Variance components QTL linkage analysis in large pedigrees

Yurii Aulchenko

February 23, 2006, am

Overview

- Variance components (VC) model
- Power of VC linkage analysis
- Power of different study designs
- Analysis of dependent traits
- Bivariate VC analysis

Expected trait value

$$E[x_i] = \mu + \sum_k \beta_k c_{ik}$$

x_i : trait value for i-th individual

c_{ik} : the value of k-th covariate for i-th individual

μ : population mean

β_k : coefficient of regression of x onto covariate c_k

Variance-covariance structure

$$Cov(x_i, x_j) = \begin{cases} \sigma_A^2 + \sigma_G^2 + \sigma_e^2 & \text{if } i = j \\ \pi_{ij}\sigma_A^2 + \phi_{ij}\sigma_G^2 & \text{if } i \neq j \end{cases}$$

σ_A^2 : variance due to QTL

σ_G^2 : residual polygenic variance

σ_e^2 : environmental variance

π_{ij} : proportion of alleles IBD between i and j at the QTL location (estimated using marker information on pedigree)

ϕ_{ij} : relationship coefficient between i and j (estimated using pedigree structure information)

Likelihood

$$log_e L = -\frac{1}{2}\left[log_e|V| - (\vec{x} - E[\vec{x}])' V^{-1} (\vec{x} - E[\vec{x}])\right]$$

where V is the variance-covariance matrix and $\vec{x}$ is the vector of that values

Parametrisation

When no covariates are included into analysis, four parameters are to be estimated:

$$\vec{\theta} = \{\mu, \sigma^2_A, \sigma^2_G, \sigma^2_e\}$$

Alternative parametrisation is:

$$\vec{\theta} = \left\{\mu, H^2_q = \left(\frac{\sigma^2_A}{\sigma^2_A + \sigma^2_G + \sigma^2_e}\right), H^2_r = \left(\frac{\sigma^2_G}{\sigma^2_A + \sigma^2_G + \sigma^2_e}\right), \sigma^2_e\right\}$$

H^2_q is QTL heritability (proportion of variance explained by the QTL)

H^2_r is residual heritability (proportion of variance explained by other genes)

Overview

- Variance components (VC) model
- Power of VC linkage analysis
- Power of different study designs
- Analysis of dependent traits
- Bivariate VC analysis

Power estimation for VC LA

Given model parameters ($\vec{\theta}$) and pedigree structure (with indication of pheno- and geno-typed individuals), the Non-Centrality Parameter (**NCP**) can be computed for the sample; **NCP** is basically the expected value of χ^2 statistics. If **NCP** is known, the

- Expected LOD score

$$\mathbf{ELOD} = \frac{\mathbf{NCP}}{2 \cdot log_e(10)}$$

- Power to reach some pre-defined LOD threshold (**Tlod**)

$$P(\mathbf{ELOD} \geq \mathbf{Tlod}) = 1 - \Phi_{\mu=\sqrt{\mathbf{NCP}}, \sigma^2=1}\left(\sqrt{2 \cdot \mathbf{Tlod} \cdot log_e(10)}\right)$$

Examples

NCP = 13.82

- $ELOD = \frac{13.82}{4.61} = 3.0$
- $P(LOD \geq 1.0) = 1 - \Phi_{\mu=\sqrt{13.82}, \sigma^2=1}\left(\sqrt{4.61}\right) = 1 - 0.06 = 0.94$
- $P(LOD \geq 3.0) = 1 - \Phi_{\mu=\sqrt{13.82}, \sigma^2=1}\left(\sqrt{13.82}\right) = 1 - 0.5 = 0.5$

NCP = 8.0

- $ELOD = \frac{8}{4.61} = 1.74$
- $P(LOD \geq 1.0) = 1 - \Phi_{\mu=\sqrt{8}, \sigma^2=1}\left(\sqrt{4.61}\right) = 1 - 0.25 = 0.75$
- $P(LOD \geq 3.0) = 1 - \Phi_{\mu=\sqrt{8}, \sigma^2=1}\left(\sqrt{13.82}\right) = 1 - 0.81 = 0.19$

NCP parameter of a pedigree

$$\text{NCP} \approx \sum_{i>j} \lambda_{ij} = \sum_{i>j} \frac{(1 + r_{ij}^2)}{(1 - r_{ij}^2)^2} \left(Var[\phi_{ij}] \cdot \left(H_q^2\right)^2 \right)$$

where the sum is taken over all pairs of pheno- and genotyped individuals i, j; λ_{ij} is thus per-pair contribution to the **NCP**

- r_{ij} is the average correlation between phenotypes of relative pair i and j; $r_{ij} = \phi_{ij} \cdot (H_r^2 + H_q^2)$
- $Var[\phi_{ij}]$ is the variance of the relationship coefficient, which is easily computed. It is equal to 1/8, 1/16 and 3/64 for sib, ancle-nephew (grandparent-avuncular) and first-cousin pairs, respectively

Limitations of the approximation

This approximation tends to over-estimate **NCP** when many remotedly related people are present in a pedigree and when the QTL effect is large. The approximation works quite well when

- $H_q^2 \leq 0.1$
- $(H_q^2 + H_r^2) > 0.2$
- There are not too many remotedly related people in the pedigree
- Pedigree size is below 30-40

Overview

- Variance components (VC) model
- Power of VC linkage analysis
- Power of different study designs
- Analysis of dependent traits
- Bivariate VC analysis

Power of a study design

Consider a study with mondey enough to phenotype and genotype 6,000 individuals. These, however, may be sampled as sib-pairs, or as larger nuclear families or as extended pedigrees. Which study design to prefer?

Let us assume that QTL heritability, H_q^2 is 10% and H_r^2 is 20%; What will be the power of different designs to reach LOD $\geq$ 3?

6,000 people in sib-pairs

- 3,000 pairs of sibs in total
- Per-pair contribution to **NCP**
  - $r_{sibs} = \frac{H_r^2+H_q^2}{2} = 0.15$
  - $\frac{(1+r_{ij}^2)}{(1-r_{ij}^2)^2} = 1.07$
  - $\frac{(1+r_{ij}^2)}{(1-r_{ij}^2)^2} \cdot Var[\phi_{ij}] \cdot \left(H_q^2\right)^2 = 1.07 \cdot \frac{1}{8} \cdot (0.1)^2 = 0.0013$
- Total expected **NCP** is $3000 \cdot 0.0013 = 4.01$
- Power to reach LOD $\geq$ 3 is 4%

6,000 people in pedigrees with 5 sibs

- 1,200 families each contributing 10 pairs (12,000 pairs of sibs in total)
- Total expected **NCP** is $12,000 \cdot 0.0013 = 16.05$
- Power to reach LOD $\geq$ 3 is 61%

6,000 people, pedigrees with 5 sibs are paired and sibs are cousing

Conclusion

Study of larger pedigrees, theoretically, leads to high power.

However, large pedigrees are hard to sample up until recent.

Sources of large pedigrees:

- Isolated populations : study can be done by comparatively small team
- Outbred populations : possible only within large nation-wide programs

Overview

- Variance components (VC) model
- Power of VC linkage analysis
- Power of different study designs
- Analysis of dependent traits
- Bivariate VC analysis

Dependence between two traits

Consider following model

Q : trait of interest

G : gene which is affecting Q

C : some other trait or factor which exibits relation with Q

There may be different reasons for relation between Q and C.

Few possible models explaining dependence

1. C is ”endophenotype”, which lies between gene **G** and trait Q: $\mathbf{G} \Rightarrow C \Rightarrow Q$
2. C is external factor: $\mathbf{G} \Rightarrow Q \Leftarrow C$
3. C modifies the effect of **G**: $\mathbf{G} \times C \Rightarrow Q$
4. **G** controls both Q and C: $Q \Leftarrow \mathbf{G} \Rightarrow C$

How to analyse these traits?

1. C is ”endophenotype”: do not analyse Q, analyse C
2. C is external factor: include C as covariate into analysis
3. C modifies the effect of **G**: stratify by C (or better perform multivariate analysis)
4. **G** controls both Q and C: analyse Q and C

Overview

- Variance components (VC) model
- Power of VC linkage analysis
- Power of different study designs
- Analysis of dependent traits
- Bivariate VC analysis

Bivariate VC analysis model

Information about two traits is analysed simultaneously. The model similar to uni-variate analysis: two sets of parameters are introduced for each trait. Additionaly, however, the correlation between genetic components of the two traits (ρ_r, ρ_q) and correlation between their enviromental components (ρ_e) are estimated.

$$\vec{\theta} = \{\mu_1, \mu_2, H^2_{q1}, H^2_{q2}, H^2_{r1}, H^2_{r1}, \sigma^2_{e1}, \sigma^2_{e2}, \rho_q, \rho_r, \rho_e\}$$

The correlations between components say how much these overlap for the two traits: e.g. $\rho_r = 1$ says that essentially the same genes control the trait

Analysis by groups (gender, age)

Bivariate analysis may be used to analyse genetic determinants of the same trait in different groups (usually these are gender or age-groups). Then the trait is represented as two new ”traits”, one for each group. As for any given person only one ”trait” value is present, environmental correlation (ρ_e) is absent from the model

Exercise a.m.

Exercise 1 till the beginning of page 5 (before "Now we will start linkage analysis of the trait qt")

- Login into other server
- Description of the data using pedigree drawings and R
- Uni- and bi-variate heritability analysis using SOLAR

PS "bit-size" determines comutational complexity of a pedigree and is computed as

$$2 \cdot number\_of\_descendants - number\_of\_founders$$