

## OLM 11.5. Species-abundance distribution under neutral dynamics

In a finite population of neutral variants without new ones occurring, drift eventually clears away all but a single variant (Ch11.2, p.240). However, if new variants can pop up from time to time, then an equilibrium number and distribution of variants may ensue (Ch11.3, p.246). This OLM demonstrates that, assuming a high number of completely neutral variants, abundance of the variants follow Fisher's log-series distribution. It should be noted here, however, that Fisher did not propose this distribution for modelling abundance distributions (Fisher et al. 1943) on the basis of some principle, but for technical reasons (OLM 11.6). The postulates of the model (a high number of absolutely neutral variants) can be approximately true for alleles, but not for species (Note 7.1, p.141), so the occurrence of new variants will be termed "mutations" hereafter.

Let us start from a general model of the abundance distribution of any kinds of different (i.e., not necessarily neutral) variants. Denote the probability of variant  $k$  having  $n$  individuals at time  $t$  by  $P_{k,n}(t)$ . The change in this probability during a time period sufficiently short to allow for a maximum of one demographic event (birth or death) to happen in the population is  $dP_{k,n}(t)/dt$ . Denote the probability of a birth event in a population of  $n$  individuals of variant  $k$  during time  $dt$  by  $B_{k,n}$ , and the corresponding death probability by  $D_{k,n}$ .

$P_{k,n}$  decreases, if

- one of  $n$  individuals dies, reducing the population size of variant  $k$  by one; the probability of this to happen is  $P_{k,n}D_{k,n}$
- one of  $n$  individuals of variant  $k$  gives birth to an offspring, increasing the size of the population of its kind by one; this occurs with a probability  $P_{k,n}B_{k,n}$

$P_{k,n}$  increases, if

- one of  $n+1$  individuals dies; the probability of this to happen is  $P_{k,n+1}D_{k,n+1}$
- one of  $n+1$  individuals of variant  $k$  gives birth to an offspring; this occurs with a probability  $P_{k,n+1}B_{k,n+1}$

Summation of all these probabilities yields

$$\frac{dP_{k,n}}{dt} = P_{k,n-1}B_{k,n-1} + P_{k,n+1}D_{k,n+1} - P_{k,n}B_{k,n} - P_{k,n}D_{k,n} \quad (11.5.1)$$

Assuming that mutations are rare events occurring during birth events, let  $B_{k,0} > 0$  in this model. If the potential number of possible variants is large, we can safely omit the rare cases of mutations into an already existing variant.

In steady state,

$$\frac{dP_{k,n}}{dt} = 0 \tag{11.5.2}$$

Figure 11.5.1 shows that this condition is met if the fluxes (arrows) into and out of the compartment representing an abundance category sum up to zero for all compartments. Since the  $P_{k,0}$  compartment has just one flux in and one out on its right side, these fluxes must be equal at equilibrium. This also means that all the remaining compartments must have equal fluxes on their right sides, too:

$$D_{k,n}P_{k,n} = B_{k,n-1}P_{k,n-1} \tag{11.5.3}$$

Rearranging Eq.(11.5.3) yields

$$P_{k,n} = P_{k,n-1} \frac{B_{k,n-1}}{D_{k,n}} = P_{0,k} \prod_{i=1}^n \frac{B_{k,i-1}}{D_{k,i}} \tag{11.5.4}$$

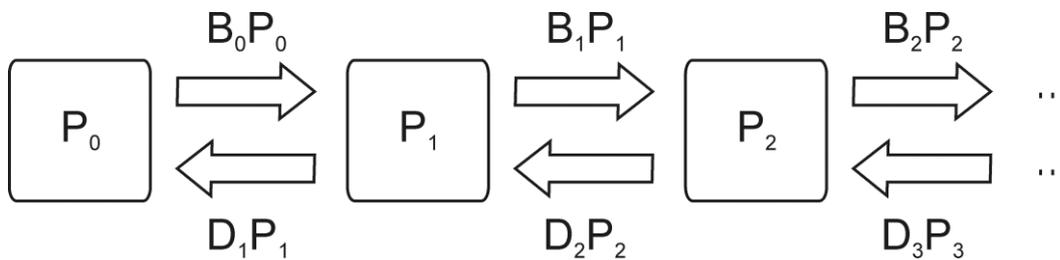


Figure 11.5.1 The probabilities of transitions between the possible abundance states of variant  $k$ , assuming that at most one individual dies or is born during each time period  $dt$ .

Compartments represent abundance categories; arrows are transition probabilities. For an explanation of notations, see text.

Assume that the different variants are neutral, i.e., their birth and death probabilities are equal. Then we get the same equilibrium distribution of abundance for all the variants, and this distribution also gives the probability that a randomly chosen variant will have exactly  $n$  individuals at a given instant:

$$P_n = P_0 \prod_{i=1}^n \frac{B_{i-1}}{D_i} \tag{11.5.5}$$

The expected number of variants falling into a certain abundance category can be obtained by multiplying this probability with the number of possible variants:

$$S_n = SP_n = SP_0 \prod_{i=1}^n \frac{B_{i-1}}{D_i} \quad (11.5.6)$$

Assume that the per capita birth and death rates do not depend on the abundance of the variant, i.e.,  $B_i=bn$  (for  $i>0$ ) és  $D_i=dn$  (for  $i>0$ ). Substituting this into Equation 11.5.6 we get

$$S_n = SP_0 B_0 \frac{b^{n-1} n^{n-1}}{d^n n^n} = \frac{P_0 B_0}{bn} \left(\frac{b}{d}\right)^n = \alpha \frac{x^n}{n} \quad (11.5.7)$$

where  $\alpha = \frac{SP_0 B_0}{b}$ , and  $x=b/d$ . The distribution we have arrived at is Fisher's log-series or logarithmic distribution. Parameter  $x$  has to be always smaller than 1, implying  $b<d$ . However, this condition does not necessarily lead to the exponential extinction of the population, since we have not considered the cases of the births of mutant offspring which can compensate for the effect of the larger death rate.

The expected number of variants ( $S_T$ ) within the population can be obtained by summing the  $S_n$  values for all  $n > 0$ :

$$S_T = \sum_{n=1}^{\infty} S_n = \alpha \sum_{n=1}^{\infty} \frac{x^n}{n} = -\alpha \ln(1-x) \quad (11.5.8)$$

(The last step includes the substitution of the Taylor series form of the  $\ln(1-x)$  function at  $x = 0$ , making use of the  $x < 1$  condition.)

For a given pair of  $x$  and  $\alpha$  parameters, the size of the population (the total number of individuals counted over all variants) can be obtained as the sum of a geometric series:

$$N = \sum_{n=1}^{\infty} n S_n = \sum_{n=1}^{\infty} \alpha x^n = \alpha \frac{x}{1-x} \quad (11.5.9)$$

The model can be parametrized in another way, too, assuming that the total size of the population ( $N$ ) is constant. Then the number of births and deaths is equal, meaning that the number of births resulting in offspring different from the parent ("mutations") is  $dN-bN$ . Denote the number of such births (i.e., the mutation rate) by  $\nu$

$$\nu = \frac{dN-bN}{bN} = \frac{d}{b} - 1 = \frac{1}{x} - 1 = \frac{1-x}{x} \quad (11.5.10)$$

Substituting this relation of  $\nu$  and  $x$  into Eq.(11.5.9) and rearranging gives  $\alpha = N\nu$ . That is, the two parameters of the distribution can be population size and mutation rate as well. Then the formula for the number of variants falling into each abundance category is somewhat more complicated:

$$S_n = \frac{N\nu}{n(\nu+1)^n} \quad (11.5.11)$$

The expected number of variants within the population at this parameter setting is

$$S_T = Nv \ln \left( \frac{v+1}{v} \right) \quad (11.5.12)$$

The number of variants occurring increases with both the mutation rate and the population size. The linear dependence of  $S_T$  on population size can be valid only as long as the assumption that all mutations give rise to a new variant never seen before holds, however.

Fisher's log-series distribution is often fitted to species-abundance data as well, and some authors consider a good fit to be proof of neutral dynamics. We have shown here that neutral dynamics indeed yields such distribution patterns, so that any species-abundance distribution substantially differing from this one must be shaped by other effects, too. The complementary statement – that a good fit of species-abundance data to the Fisher log-series distribution is proof of neutral dynamics – needs not be true, however. We have demonstrated in Ch11.3 (p.246) that each of a number of theoretical distributions may fit equally well to the same observed empirical distribution. In such cases the species-abundance pattern may not be sufficient to identify the mechanism it was generated by. The most we can get from such studies is excluding mechanism which could not have possibly produced them.

## References

Fisher, R.A., Corbet, A.S. and Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, 12(1): 42-58.