Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test (Michael C. Wu et al., 2011)

Xu (Rita) Shi

Presentation 1: Introduction, Motivation and Overview

November 5, 2013

Outline

Challenge in rare-variant association test

Limitations for previous methods

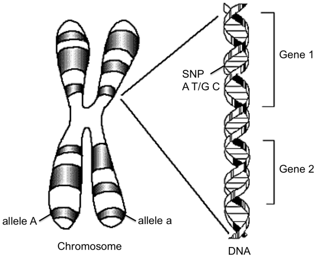
- Burden test
- C-alpha test

Overview of Sequence Kernel Association Test (SKAT)

- Flexible
- Computationally efficient

Bonus example

Background Knowledge



- Minor Allele Frequency (MAF)

  Frequency at which least common allele occurs in a given population

- Rare variant

  Variant with a MAF less than 1-5% (in SKAT: 3%)

- Genome Wide Association Study (GWAS)

  An examination of many **common** genetic variants in different individuals to see if any variant is associated with a trait

Challenge in rare-variant association test

Subject	V1	V2	V3	Disease
1	1	1	0	1
2	0	0	1	1
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0

- Multiple rare variants within the same functional unit e.g. exon of a gene — unexplained genetic component of complex traits (Missing heritability)
- Traditional association test of common variants: underpowered unless sample sizes or effect sizes are very large
- Need to collectively analyze instead of individually

Limitations for previous methods

- Burden test: summarize/collapse rare variants in a region

  $I_{ij} = 0,\ 1$ or 2 – number of minor alleles at variant j for individual i
  r_i – number of variants that carry at least one copy of the minor allele
  n_i – total number of rare variants, $r_i = \sum_{j=1}^{n_i} 1(I_{ij} > 0)$

  Count-Based Proportion: $E(Y_i) = \beta_0 + \lambda \frac{r_i}{n_i} + \beta X_i$

  Dichotomize (Cohort Allelic Sum Test, CAST):
  $E(Y_i) = \beta_0 + \lambda 1(r_i > 0) + \beta X_i$

  Weighted Sum Test (WST): $logit\ P(Y_i = 1) = \beta_0 + r_i^* + \beta X_i + \epsilon_i$,
  where $r_i^* = \sum_{j=1}^{n_i} \frac{I_{ij}}{w_j}$

- Limitation: assume that all rare variants influence the phenotype in the same direction and with the same magnitude

Limitations for previous methods

- C-alpha test:

  Compares the expected variance to the actual variance of the distribution of allele frequencies

- Limitation:

  Only apply to case-control data

  Cannot adjust for covariate (population stratification)

  Need permutation sometimes (computationally expensive)

Overview of Sequence Kernel Association Test (SKAT)

- Model: $logit\ P(Y_i = 1) = \beta_0 + \sum_{j=1}^{p} b_j G_{ij} + \beta X_i$
  or $E(Y_i) = \beta_0 + \sum_{j=1}^{p} b_j G_{ij} + \beta X_i$

  i: individual, i = 1, ..., n

  j: p genetic variants within a functional region – common and rare

  $X_i = (X_{i1}, X_{i2}, ..., X_{im})$: covariates, e.g. age, gender

  Y_i: phenotype (dichotomous or continuous)

  $G_i = (G_{i1}, G_{i2}, ..., G_{ip})$: genotype, $G_{ij} = 0$, 1 or 2

- Assumption: $b_j \sim (0, (w_j\sigma)^2)$

  $H_0 : \boldsymbol{b} = 0 \Leftrightarrow H_0 : \sigma^2 = 0$ (variance-component test)

- Allows for different directions and magnitudes of genetic effects

Overview of Sequence Kernel Association Test (SKAT)

- Variance-component score statistic:

$$
\begin{aligned}
Q &= (y-\hat{y}_0)'K(y-\hat{y}_0), \hat{y}_0\text{: fitted value under } H_0 \\
&= ||WG'(y-\hat{y}_0)||^2 \\
&\text{kernel } K = GWWG', \text{weight } W = diag(w_j) \\
&= \sum_{j=1}^{p} w_j^2 \;\; ||\boldsymbol{G_j}(\boldsymbol{y}-\boldsymbol{\hat{y}_0})||^2 \\
&= \sum_{j=1}^{p} w_j^2 \;\; \sum_{i=1}^{n} [G_{ij}(y_i-\hat{y}_0)]^2 \\
&\sim \text{mixture of } \chi_1^2
\end{aligned}
$$

Weighted sum of individual score statistic $\boldsymbol{S_j} = \boldsymbol{G_j}(\boldsymbol{y}-\boldsymbol{\hat{y}_0})$

- Only requires fitting the null model

Overview of Sequence Kernel Association Test (SKAT)

- SKAT: weigited sum of score statistic

$logit\ P(Y_i = 1) = \beta_0 + \sum_{j=1}^{p} \boldsymbol{b_j} G_{ij} + \beta X_i$

$$Q_j = \sum_{j=1}^{p} w_j^2 \sum_{i=1}^{n} [G_{ij}(y_i - \hat{y_0})]^2$$

- Burden: weighted sum of genetic variants:

$logit\ P(Y_i = 1) = \beta_0 + \sum_{j=1}^{p} \boldsymbol{w_j} \beta_{\boldsymbol{B}} G_{ij} + \beta X_i$

$$Q = [\sum_{i=1}^{n} (y_i - \hat{y_0})\ (\sum_{j=1}^{p} w_j G_{ij})\ ]^2$$

Overview of Sequence Kernel Association Test (SKAT)

- Choice of weight: w_j

  $Beta((MAF_j); 1, 25)$

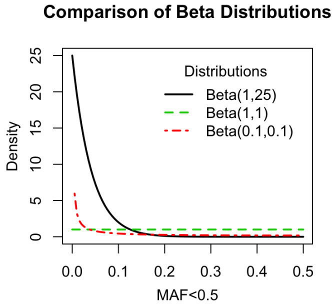
  Upweight rare-variant
  downweight common-variant

  $Beta((MAF_j); 1, 1)$

  Uniform[0, 1]

  $Beta((MAF_j; 0.1, 0.1)$

  steeper than $Beta(1, 25)$ at (0, 0.01)



Overview of Sequence Kernel Association Test (SKAT)

- Choice of kernel function: $(K(G_i, G_{i'}))_{n\times n}$ positive semidefinite

  Weighted linear kernel function $K(G_i, G_{i'}) = \sum_{j=1}^{p} w_j^2 G_{ij} G_{i'j}$

  Linear genetic effects

  Weighted quadratic kernel $K(G_i, G_{i'}) = (1 + \sum_{j=1}^{p} w_j G_{ij} G_{i'j})^2$

  Both linear and quadratic genetic effects

  $logit\ P(Y_i = 1) = \beta_0 + f(G_i) + \beta X_i,\ H_0 : f(G) = 0$

  Weighted IBS kernel $K(G_i, G_{i'}) = \sum_{j=1}^{p} w_j IBS(G_{ij}, G_{i'j})$

  Identity by state (IBS), number of alleles that share IBS

  Free of assumption of additivity, allows for interaction between variants

Sequence Kernel Association Test (SKAT)

- Test for association between phenotype and a collection of rare and common variants in sequencing-based association studies
- Robust to direction and magnitude
- Allow for covariate adjustment
  Works for both continuous and dichotomous phenotype
- Only need to fit null model
  No permutation needed for p-value
- f(G): allow for epistatistic effects (interaction between genetic variants); family data
  Y: regression based, easily extended to survival, longitudinal and multivariate phenotypes

Next Steps

- Estimation of power and sample size
- Simulations; compare to previous methods
- Application to Dallas Heart Study Data? (still waiting for response)
- Math in appendix A

Thank you!

References

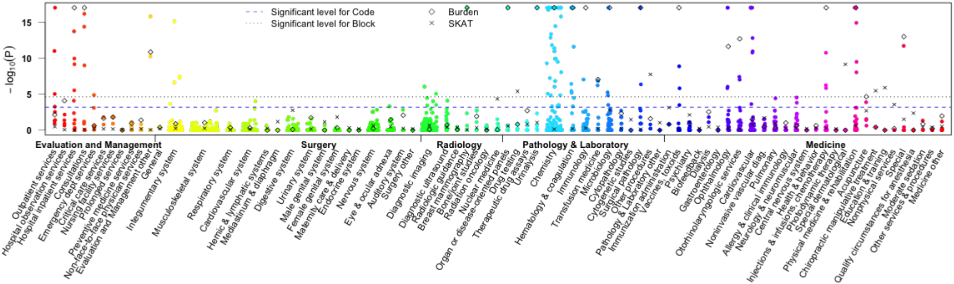
- Wu, M. et. al. Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*, 92: 841-853, 2013.

- Morris,A.P., and Zeggini, E. Anevaluationofstatistical approaches to rare variant analysis in genetic association studies. *Genetic Epidemiology*, 34: 188-193, 2010.

- Li, B., and Leal, S.M. Methods for detecting associa- tions with rare variants for common diseases: application to analysis of sequence data. *American Journal of Human Genetics*, 83: 311-321, 2008.

- Madsen, B.E., and Browning, S.R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genetics*, 2009 Feb; 5(2):e1000384.

- Neale, B.M. et al. Testing for an unusual distribution of rare variants. *PLoS Genetics*, 2011 Mar;7(3):e1001322.



Significant level for Code
Significant level for Block
Burden
SKAT
−log10(P)
0
5
10
15
Evaluation and Management
Surgery
Radiology
Pathology & Laboratory
Medicine
Outpatient services
Hospital observation services
Hospital inpatient services
Consultations
Emergency dept services
Critical care services
Nursing facility services
Prolonged services
Preventive medicine services
Non-face-to-face physician services
Evaluation and Management other
General
Integumentary system
Musculoskeletal system
Respiratory system
Cardiovascular system
Hemic & lymphatic systems
Mediastinum & diaphragm
Digestive system
Urinary system
Male genital system
Female genital system
Maternity care & delivery
Endocrine system
Nervous system
Eye & ocular adnexa
Auditory system
Surgery other
Diagnostic imaging
Diagnostic ultrasound
Radiologic guidance
Breast mammography
Bone/joint studies
Radiation oncology
Nuclear medicine
Organ or disease-oriented panels
Drug testing
Therapeutic drug assays
Urinalysis
Chemistry
Hematology & coagulation
Immunology
Transfusion medicine
Microbiology
Cytopathology
Cytogenetic studies
Surgical pathology
Other procedures
Pathology & Laboratory other
Immunization administration
Vaccines, toxoids
Psychiatry
Biofeedback
Dialysis
Gastroenterology
Ophthalmology
Otorhinolaryngologic services
Cardiovascular
Noninvasive vascular diag.
Pulmonary
Allergy & clinical immunology
Neurology & neuromuscular
Central nervous system
Health & behavior
Injections & infusions, chemotherapy
Photodynamic therapy
Special dermatological
Physical medicine & rehabilitation
Acupuncture
Chiropractic manipulative treatment
Education & training
Nonphysical services
Special
Qualify circumstances for anesthesia
Moderate sedation
Other services & procedures
Medicine other