PBG 200A Notes

Sam Fleischer

November 7, 2016

1 Questions from Ch. 3 notes pgs 42-50

- Are we assuming the mean of the population is $\mu = 0$? More specifically, are we assuming nobody in the population has the gene?
- In figure 17, why don't we see only positive phenotypes?
- Question 1A:

$$\text{cov[half-sibs]} = 0.25 = 2F_{\text{half-sibs}}V_A \qquad F_{\text{half-sibs}} = \frac{1}{8} \quad \Longrightarrow \quad V_A = 1 \quad \Longrightarrow \quad h^2 = \frac{V_A}{V} = \frac{1}{4}$$

- Question 1B: ?
- Who's Galton?

2 End of the Last Leture

- Generally $\frac{D_N}{D_S} < 1$, which is consistent with the idea that most fixed mutations are synonymous

3 Inconsistencies with Neural Theory

- Some data suggests that there are fewer nonsynonymous polymorphisms w.r.t synonymous poymorphisms than there are nonsynonymous fixed alleles w.r.t synonymous fixed alleles.
- Set α equal to the number of mutations due to selection, and so $1 - \alpha$ is the number of mutations due to drift.

$$(1-\alpha)D_N = t_{\text{Div}}\mu_N$$
$$\alpha = 1 - \frac{D_S P_N}{D_N P_S}$$

4 Incomplete Lineage Sorting

- There are discrepancies between gene trees and the species tree. This means geneticists must use multiple gene trees in order to determine phylogeny.
- This happens if (for example) humans and chimps don't coalesce in the human/chimp common ancestor population
- $$\mathbb{P}[\text{fail to coalesce in the human/chimp common ancestor population}] = \left(1 - \frac{1}{2N_A}\right)^t$$

  where t is measured in generations. So

$$\mathbb{P}[\text{incomplete lineage sorting}] = \frac{2}{3}\left(1 - \frac{1}{2N_A}\right)^t \approx \frac{2}{3}\exp\left[-\frac{t}{2N_A}\right]$$

  So given t, we can find N_A, and thus there is a 30% probabiility of ILS.

5 ILS vs. Introgression

- Look at genes which disagree with the species tree. Consider the proportion of those which show human population 1 closer with Neandethal than with human population 2, and vice-versa. They should be approximately the same. N_{ABBA} vs. N_{BABA}.

$$D = \frac{\#(\text{ABBA}) - \#(\text{BABA})}{\#(\text{ABBA}) + \#(\text{BABA})}$$

6 Phenotypic Evolution

6.1 Phenotypic Resemblance between relatives

- Covariance, (Pearson) correlations, and slopes of linear regression

$$\begin{aligned}
\text{cov}[X, Y] &= \frac{1}{m}\sum_{i=1}^{m}\big(X_i - \overline{X}\big)\big(Y_i - \overline{Y}\big) = \overline{XY} - \overline{X}\cdot\overline{Y} = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
\text{cov}[X, X] &= \text{var}X \\
\text{cor}[X, Y] &= \frac{\text{cov}[X, Y]}{\sqrt{\text{var}X\,\text{var}Y}} \\
\text{slope}(Y \sim X) &= \frac{\text{cov}[X, Y]}{\text{var}X}
\end{aligned}$$

6.2 Mendel, Galton, Bateson, Fisher

- Galton: Regression toward the mean of heights of offspring when compared to their parents heights.
- Bateson: Populations were evolving through large effect mutations.
- Fisher: We can solve the problem of continuous variation by realizing that traits are blends of many Mendelian traits.

6.3 Example: BMI

- approximately normal distribution over the population
- Phenotypes are always due to the interaction of geners and environments

6.4 Resemblance between relatives in quantitative traits

- A trait with L loci.
- each segregating an allele A_1 at frequency $p\ell$.
- genotype at locus ℓ is $0, 1, 2$ with probs $p_\ell^2, 2p_\ell(1-p_\ell), (1-p_\ell)^2$.

$$X_{Ai} = \sum_{\ell=1}^{L} a_l G_{\ell,i} \quad \Longrightarrow \quad X_{P,i} = X_{Ai} + X_{Ei}$$

- So X_A is normally distributed at large enough L. It's reasonable to assume X_E is normally distributed. Thus X_P is normally distributed with

$$N(\mu_A + \mu_E, V_P) \qquad V_P = V_E + V_A$$

and define heritability (in the narrow sense) h by

$$h^2 \coloneqq \frac{V_A}{V_P}$$