Lecture 3. Inbreeding coefficient as correlation. HWE for multiple alleles. HWE for X-linked genes.

1.4 Inbreeding coefficient as correlation

Genotype A_1A_2 sampled at random

$\mathrm{P}(A_1 = A_2 = A) = D$, $\mathrm{P}(A_1 = A_2 = a) = R$

$\mathrm{P}(A_1 = A, A_2 = a) = \mathrm{P}(A_1 = a, A_2 = A) = H/2$

F = correlation coeff. between $1_{\{A_1=A\}}$ and $1_{\{A_2=A\}}$

$F = 0$: independent alleles

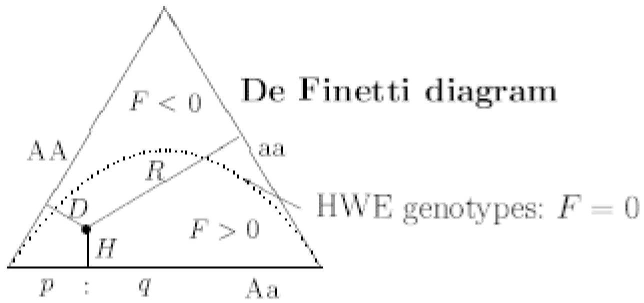
random genotype sampling = random allele sampling

$F > 0$: positive dependence

attraction of A to A and a to a, deficit of heterozygotes

$F < 0$: negative dependence

repulsion case, excess of heterozygotes



Ex 8: selfing

Mating genotypes: $AA \times AA$, $Aa \times Aa$, $aa \times aa$

$D_1 = D_0 + \frac{H_0}{4}$, $R_1 = R_0 + \frac{H_0}{4}$, $H_1 = \frac{H_0}{2}$

$D_t = p_0 - H_0 \cdot (0.5)^{t+1}$, $R_t = q_0 - H_0 \cdot (0.5)^{t+1}$

$H_t = H_0 \cdot (0.5)^t$, completely inbred line $F_t \to 1$

1.5 HWE for multiple alleles

One locus with k alleles $A_1, A_2, A_3, \ldots, A_k$

genotype frequencies: $p_{11}, p_{12}, p_{13}, p_{23}, p_{33}, \ldots$

Number of possible genotypes

number of heterozygotes + number of homozygotes

$= \binom{k}{2} + k = \frac{k(k+1)}{2}$

Allele frequencies: $p_1, p_2, p_3, \ldots, p_k$

$p_i = p_i^2 + \frac{1}{2}\Sigma_{j\neq i}\, p_{ij}$

HWE genotype freqencies uniquely define p_i

A_1A_1	A_1A_2	A_1A_3	A_2A_2	A_2A_3	A_3A_3	...
p_1^2	$2p_1p_2$	$2p_1p_3$	p_2^2	$2p_2p_3$	p_3^2	...

HWE heterozygosity $H = 1 - p_1^2 - \ldots - p_k^2$

Ex 11: ABO blood groups

Three alleles and four phenotypes = blood groups

A = $\{AA, AO\}$, AB = $\{AB\}$

B = $\{BB, BO\}$, O = $\{OO\}$

Spanish Basques sample

Blood group	A	B	O	AB	Total
observed counts	724	110	763	20	n=1617
expected counts	710.7	94.8	776.12	35.4	n=1617

EM estimates of allele frequencies

$\hat{p}_A = 0.2661$, $\hat{p}_B = 0.0411$, $\hat{p}_O = 0.6928$

$X^2 = 9.58$, df $= 4 - 3 = 1$, $\sqrt{9.58} = 3.1$

reject HWE (possibly due to immigration)

Papago Indians, Arizona

Blood group	A	O	B	AB	Total
observed counts	37	563	0	0	n=600

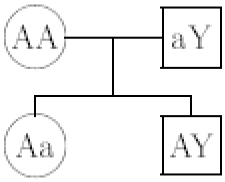
Estimated allele frequencies under HWE

$\hat{p}_B = 0$, $\hat{p}_O = \sqrt{\frac{563}{600}} = 0.97$, $\hat{p}_A = 0.03$

different frequencies in two populations, why?

1.6 HWE for X-linked genes

One gene on the X chromosome, two alleles A and a



Allele A frequencies in males p_{m} and females p_{f}

dynamics of the frequencies: $p'_{\mathrm{m}} = p_{\mathrm{f}}$, $p'_f = \frac{p_{\mathrm{m}} + p_{\mathrm{f}}}{2}$

Equilibrium frequencies: $\hat{p}_{\mathrm{m}} = \hat{p}_{\mathrm{f}} = \frac{p_{\mathrm{m}} + 2p_{\mathrm{f}}}{3}$

HWE: $D_{\mathrm{f}} = p^2$, $H_{\mathrm{f}} = 2pq$, $R_{\mathrm{f}} = q^2$, $p_{\mathrm{m}} = p_{\mathrm{f}} = p$

Recessive X-linked traits

affected males to females ratio $q_{\mathrm{m}}/R_{\mathrm{f}} = q/q^2 = 1/q$

Ex 13: color blindness

green blindness: $q = 0.05$, red blindness: $q = 0.01$

affected males to females ratios: 20 and 100

Ex 14: Xg blood group

X-linked gene with two alleles: $A = \mathrm{Xg}^a$ and $a = \mathrm{Xg}$

two blood types	Xg(a+)	Xg(a−)
female genotypes	$\mathrm{Xg}^a/\mathrm{Xg}^a$, $\mathrm{Xg}^a/\mathrm{Xg}$	Xg/Xg
male genotypes	Xg^a/Y	Xg/Y

British sample: female counts || male counts

	Xg(a+)	Xg(a−)	Total	Xg(a+)	Xg(a−)	Total
obs	967	102	1069	667	346	1013
exp	956.1	112.9	1069	683.8	329.2	1013

EM estimates: $\hat{p} = 0.675$, $\hat{q} = 0.325$

$X^2 = 2.45$, df $= 4 - 2 - 1 = 1$, $\sqrt{2.45} = 1.57$

not significant P-value $= 0.12$, do not reject HWE

Literature:

1. D.L.Hartl, A.G.Clarc. Principle of population genetics. Sinauer Associates, 2007.
2. R.Nielson, M. Statkin. An introduction to population genetics: theory and applications, Sinauer Associates. 2013.