Selection Components Analysis

Introduction

Consider the steps in a transition from one generation to the next, starting with a newly formed zygote:

- Zygote
- Adult — Survival from zygote to adult may differ between the sexes.
- Breeding population — Adult genotypes may differ in their probability of mating, and the differences may be different in males and females
- Newly formed zygotes

When the transition from one stage to the next depends on genotype, then selection has occurred at that stage. Thus, to determine whether selection is occurring we construct expectations of genotype or allele frequencies at one stage based on the frequencies at the immediately preceding stage assuming that no selection has occurred. Then we compare observed frequencies to those expected without selection. If they match, we have no evidence for selection. If they don't match, we do have evidence for selection.

As we've already seen, it's conceptually easy (if often experimentall difficult) to detect and measure selection if we can assay genotypes non-destructively at appropriate stages in the life-cycle. What if we can't? Well, there's a very nice approach known as *selection components analysis* that generalizes the approach to estimating relative viabilities that we've already seen.[1]

[1] Christiansen, F. B., and O. Frydenberg. 1973. Selection component analysis of natural polymorphisms using population samples including mother-offspring combinations. *Theor. Popul. Biol.* 4:425–445.

The Data

Pregnant mothers are collected. One offspring from each mother is randomly selected and its genotype determined. In addition, the genotypes of a random sample of non-reproductive ("sterile") females and adult males are determined. The data can be summarized as follows:

	Offspring					
Mother	A_1A_1	A_1A_2	A_2A_2	$\sum$	"Sterile" Females	Males
A_1A_1	C_{11}	C_{12}	—	F_1	S_1	M_1
A_1A_2	C_{21}	C_{22}	C_{23}	F_2	S_2	M_2
A_2A_2	—	C_{32}	C_{33}	F_3	S_3	M_3
Total				F_0	S_0	M_0

Given the total sample size for mother-offspring pairs, "sterile" females, and males, how many free parameters are there? How many frequencies would we need to know to reproduce the data?

6	for mother-offspring pairs
2	for "sterile" females
2	for males
10	total

The Analysis

H_1: Half of the offspring from heterozygous mothers are also heterozygous.
Under H_1

$$\begin{aligned}
\gamma_{21} &= (1/2)(F_2/F_0)(C_{21}/(C_{21}+C_{23})) \\
\gamma_{22} &= (1/2)(F_2/F_0) \\
\gamma_{23} &= (1/2)(F_2/F_0)(C_{23}/(C_{21}+C_{23}))
\end{aligned}$$

Under H_1, γ_{22} can be predicted just from the frequency of heterozygous mothers in the sample. Thus, only 9 parameters are needed to describe the data under H_1. Since 10 are required if we reject H_1 we can use a likelihood ratio test with one degree of freedom to see whether the above estimates provide an adequate description of the data.

If H_1 is rejected, we can conclude that there is either gametic selection or segregation distortion in A_1A_2 females.

H_2: **The frequency of transmitted male gametes is independent of the mother's genotype.**
Under H_2

$$\begin{aligned} p_m &= (C_{11} + C_{21} + C_{32})/(F_0 - C_{22}) \\ q_m &= (C_{12} + C_{23} + C_{33})/(F_0 - C_{22}) \end{aligned}$$

The expected frequency of the various mother-offspring combinations is

	A_1A_1	A_1A_2	A_2A_2
A_1A_1	$\phi_1 p_m$	$\phi_1 q_m$	—
A_1A_2	$(1/2)\phi_2 p_m$	$(1/2)\phi_2$	$(1/2)\phi_2 q_m$
A_2A_2	—	$\phi_3 p_m$	$\phi_3 q_m$

where $\phi_i = F_i/F_0$. Under H_2 only the female genotype frequencies and the male gamete frequencies are needed to describe the mother-offspring data. That's a total of $2 + 1 + 2 + 2 = 7$ frequencies needed to describe *all* of the data. Since H_1 needed 9, that gives us 2 degrees of freedom for our likelihood ratio test of H_2 given H_1.

If H_2 is rejected, we can conclude that there is some form of non-random mating in the breeding population or female-specific selection of male gametes.

H_3: **The frequency of the transmitted male gametes is equal to the allele frequency in adult males.**
Under H_3 the maximum likelihood estimates for p_m and q_m cannot be found explicitly, they are a complicated function of p_m and q_m as defined under H_2 and of M_1, M_2, and M_3. Under H_3, however, we no longer need to account separately for the gamete frequency in males, so a total of $2 + 2 + 2 = 6$ frequencies is needed to describe the data. Since H_2 needed 7, that gives us 1 degree of freedom for our likelihood ratio test of H_3 given H_2.

If H_3 is rejected, we can conclude either that males differ in their ability to attract mates (i.e., there is sexual selection) or that male gametes differ in their ability to accomplish fertilization (e.g., sperm competition), or that there is segregation distortion in A_1A_2 males.

H_4: **The genotype frequencies of reproductive females are the same as those of "sterile" females.**

Under H_4 the maximum likelihood estimates for the genotype frequencies in females are

$$\phi_i = (F_i + S_i)/(F_0 + S_0)$$

Under H_4 we no longer need to account separately for the genotype frequencies in "sterile" females, so a total of $2 + 2 = 4$ frequencies is needed to describe the data. Since H_3 needed 6, that gives us 2 degrees of freedom for our likelihood ratio test of H_4 given H_3.

If H_4 is rejected, we can conclude that females differ in their ability to reproduce successfully.

H_5: **The genotype frequencies of adult females and adult males are equal.**

Under H_5 the maximum likelihood estimates for the adult genotype frequencies can not be found explicitly. Instead, they are a complicated function of almost every piece of information that we have. Under H_5, however, we no longer need to account separately for the genotype frequencies in females and males, so a total of 2 frequencies is needed to describe the data. Since H_4 needed 4, that gives us 2 degrees of freedom for our likelihood ratio test of H_5 given H_4.

If H_5 is rejected we can conclude that the relative viabilities of the genotypes are different in the two sexes. (We have assumed implicitly throughout that the locus under study is an autosomal locus. Notice that rejection of H_5 is consistent with *no* selection in one sex.)

H_6: **The genotype frequencies in the adult population are equal to those of the zygote population.**

Under H_6 the maximum-likelihood estimator for the allele frequency in the population is

$$p = \frac{((C_{11} + C_{21} + C_{32}) + 2(F_1 + S_1 + M_1) + (F_2 + S_2 + M_2))}{((F_0 - C_{21}) + F_0 + S_0 + M_0)}$$

Under H_6 the genotype frequencies in our original table can be summarized as follows:

Mother	A_1A_1	A_1A_2	A_2A_2	$\sum$	"Sterile" Females	Males
A_1A_1	p^3	p^2q	0	p^2	p^2	p^2
A_1A_2	p^2q	pq	pq^2	$2pq$	$2pq$	$2pq$
A_2A_2	0	pq^2	q^3	q^2	q^2	q^2

In short, under H_6 only one parameter, the allele frequency, is required to describe the entire data set. Since under H_5 needed two parameters, our likelihood ratio test of H_6 given H_5 will have one degree of freedom.

If H_6 is rejected, we can conclude that genotypes differ in their probability of survival from zygote to adult, i.e., that there is viability selection. If H_1–H_6 are accepted, we have no evidence that selection is happening at any stage of the life cycle at this locus and no evidence of non-random mating with respect to genotype at this locus.

An example

This data is from a 2-allelic esterase polymorphism in our old friend *Zoarces viviparus*, the eelpout. The observations are in roman type in the table below. The numbers in italics are those expected if hypotheses H_1–H_6 are accepted.

Mother	A_1A_1	A_1A_2	A_2A_2	$\sum$	"Sterile" Females	Males
	41	70	—	111	8	54
A_1A_1	*39.0*	*67.0*	—	*106.0*	*9.3*	*58.4*
	65	173	119	357	32	200
A_1A_2	*67.0*	*181.9*	*114.9*	*363.8*	*32.1*	*200.5*
	—	127	187	314	29	177
A_2A_2	—	*114.9*	*197.3*	*312.2*	*27.6*	*172.1*
	106	370	306	782	69	431
Sum	*106.0*	*363.8*	*312.2*	—	—	—

The results of the series of hypothesis tests is as follows:

Hypothesis	Degrees of freedom	χ^2	P	50% power point
H_1	1	0.34	>0.50	0.05
H_2	2	1.37	>0.50	$\leq$ 0.09
H_3	1	0.98	>0.30	$\leq$ 0.05
H_4	2	0.37	>0.50	$\leq$ 0.10
H_5	2	0.22	>0.80	$\leq$ 0.05
H_6	1	0.09	>0.70	0.03

We conclude from this analysis that there is no evidence of selection on the genetic variation at the esterase locus in *Zoarces viviparus* and that there is no evidence of non-random mating with respect to genotype at this locus. The power calculations increase our confidence that if there is selection happening, the differences among genotypes are on the order of just a few percent.

Creative Commons License