

BIOSTAT 572: Update Talk

Elisa Sheng

Biostatistics Department, University of Washington

May 1, 2012

Recap of Intro Talk

Z = phenotype

g_i = genotype at SNP _{i} (0,1, or 2)

ϵ = environment

\dots = gene interactions (and gene-environment interactions)

$$Z = \alpha + \sum_i \beta_i f_i(g_i) + \epsilon + \dots$$

Assumptions: Part I

1. main effects are additive ($f_i(g_i) = g_i$)
2. SNPs are in linkage equilibrium (g_i are uncorrelated)
3. the two alleles at each SNP are independent ($g_i \sim \text{Binomial}(2, \pi_i)$ for all i)
4. the phenotype has been normalized ($Z \sim (0, 1)$)

So,

$$Z = \alpha + \sum_i \beta_i g_i + \epsilon + \dots$$

$$h^2 = \sum_i 2\beta_i^2 \pi_i (1 - \pi_i)$$

What is h_{all}^2 ?

$$h_{known}^2 = \sum_{i \in \{\text{known SNPs}\}} 2\beta_i^2 \pi_i (1 - \pi_i)$$

$$h_{unknown}^2 = \sum_{i \in \{\text{unknown SNPs}\}} 2\beta_i^2 \pi_i (1 - \pi_i)$$

$$h_{all}^2 = h_{known}^2 + h_{unknown}^2$$

The Brave New Method

The Theorem: $h_{all}^2 = (1 - \kappa_0)\rho'(\kappa_0)$

- $\kappa_{1,2}$ is the *kinship coefficient*, the proportion of genome shared by individuals 1 and 2
- κ_0 is the average proportion of genome-sharing in the population
- $\rho(k) = E(Z_i Z_j | \kappa_{i,j} = k)$, the average phenotypic correlation between individuals who share k of their genome
- $\rho'(\kappa_0)$ is the 'slope' of ρ at κ_0

Assumptions: Part II

Let $\kappa_{1,2}^i$ be an indicator variable for genome sharing at SNP i

$$\kappa_{1,2} = \frac{1}{n} \sum_{i=1}^n \kappa_{1,2}^i$$

5. rate of genome sharing is constant over the genome
6. population is unstructured

$$E(\kappa_{1,2}) = \kappa_0 \text{ for all pairs of individuals}$$

$$E(\kappa_{1,2}^i = 1 | \kappa_{1,2} = k) = E(\kappa_{1,2}^i = 1)$$

Assumptions: Part III

Let H_{1i}, H_{2i} be one allele at SNP_{*i*} from individuals 1 and 2

7. H_{1i}, H_{2i} are conditionally independent when $\kappa_{1,2} = \kappa_0$

$$E(H_{1i}, H_{2i} | \kappa_{1,2} = \kappa_0) = E(H_{1i})E(H_{2i})$$

Note: This property can be derived using assumptions 5-6

Phenotypic Correlation

$$\begin{aligned} & E(Z_1 Z_2 | \kappa_{1,2} = k) \\ &= E\left(\left(\alpha + \sum_i \beta_i g_{1i} + \epsilon_1\right)\left(\alpha + \sum_i \beta_i g_{2i} + \epsilon_2\right) \mid \kappa_{1,2} = k\right) \\ &= -\alpha^2 + \sum_i \beta_i E(g_{1i} g_{2i}) \end{aligned}$$

The key ingredients are:

- no gene-gene or gene-environment interactions (we ignore them for narrow sense heritability)
- Z is normalized implies $\alpha = -\sum_i \beta_i E(g_{1i})$
- $\epsilon \sim iid(0, \sigma_\epsilon^2)$

Phenotypic Correlation

$$\begin{aligned} & E(g_{1i}g_{2i}) \\ &= 4 \left(k\pi_i + \frac{(\pi_i - \pi_i\kappa_0)(1 - k)}{1 - \kappa_0} \right) \\ &= k \frac{\pi_i(1 - \pi_i)}{1 - \kappa_0} + \frac{\pi_i(\pi_i - \kappa_0)}{1 - \kappa_0} \end{aligned}$$

The key ingredients here are assumption 5-7

Phenotypic Correlation

$$\begin{aligned} & E(Z_1 Z_2 | \kappa_{1,2} = k) \\ &= -\alpha^2 + \sum_i \beta_i E(g_{1i} g_{2i}) \\ &= -\alpha^2 + \sum_i \beta_i \frac{\pi_i(\pi_i - \kappa_0)}{1 - \kappa_0} + \frac{k}{2(1 - \kappa_0)} \sum_i 2\beta_i \pi_i (1 - \pi_i) \\ &= \textit{intercept} + \frac{h_{all}^2}{2(1 - \kappa_0)} k \end{aligned}$$

So

$$h_{all}^2 = 2(1 - \kappa_0) \times \textit{slope}$$

Next Steps

- Simulations: hopefully get the simulated data from the authors
- Maybe use a real dataset to apply the method
- Explain relationships between assumptions and corresponding properties (e.g. conditional/unconditional independence)
- Show assumption 7 can be derived from assumptions 5-6