

Testing variance components in association mapping using Score/SKAT

Lian Lian

Michigan State University

November 12, 2015

Review of Association Mapping

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_i\beta_i + \mathbf{e}$$

- For each marker i , use t-test on β_i
- Low power for rare variants (allele frequency 1%-5%)
- Low power for markers with small effects

Sequential Kernel Association Test (SKAT)

Combined test for nearby markers

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \mathbf{e}$$

- $\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{I}_\tau)$
- $H_0 : \beta_i = 0 \text{ for all } i \Leftrightarrow \tau = 0$
- Only one test for each window

Sequential Kernel Association Test

Test $\tau = 0$

$$\mathbf{y} = X\boldsymbol{\alpha} + Z\boldsymbol{\beta} + \mathbf{e}, \boldsymbol{\beta} \sim N(0, I\tau)$$

- Likelihood ratio test

- Fit $H_0 : \mathbf{y} = X\boldsymbol{\alpha} + \mathbf{e} \rightarrow l_n(\theta_0)$
- Fit $H_A : \mathbf{y} = X\boldsymbol{\alpha} + Z\boldsymbol{\beta} + \mathbf{e} \rightarrow l_n(\hat{\theta}_n)$
- $2(l_n(\hat{\theta}_n) - l_n(\theta_0)) \sim \chi^2$

- SKAT

- Test based on score
- Only need to fit H_0

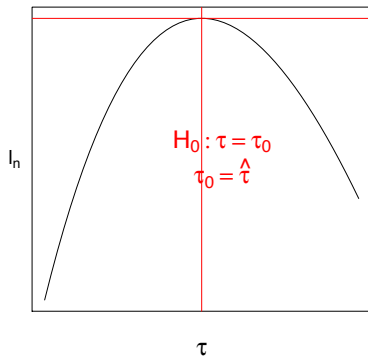
Score

Score: the derivative of log likelihood l_n at $\tau = \tau_0$

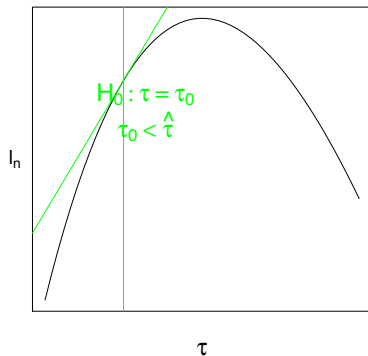
$$S = \left. \frac{\partial l_n}{\partial \tau} \right|_{\tau = \tau_0}$$

where l_n is the log likelihood from the data.

log likelihood from a data set



log likelihood from a data set



$$y = \mathbf{X}\alpha + \mathbf{Z}\beta + \mathbf{e}$$

- $y \sim N(\mathbf{X}\alpha, V)$
where $V = K\tau + I\sigma^2$, $K = \mathbf{Z}\mathbf{Z}'$
- log REML:

$$\begin{aligned} l_n &= -1/2 \log(|V|) - 1/2 \log(|\mathbf{X}'V^{-1}\mathbf{X}|) \\ &\quad - 1/2 (\mathbf{y} - \mathbf{X}\hat{\alpha})' V^{-1} (\mathbf{y} - \mathbf{X}\hat{\alpha}) \\ \hat{\alpha} &= (\mathbf{X}'V^{-1}\mathbf{X})^{-1} \mathbf{X}'V^{-1}\mathbf{y} \end{aligned}$$

$$\begin{aligned} &= \frac{\partial l_n}{\partial \tau} \\ &= 1/2(y - X\hat{\alpha})^T V^{-1} K V^{-1} (y - X\hat{\alpha}) - 1/2 \text{tr}(K P) \\ &P = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} \end{aligned}$$

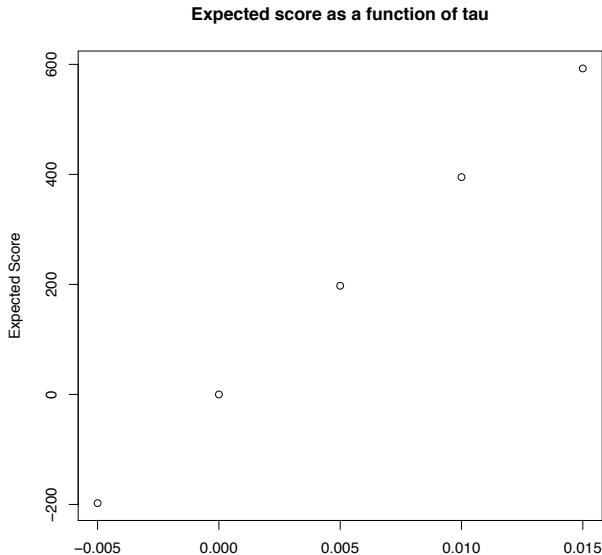
$$\begin{aligned} S &= \left. \frac{\partial l_n}{\partial \tau} \right|_{\tau=0} \\ &= 1/2(Y - X\hat{\alpha}_0)^T K (Y - X\hat{\alpha}_0) / \sigma^4 - 1/2 \text{tr}(K P_0) \\ &\hat{\alpha}_0 = (X^T X)^{-1} X^T y \\ &P_0 = (I - X(X^T X)^{-1} X^T) / \sigma^2 = (I - H_0) / \sigma^2 \\ &H_0 = X(X^T X)^{-1} X^T \end{aligned}$$

Relationship between Score and parameter of interest τ

$$\begin{aligned}E(S) &= 1/2\text{tr}[KE[(Y - X\hat{\alpha}_0)(Y - X\hat{\alpha}_0)^T]/\sigma^4 - 1/2\text{tr}[K(I - H_0)]/\sigma^2 \\&= 1/2\text{tr}[K(I - H_0)V(I - H_0)]/\sigma^4 - 1/2\text{tr}[K(I - H_0)]/\sigma^2 \\&= 1/2\text{tr}[K(I - H_0)(K\tau + I\sigma^2)(I - H_0)]/\sigma^4 - 1/2\text{tr}[K(I - H_0)]/\sigma^2 \\&= 1/2\text{tr}[K(I - H_0)K\tau(I - H_0)]/\sigma^4 \\&\quad + 1/2\text{tr}[K(I - H_0)(I - H_0)I\sigma^2]/\sigma^4 \\&\quad - 1/2\text{tr}[K(I - H_0)\sigma^2]/\sigma^4 \\&= 1/2\text{tr}[K(I - H_0)K(I - H_0)\tau]/\sigma^4 \\&= 1/2\text{tr}\{[K(I - H_0)]^2\}\tau/\sigma^4\end{aligned}$$

It is a linear function in τ ! The higher the τ , the higher the expected score!

Score as a function of τ



Replace σ^2 with the estimated value $\hat{\sigma}^2$

$$S = \left. \frac{\partial l_n}{\partial \tau} \right|_{\tau=0} = 1/2(Y - X\hat{\alpha}_0)^T K(Y - X\hat{\alpha}_0)/\hat{\sigma}^4 - 1/2\text{tr}(KP_0)$$

$$S \sim N(0, 1/2\text{tr}(KP_0KP_0))$$

SKAT

[Wu et al., 2011] Defined Q and as:

$$Q = 1/2(Y - X\hat{\alpha}_0)^T K(Y - X\hat{\alpha}_0)/\hat{\sigma}^2$$

Q increases as τ increases.

$$Q \sim \sum \lambda_i \chi_{1,i}^2$$

Extend SKAT to more random components

Original SKAT test

- $H_0: F$
- $H_1: F + R_1$

Models with more than one random effect

- One random component in H_0
 - $H_0: F + R_1$
 - $H_1: F + R_1 + R_2$
 - Examples: common + rare variants, background + window
- Two random effects in H_0
 - $H_0: F + R_1 + R_2$
 - $H_1: F + R_1 + R_2 + R_3$
 - Examples: $G \times E: G + M + E + M \times E$
- A more general case: any number of random effect in the NULL model

Models with more than one random effect

$$y = \mu + \mathbf{X}\boldsymbol{\alpha} + \sum \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}$$

- $E(\mathbf{b}_i) = \mathbf{0}$, $V(\mathbf{b}_i) = I\tau_i$
 - $H_0 : \mathbf{b}_i = \mathbf{0}$, $\Leftrightarrow \tau_i = 0$
 - SKAT/Score test?

Score with REML for τ_i

$$S_i = \left. \frac{\partial l_n}{\partial \tau_i} \right|_{\tau_i=0} = 1/2(Y - X\hat{\alpha}_{-i})^T V_{-i}^{-1} K_i V_{-i}^{-1} (Y - X\hat{\alpha}_{-i}) - 1/2 \text{tr}(K_i P_{-i})$$

$$V_{-i} = \sum_{j \neq i} K_j \tau_j + I \sigma_e^2$$

Define Q and as:

$$Q = 1/2(Y - X\hat{\alpha}_{-i})^T V_{-i}^{-1} K_i V_{-i}^{-1} (Y - X\hat{\alpha}_{-i})$$

Relationship between Score and τ_i

$$\begin{aligned} E(S_i) &= 1/2 \text{tr}[V_{-i}^{-1} K_i V_{-i}^{-1} E[(Y - X\hat{\alpha}_{-i})(Y - X\hat{\alpha}_{-i})^T] \\ &\quad - 1/2 \text{tr}[K_i V_{-i}^{-1} (I - H_{-i})] \\ &= 1/2 \text{tr}[V_{-i}^{-1} K_i V_{-i}^{-1} (I - H_{-i}) V (I - H_{-i})^T] \\ &\quad - 1/2 \text{tr}[K_i V_{-i}^{-1} (I - H_{-i})] \\ &= 1/2 \text{tr}[V_{-i}^{-1} K_i V_{-i}^{-1} (I - H_{-i}) (V_{-i} + K_i \tau_i) (I - H_{-i})^T] \\ &\quad - 1/2 \text{tr}[K_i V_{-i}^{-1} (I - H_{-i})] \\ &= 1/2 \text{tr}[V_{-i}^{-1} K_i V_{-i}^{-1} (I - H_{-i}) (K_i \tau_i) (I - H_{-i})^T] \\ &= 1/2 \text{tr}\{[V_{-i}^{-1} (I - H_{-i}) K_i]^T V_{-i}^{-1} (I - H_{-i}) K_i\} \tau_i \end{aligned}$$

$$H_{-i} = X(X^T V_{-i}^{-1} X)^{-1} X^T V_{-i}^{-1}$$

$E(S_i)$ is a linear function in τ_i !

The higher the τ_i , the higher the expected score!

$$\begin{aligned} Q &= (Y - X\hat{\alpha}_{-i})^T V_{-i}^{-1} K_i V_{-i}^{-1} (y - X\hat{\alpha}_{-i}) \\ &\sim \sum \lambda_i \chi_{1,i}^2 \end{aligned}$$

- λ_i from $\text{eigen}(Z_i^T V_{-i}^{-1} (I - H_{-i}) Z_i)$

For comparison, we also included Score test. The distribution of Score is:

$$S \sim N(0, 1/2 \text{tr}(K_i P_{-i} K_i P_{-i}))$$

Implemented methods

- Interface for association mapping (GWAS)
- Testing a matrix or variable as fixed or random in standard mixed model setting (fitNULL, testZ, testX)

Mouse data set

- `mouse.X(1825 × 10346)`: Marker genotypes
- `mouse.pheno(1825 × 17)`: data.frame for phenotypes
- `mouse.G(1825 × 1825)`: Genomic relationship matrix
- `mouse.eigenG($U1,$d1)`: Eigen decomposition for G matrix

Example 1: GWAS by SKAT/Score: no random component in the NULL model

GWAS function from SKAT2

```
GWAS(formula=Obesity.BMI~GENDER,  
GxE.formula=~(1|mouse.X[,1:100]),  
data=mouse.pheno, methods="SKAT")
```

SKAT function from the original SKAT package

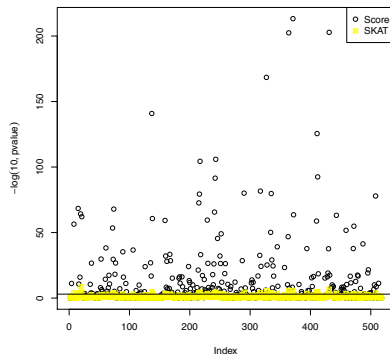
```
obj<-SKAT_Null_Model(Obesity.BMI~GENDER,  
out_type="C", data=mouse.pheno)  
p=matrix(0,5,1)  
for(i in 1:5){  
  Zi=mouse.X[, (i-1)*20+1:20]  
  p[i,]=SKAT(Zi, obj, weights=rep(1,ncol(Zi)),  
    is_check_genotype=F)$p.value  }
```

Results from Example 1

	SKAT2	SKAT
1	0.36	0.36
2	0.48	0.48
3	0.28	0.28
4	0.01	0.01
5	0.29	0.29

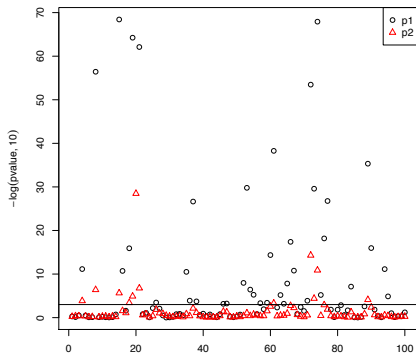
Example 1 continued: Score vs SKAT

```
p1=GWAS(formula=Obesity.BMI~GENDER,  
GxE.formula=~(1|mouse.X),  
data=mouse.pheno,methods=c("Score","SKAT"))$p.value
```



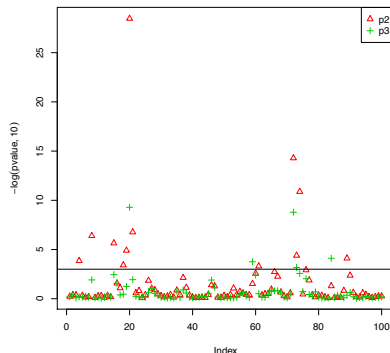
Example 2: GWAS by Score: fit cage as random effect in the in the NULL model.

```
pheno$cage=as.factor(pheno$cage)
p2=GWAS(formula=Obesity.BMI~GENDER+(1|cage),GxE.formula
=~(1|mouse.X[,1:2000]),data=mouse.pheno)$p.value
```



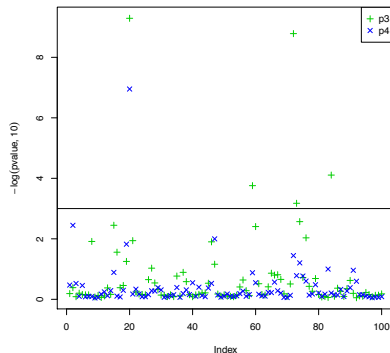
Example 3: GWAS by Score: fit Genomic background in the NULL model

```
p3=GWAS(formula=Obesity.BMI~GENDER  
+.eigenG(mouse.eigenG),GxE.formula  
=~(1|mouse.X[,1:2000]),data=mouse.pheno)$p.value
```



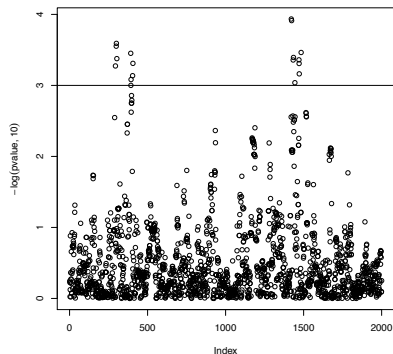
Example 4: GWAS by Score: test GxE

```
p4=GWAS(formula=Obesity.BMI~GENDER+.eigenG(mouse.eigenG),  
GxE.formula=~(1|mouse.X[,1:2000])+  
(1|mouse.X[,1:2000]:GENDER),  
data=pheno)$p.value
```



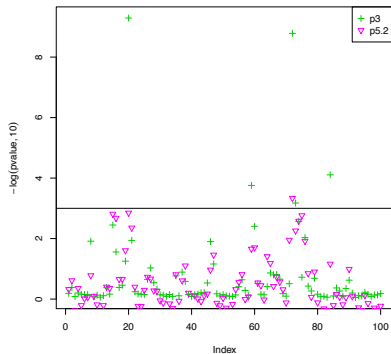
Example 5: GWAS Single SNP markers as fixed effects

```
p5=GWAS(formula=Obesity.BMI~GENDER+.eigenG(mouse.eigenG),  
GxE.formula=~mouse.X[,1:2000],data=pheno, setsize=1)$p.value
```



Example 5 continue: min p-value in a window of Single SNP tests

```
p5.2=GWAS(formula=Obesity.BMI~GENDER  
+.eigenG(mouse.eigenG),  
GxE.formula=~mouse.X[,1:2000],data=pheno)$p.value
```



Thank You



Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011).
Rare-variant association testing for sequencing data with the sequence
kernel association test.
The American Journal of Human Genetics, 89(1):82–93.