

## 5.5 How an advantageous mutation spreads through a population

Almost every probability textbook treats the basic setup of genetics – an individual has two genes at a site, inheriting one from each parent, this one chosen at random from the parent’s two. It is noteworthy that this is one of very few interesting real-world phenomena that (as far as scientists understand) really is “physically random” like coin tossing. Introductory textbooks typically treat the dominant/recessive setting, for instance in simplified models of eye color. Their theory gives predictions for genetic type of child in terms of genetic type of parents; and (via Bayes rule and empirical population frequencies) predictions for child’s eye color in terms of parents’ eye color.

Instead of that familiar topic, I will describe toy models in population genetics. Let me approach the topic obliquely, via a little history. The *fact* that empires have risen and fallen throughout human history is hardly controversial. People sometimes propose general explanations for how this happens – divine favor, racial superiority, class struggle, technological superiority, societal ethics, ecological collapse – but none is not widely accepted, and indeed are generally taken to reflect prejudices of the era when they were formulated. By a rough analogy, in the middle of the nineteenth century, once dinosaur and other fossils were being discovered, the proposition that life on Earth has been in existence for a very long time, that earlier species had become extinct and that other species had originated – this proposed *fact* wasn’t particularly controversial. But the point of Darwin’s idea of “evolution by natural selection” was that there is one *explanation* of this process – natural selection. Darwin and his nineteenth century followers did not have our current notion of genetics and did not seek a mathematical formulation of their theory. And indeed they were aware that there was a difficulty with the whole idea, if approached from a certain common sense view of heredity (“paint mixing”, below). Let me first describe the difficulty, and then show how it is resolved in the correct theory of genetics.

### 5.5.1 If heredity were like paint mixing

Observation of animal breeding might suggest offspring are a mixture of parents, like a mixture of blue and yellow paint makes green paint. Of course this couldn’t be the whole story, or every individual in a population would be identical by heredity, but (unaware of genetics) we might imagine heredity working as “mixture of parents, plus individual randomness”. And indeed this kind of “additive” model does correctly predict the behavior of some real-world quantitative characteristics, for instance height in humans ([15] sec. 25.3). However, let us consider a model for how natural selection might work on a novel heritable trait, if heredity were like paint mixing. We’ll give a model that ignores randomness (both in number of offspring and assumed “individual randomness”), but incorporating randomness doesn’t change the conclusions.

**A paint mixing model.** One individual (in generation 0, say) has a new characteristic giving selective advantage  $\alpha$ , meaning that the mean number of offspring reaching maturity is  $2(1 + \alpha)$  instead of 2. Each offspring (generation 1) has only half of the characteristic (this is the “like paint mixing” assumption), so has selective advantage  $\alpha/2$ , so each generation 1 offspring has mean number  $2(1 + \frac{1}{2}\alpha)$  offspring in generation 2, and these generation 2 individuals has a quarter of the characteristic. So the “penetration” (sum over individuals of their proportion of the characteristic) of the characteristic in successive generations is

generation	0	1	2
mean number individuals	1	$2(1 + \alpha)$	$4(1 + \alpha)(1 + \frac{1}{2}\alpha)$
proportion of characteristic	1	$\frac{1}{2}$	$\frac{1}{4}$
penetration	1	$1 + \alpha$	$(1 + \alpha)(1 + \frac{1}{2}\alpha)$

As time passes the mean penetration increases, not indefinitely but only to a finite limit

$$\beta(\alpha) = \prod_{i=0}^{\infty} (1 + 2^{-i}\alpha)$$

which for small  $\alpha$  is approximately  $1 + 2\alpha$ . This value doesn’t depend on the population size ( $N$ , say). So the key conclusion is that the effect of a single appearance of a new characteristic would be, after many generations, that each individual in the population gets a proportion around  $(1 + 2\alpha)/N$  of the characteristic.

This conclusion is bad news for a theory of natural selection, because it implies that to become “fixed” in a population, a new characteristic would have to reappear many times – order  $N$  times – even when it provides a selective advantage.

### 5.5.2 The genetic model

How does genetics really work? What I describe here is less realistic than the usual toy model (the Wright-Fisher model – see section 5.7.1) but leads to the same formula. We consider genes rather than individuals, so there are  $2N$  genes in each generation. On average, a gene has 1 copy in the next generation, with some s.d. ( $= \sigma$ , say). For a new allele (the *alleles* are the possible forms of a given gene) which confers a small selective advantage, we suppose the average number of copies becomes  $\mu = 1 + \alpha$ . This can only be true while the penetration (proportion of all genes which are that allele) is small, which is an unrealistic aspect of the model.

So suppose a mutation creates a new allele with small selective advantage  $\alpha$ . Then the number of copies behaves (while penetration is small, at least) as a just supercritical Galton-Watson process described in section 5.6. In particular, either the new allele disappears from the population quite quickly (extinction, in the Galton-Watson terminology) or the penetration grows and eventually becomes *fixed* in the population – every gene is this allele. So the formula (5.3)

for survival probability of just supercritical Galton-Watson processes carries over to the present setting

For a single mutation giving a gene with small selective advantage  $\alpha$ , the chance that the gene becomes fixed in the population is about  $\frac{2\alpha}{\sigma^2}$ .

This conclusion is much better news for a theory of natural selection, because now the population size doesn't matter. If the chance above were 1/10, say, then an advantageous mutation needs to reappear only 10 or 20 times to be likely to become "fixed" in the population, regardless of how large the population size  $N$  is.

The whole process of an allele becoming fixed in this way is called a *selective sweep*. Once a sweep is under way, the number of copies grows at rate  $\alpha$  per generation, and so

$$\text{duration of a selective sweep} \approx \frac{\log(2N)}{\alpha} \text{ generations} . \quad (5.1)$$

## 5.6 A classic toy model: the Galton-Watson process

This is used as a toy model in many different settings. To have a concrete language in this section we talk about "individuals" and "offspring". To visualize individuals and offspring, you can either imagine asexual reproduction or look only at males or only at females in a two-sex species like humans. Note that in the use in section 5.5.2 the "individuals" are genes of a particular allelic type (xxx other examples).

The assumption is that there is a probability distribution  $\mathbf{p} := (p_i, i = 0, 1, 2, \dots)$  and that each individual in a generation has a random number of offspring in the next generation, this number being picked from  $\mathbf{p}$  independently for different parents. Write  $\mu$  and  $\sigma$  for the mean and s.d. of the number of offspring. By default we assume the process starts with 1 individual in generation 0; so there is some random number  $Z_n \geq 0$  of individuals in each generation  $n = 0, 1, 2, 3, \dots$ . There are two logical possibilities for what might happen in the long run:

- either "extinction" meaning  $Z_n = 0$  for all large  $n$
- or "survival", meaning  $Z_n \geq 1$  for all  $n$ .

One of the highlights of an undergraduate course in stochastic processes is the following xxx cite

**Theorem.** (a) If  $\mu < 1$  then  $P(\text{extinction}) = 1$ .

(b) If  $\mu > 1$  then  $\rho = P(\text{extinction}) < 1$  and is the solution of the equation

$$\rho = \Phi(\rho) \quad (5.2)$$

where  $\Phi$  is the probability generating function defined by

$$\Phi(z) = \sum_{i=0}^{\infty} p_i z^i.$$

Keep in mind that the “independence” assumptions are tantamount to assuming there is no “interaction” between individuals and that there are no external constraints on population size – both assumptions are rather unrealistic in most contexts.

(xxx somewhere mention: Galton’s study of extinction of family names, mitochondrial Eve, etc)

I won’t repeat the textbook derivation of the Theorem, but I will derive an interesting approximate formula for a particular setting. The cases  $\mu < 1$ ,  $\mu = 1$ ,  $\mu > 1$  are called *subcritical*, *critical*, *supercritical*. I want to consider the “just supercritical” case where  $\mu > 1$  but  $\mu - 1$  is small.

For a just supercritical Galton-Watson process,  $P(\text{survival}) \approx \frac{2(\mu-1)}{\sigma^2}$ . (5.3)

This is often not mentioned in textbooks, so let me give a derivation.

### 5.6.1 Derivation of formula (5.3)

Textbook facts about the probability generating function for the random number  $X$  of offspring are

$$\Phi(1) = 1, \quad \Phi'(1) = \mu, \quad \Phi'' = E[X(X-1)] = \sigma^2 + \mu^2 - \mu \approx \sigma^2$$

the approximation holding because  $\mu \approx 1$ .

We want the survival probability  $\bar{\rho} = 1 - \rho$ . The equation in the Theorem,  $\rho = \Phi(\rho)$ , can be rewritten in terms of  $\bar{\rho}$  as  $h(1 - \bar{\rho}) = 0$ , where  $h(x) = \Phi(x) - x$ . Consider the series expansion: for small  $x$ ,

$$h(1 - x) \approx h(1) - xh'(1) + \frac{1}{2}x^2h''(1).$$

Since  $h(1) = 0$ ,  $h'(1) = \mu - 1$ ,  $h''(1) \approx \sigma^2$  the rewritten equation becomes

$$0 \approx -\bar{\rho}(\mu - 1) + \frac{1}{2}\bar{\rho}^2\sigma^2$$

and solving for  $\bar{\rho}$  gives the stated formula (5.3).

### 5.6.2 Phase transitions, scaling laws and universality

The phrase *phase transition* has a particular meaning in physics

In thermodynamics, *phase transition* is the transformation of a thermodynamic system from one phase to another. The distinguishing characteristic of a phase transition is an abrupt change in one or more physical properties, in particular the heat capacity, with a small change in a thermodynamic variable such as the temperature. In the English vernacular, the term is most commonly used to describe transitions between solid, liquid and gaseous states of matter, in rare cases including plasma. [67]

In addition to this strict meaning within physics, the phrase *phase transition* has acquired a broader meaning within probability models in any discipline. Suppose the model has a parameter  $\lambda$  and suppose there is a *critical value*  $\lambda_c$  such that the behaviors when  $\lambda < \lambda_c$  and when  $\lambda > \lambda_c$  are qualitatively different. Then one says there is a phase transition at  $\lambda_c$ .

The Galton-Watson model provides an illustration, with the parameter  $\mu =$  mean number of offspring. For  $\mu < 1$  survival is impossible whereas for  $\mu > 1$  survival is possible, so one could speak of a phase transition at critical value  $\mu_c = 1$ .

Continuing toward using the Galton-Watson process to illustrate our next conceptual point, in the subcritical ( $\mu < 1$ ) case an easy argument shows that the mean population size in generation  $n$  is  $EZ_n = \mu^n$ , and hence the total population before extinction, say  $Z := \sum_{n \geq 0} Z_n$ , satisfies

$$EZ = \frac{1}{1-\mu}. \quad (5.4)$$

We can summarize the two different facts (5.3,5.4) as

mean total population  $\propto (1-\mu)^{-1}$  as  $\mu \uparrow 1$ ; survival probability  $\propto (\mu-1)^1$  as  $\mu \downarrow 1$ .

This illustrates a general observation about probability models that has been a central theme of statistical physics. In a model with a phase transition, take some “natural statistic”  $s(\lambda)$  measuring quantitatively some feature that has a qualitative change at the critical point. Typically such a statistic tends to 0 or  $\infty$  as the critical point is approached from one side, and moreover follows a power law

$$s(\lambda) \propto |\lambda - \lambda_c|^\gamma \text{ as } \lambda \rightarrow \lambda_c \text{ from one side}$$

for some *scaling exponent*  $\gamma$  which is often an integer or a simple fraction. So (5.3,5.4) illustrate this *scaling law* behavior, with scaling exponents  $-1$  and  $1$ .

In very simple models (again illustrated by the Galton-Watson model) we can do exact calculations of some statistic  $s(\lambda)$  and then obtain a scaling exponent from an explicit formula. Such models are usually oversimplified; if we study more complicated models, the values of the statistics  $s(\lambda)$  and critical points  $\lambda_c$  will change. But the *universality* paradigm asserts that the numerical value of the scaling exponent will often remain unchanged.

The range of validity of this paradigm is a topic on ongoing theoretical research. The conceptual idea is that the intrinsic mathematical nature of what

we're modeling is captured better by the scaling exponents of natural statistics than by other aspects of particular models.

xxx real-world applicability less clear – motivates search for power laws.

xxx contact process as example?