Snp HeRitability Estimation Kit (SHREK) Manual

Choi Shing Wan, Sam

`choishingwan@gmail.com`

May 27, 2015

Contents

Chapter 1

Background

Chapter 2

Installation

2.1   Dependencies

The programme is dependent on the ISO C++ 2011 standard and Eigen library.

2.2   Tutorial

Chapter 3

Input formats

3.1 Test statistic file

3.2 Linkage file

The linkage file should be in the binary ped format.

Chapter 4

Examples

Chapter 5

Assumptions and limitations

5.1 Assumptions

5.1.1 Calculation of variance

There are multiple assumptions when we calculate the variance. First, we assume a big sample size. Only when the sample size is big, can we approximate the distribution of the effects to follow the χ^2 distribution. Therefore, when the sample size is small, our variance estimate can be biased.

Secondly, for simplicity of implementation, when calculating the variance covariance matrix of the effects, we uses the **maximum sample size** as the n. So, for example, if 9 out of 10 Snps' association were conducted on 100 samples, yet the remaining 1 Snp were conducted on 10,000 samples, the programme will still use $n = 10000$ in the calculation of the variance covariance of **all** the Snp pairs.

5.2 Limitations

5.2.1 Direction of effect

If one doesn't provide the direction of effect, the programme will use the direction from the test-statistic. However, if only p-values are provided, all effect will be assumed to be positive. This will lead to a positive bias in the variance explained. Therefore it is advised to always provide the direction of effect.

5.2.2 Perfect LD

In the case where two or more Snps are in perfect LD, we cannot determine the true underlying heritability explained of each Snps. Therefore, in SHREK, Snps in perfect LD will be grouped together and formed one Snp where its effect is the mean of all Snps in the group. Decomposition will then be performed on this Snp and the resulting heritability will be **equally distributed** among the original group of Snps.

So for example: Assuming there are 5 Snps, $Snp_1, Snp_2, Snp_3, Snp_4 and Snp_5$ where Snp_1, Snp_2 and Snp_3 are all in perfect LD each with effect f_1, f_2 and f_3 respectively. Then we will first get Snp_μ with effect as $f_\mu = \frac{f_1 + f_2 + f_3}{3}$. As Snp_1, Snp_2 and Snp_3 has

the same LD structure with all other Snps (as they are in perfect LD), we then use the LD structure of Snp_1 as a representative of the LD of Snp_μ with all other Snps. This will then gives us

$$\begin{pmatrix} R_{11} & R_{14} & R_{15} \\ R_{14} & R_{44} & R_{45} \\ R_{15} & R_{45} & R_{55} \end{pmatrix} \begin{pmatrix} h_\mu \\ h_d \\ h_e \end{pmatrix} = \begin{pmatrix} f_\mu \\ f_d \\ f_e \end{pmatrix}$$

Finally, we will get

$$h_1 = \frac{h_\mu}{3}$$
$$h_2 = \frac{h_\mu}{3}$$
$$h_3 = \frac{h_\mu}{3}$$

It is important to note that for the variance, it is **NOT** equally divided among the Snps. Instead, it is something like

$$var(h_1) = var(h_\mu)$$
$$var(h_2) = var(h_\mu)$$
$$var(h_3) = var(h_\mu)$$