

ESTIMATING VIABILITY

Introduction

Being able to make predictions with known (or estimated) viabilities, doesn't do us a heck of a lot of good unless we can figure out what those viabilities are. Fortunately, figuring them out isn't too hard.¹ If we know the number of individuals of each genotype before selection, it's really easy as a matter of fact.² Consider that our data looks like this:

Genotype	A_1A_1	A_1A_2	A_2A_2
Number in zygotes	$n_{11}^{(z)}$	$n_{12}^{(z)}$	$n_{22}^{(z)}$
Viability	w_{11}	w_{12}	w_{22}
Number in adults	$n_{11}^{(a)} = w_{11}n_{11}^{(z)}$	$n_{12}^{(a)} = w_{12}n_{12}^{(z)}$	$n_{22}^{(a)} = w_{22}n_{22}^{(z)}$

In other words, estimating the absolute viability simply consists of estimating the probability that an individuals of each genotype that survive from zygote to adult. The maximum-likelihood estimate is, of course, just what you would probably guess:

$$w_{ij} = \frac{n_{ij}^{(a)}}{n_{ij}^{(z)}} \quad ,$$

Since w_{ij} is a probability and the outcome is binary (survive or die), you should be able to guess what kind of likelihood relates the observed data to the unseen parameter, namely, a binomial likelihood. In Stan notation:³

```
n_11_adult ~ binomial(n_11_zygote, w_11)
n_12_adult ~ binomial(n_12_zygote, w_12)
n_22_adult ~ binomial(n_22_zygote, w_22)
```

¹I almost said that it was easy, but that would be going a bit too far.

²And in the very next sentence I contradicted the last footnote. But it really is easy to estimate viabilities if we can genotype individuals before and after selection.

³You knew you were going to see this again, didn't you?

Estimating relative viability

To estimate absolute viabilities, we have to be able to identify genotypes non-destructively, because we have to know what their genotype was both *before* the selection event and *after* the selection event. That's fine if we happen to be dealing with an experimental situation where we can do controlled crosses to establish known genotypes or if we happen to be studying an organism and a trait where we can identify the genotype from the phenotype of a zygote (or at least a very young individual) and from surviving adults.⁴ What do we do when we can't follow the survival of individuals with known genotype? Give up?⁵

Remember that to make inferences about how selection will act, we only need to know *relative* viabilities, not *absolute* viabilities.⁶ We still need to know something about the genotypic composition of the population before selection, but it turns out that if we're only interested in relative viabilities, we don't need to follow individuals. All we need to be able to do is to score genotypes and estimate genotype frequencies before and after selection. Our data looks like this:

Genotype	A_1A_1	A_1A_2	A_2A_2
Frequency in zygotes	$x_{11}^{(z)}$	$x_{12}^{(z)}$	$x_{22}^{(z)}$
Frequency in adults	$x_{11}^{(a)}$	$x_{12}^{(a)}$	$x_{22}^{(a)}$

We also know that

$$\begin{aligned} x_{11}^{(a)} &= w_{11}x_{11}^{(z)}/\bar{w} \\ x_{12}^{(a)} &= w_{12}x_{12}^{(z)}/\bar{w} \\ x_{22}^{(a)} &= w_{22}x_{22}^{(z)}/\bar{w} \quad . \end{aligned}$$

Suppose we now divide all three equations by the middle one:

$$\begin{aligned} x_{11}^{(a)}/x_{12}^{(a)} &= w_{11}x_{11}^{(z)}/w_{12}x_{12}^{(z)} \\ 1 &= 1 \\ x_{22}^{(a)}/x_{12}^{(a)} &= w_{22}x_{22}^{(z)}/w_{12}x_{12}^{(z)} \quad , \end{aligned}$$

⁴How many organisms and traits can you think of that satisfy this criterion? Any? There is one other possibility: If we can identify an individual's genotype after it's dead *and* if we can construct a random sample that includes both living and dead individuals *and* if we the probability of including an individual in the sample doesn't depend on whether that individual is dead or alive, then we can sample a population after the selection event and score genotypes both before and after the event from one set of observations.

⁵Would I be asking the question if the answer were "Yes"?

⁶At least that's true until we start worrying about how selection and drift interact.

or, rearranging a bit

$$\frac{w_{11}}{w_{12}} = \left(\frac{x_{11}^{(a)}}{x_{12}^{(a)}} \right) \left(\frac{x_{12}^{(z)}}{x_{11}^{(z)}} \right) \quad (1)$$

$$\frac{w_{22}}{w_{12}} = \left(\frac{x_{22}^{(a)}}{x_{12}^{(a)}} \right) \left(\frac{x_{12}^{(z)}}{x_{22}^{(z)}} \right) . \quad (2)$$

This gives us a complete set of relative viabilities.

Genotype	A_1A_1	A_1A_2	A_2A_2
Relative viability	$\frac{w_{11}}{w_{12}}$	1	$\frac{w_{22}}{w_{12}}$

If we use the maximum-likelihood estimates for genotype frequencies before and after selection, we obtain maximum likelihood estimates for the relative viabilities.⁷ If we use Bayesian methods to estimate genotype frequencies before and after selection (including the uncertainty around those estimates), we can use these formulas to get Bayesian estimates of the relative viabilities (and the uncertainty around them).

An example

Let's see how this works with some real data from Dobzhansky's work on chromosome inversion polymorphisms in *Drosophila pseudoobscura*.⁸

Genotype	ST/ST	ST/CH	CH/CH	Total
Number in larvae	41	82	27	150
Number in adults	57	169	29	255

You may be wondering how the sample of adults can be larger than the sample of larvae. That's because to score an individual's inversion type, Dobzhansky had to kill it. The

⁷If anyone cares, it's because of the invariance property of maximum-likelihood estimates. If you don't understand what that is, don't worry about it, just trust me. Or if you want to know what the invariance principle is, ask.

⁸Taken from [1].

numbers in larvae are based on a sample of the population, and the adults that survived were not genotyped as larvae. As a result, all we can do is to estimate the relative viabilities.

$$\frac{w_{11}}{w_{12}} = \left(\frac{x_{11}^{(a)}}{x_{12}^{(a)}} \right) \left(\frac{x_{12}^{(z)}}{x_{11}^{(z)}} \right) = \left(\frac{57/255}{169/255} \right) \left(\frac{82/150}{41/150} \right) = 0.67$$

$$\frac{w_{22}}{w_{12}} = \left(\frac{x_{22}^{(a)}}{x_{12}^{(a)}} \right) \left(\frac{x_{12}^{(z)}}{x_{22}^{(z)}} \right) = \left(\frac{29/255}{169/255} \right) \left(\frac{82/150}{27/150} \right) = 0.52 \quad .$$

So it looks as if we have balancing selection, i.e., the fitness of the heterozygote exceeds that of either homozygote.

We can check to see whether this conclusion is statistically justified by comparing the observed number of individuals in each genotype category in adults with what we'd expect if all genotypes were equally likely to survive.

Genotype	<i>ST/ST</i>	<i>ST/CH</i>	<i>CH/CH</i>
Expected	$\left(\frac{41}{150}\right) 255$	$\left(\frac{82}{150}\right) 255$	$\left(\frac{27}{150}\right) 255$
	69.7	139.4	45.9
Observed	57	169	29
$\chi_2^2 = 14.82, P < 0.001$			

So we have strong evidence that genotypes differ in their probability of survival.

We can also use our knowledge of how selection works to predict the genotype frequencies at equilibrium:

$$\frac{w_{11}}{w_{12}} = 1 - s_1$$

$$\frac{w_{22}}{w_{12}} = 1 - s_2 \quad .$$

So $s_1 = 0.33$, $s_2 = 0.48$, and the predicted equilibrium frequency of the *ST* chromosome is $s_2/(s_1 + s_2) = 0.59$.

Now all of those estimates are maximum-likelihood estimates. Doing these estimates in a Bayesian context is relatively straightforward and the details will be left as an exercise.⁹ In outline we simply

1. Estimate the genotype frequencies before and after selection as samples from a multinomial.

⁹In past years Project #3 has consisted of making Bayesian estimates of viabilities from data like these and predicting the outcome of viability selection.

2. Apply the formulas from equations (1) and (2) to calculate relative viabilities and selection coefficients.
3. Determine whether the 95% credible intervals for s_1 or s_2 overlap 0.¹⁰
4. Calculate the equilibrium frequency from $s_2/(s_1+s_2)$, if $s_1 > 0$ and $s_2 > 0$.¹¹ Otherwise, determine which fixation state will be approached.

In the end you then have not only viability estimates and their associated uncertainties, but a prediction about the ultimate composition of the population, associated with an accompanying level of uncertainty.

References

- [1] Th. Dobzhansky. Genetics of natural populations. XIV. A response of certain gene arrangements in the third chromosome of *Drosophila pseudoobscura* to natural selection. *Genetics*, 32:142–160, 1947.

Creative Commons License

These notes are licensed under the Creative Commons Attribution License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

¹⁰Meaning that we don't have good evidence for selection either for or against the associated homozygotes, relative to the heterozygote.

¹¹In practice, this gets a little complicated. What typically happens is that in some samples from the posterior the heterozygote is intermediate in fitness, meaning that one of the two homozygotes is unconditionally favored. That makes calculating the posterior distribution for the equilibrium frequency a bit complicated. We'll avoid those complications in this year's project.