{1}{a_{22}} \end{pmatrix}$$

Note that the covariance matrix can also be derived using the multivariate normal approximation to the distribution of the observed counts.

The Wald test for testing $H_0 : \lambda_{11} = \lambda_{12} = \lambda_{21} = \lambda_{22} = 1$ is given by

$$Wald = \hat{\theta}_\lambda^T \left[\widehat{\text{cov}}(\hat{\theta}_\lambda) \right]^{-1} \hat{\theta}_\lambda$$

Under the null hypothesis of no gene-gene interaction, *Wald* follows the chi-squared distribution with 4 *df* asymptotically.