OLM 2.3. Likelihood function and model selection

Modelling and model selection are essential components of ecological research as we explained in Ch2.7 (p.27). Models' parameters are often estimated by maximum likelihood estimators and model selection is often based on information theory functions calculated from likelihood that is why we briefly introduced the concept of likelihood there. Here we provide the simplest possible numerical example for illustration of the concepts and their application. For further reading we recommend (Burnham and Anderson 2013).

The *likelihood* of a parameterised model is the probability of obtaining a given set of observed data using the model with the specified parameters. In the book we illustrated the meaning of this statement by a simple case of a cohort of ten cloned individuals in identical environments of whom two dies in a certain period of time (TBE, p.29). As this situation corresponds to a *binomial (Bernoulli) experiment* we may calculate the probability of this observation for a series of survival probabilities (p) which provides the *likelihood function*, $L(p)$. The model used is based on the binomial distribution also given in Eq. (2.1).

The likelihood function calculated for a binomial experiment with 10 individuals (Table 2.3.1.b) shows that the probability of the event of exactly 2 individuals dying (i.e., k = 8 surviving) out of 10 if each of them has a chance of p = 0.8 of survival is

$$L(p = 0.8) = \mathrm{P}(N = 10, k = 8, p = 0.8) = 0.30 \qquad (2.3.1)$$

which is rather small. However, the same outcome would be even more unlikely for any p different from 0.8 (Table 2.3.1..b). That is, 0.8 is the maximum likelihood estimation for p, the survival probability.

Table 2.3.1. The likelihood estimation of survival probability.
a) The simplest data table. b) The $L(p)$ probability of the outcome given in the data table, calculated for different survival probabilities (p), assuming that the fate of each individual is an outcome of a binomial trial.

(a)

alive	dead	all individ.
8	2	10

(b)

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	**0.8**	0.9
$L(p)$	0.00	0.00	0.00	0.01	0.04	0.12	0.23	**0.30**	0.19

Nothing guarantees, however, that 8/10 is indeed the chance of survival, since the fates of 10 individuals comprise a very small number of realizations for the stochastic death/survival process. Expressing this in statistical terms, a 10-element sample is a very small sample of the statistical population. Assume now that the experiment is repeated with 200 individuals, and the number of survivors was 166, i.e., 34 individuals died. Then the relative frequency of

survivors is 166/200 = 0.83. For comparing the likelihood functions of the two samples we need to normalize the likelihood values with their maxima: $L(p)/L_{max}$, and then plot the normalized functions against p (Figure 2.9). The position of the maximum has hardly changed with sample size, but the larger sample yields a likelihood function with a much narrower range of plausible values.

The state of individuals (e.g. their gender) may affect their fates. Suppose that the sample of 200 individuals consists of 102 males, of which 82 survives, and of 98 females with 84 survivors. Then we have the distribution given in Table 2.3.2.

Table 2.3.2. A simple contingency table, with two different variables corresponding to rows and columns. The numbers in the cells are frequencies.

	Alive	Dead
Female	84	14
Male	82	20

This situation differs from the binomial trial only in that the random experiments have four different outcomes corresponding to the cells of Table 2.3.2 instead of the two in Table 2.3.1..a. Repeating the random experiment 200 times and again assuming that the fates of the individuals are independent of each other, the probability of the realized outcome can be calculated from the multinomial distribution:

$$L(p_1,p_2,p_3,p_4) = P(84,14,82,20) = \frac{200!}{84!14!82!20!} p_1^{84},p_2^{14},p_3^{82},p_4^{20} \qquad (2.3.2)$$

L has the highest value (0.000475) when the p_i probabilities are equal to the relative frequencies of the corresponding categories, i.e., p_1=0.42, p_2=0.07, p_3=0.41, p_4=0.1. Note that the model has only three parameters, since the sum of four probabilities has to be 1.

We may ask whether males and females are different in their survival probabilities based on the sample of Table 2.3.2. If there is no difference, then we may use the same survival probability for both genders estimated from the pooled data, thus reducing the number of model parameters. The likelihood obtained for the pooled model is L = 0.00038, somewhat smaller than for the model making the gender distinction. However, the pooled model has only two free parameters instead of three, and the standard deviation of the survival probability is smaller than in the model distinctive for gender. If the two models are embedded, i.e., if the model with fewer free parameters is a special case of the other one (which condition is satisfied in our example), then we may use the *log likelihood ratio test* based on the log ratio of the L values of the two models to decide (with a hypothesis test) if the model with more parameters fits to the data significantly better than the other one. By

calculating the test in our example we may conclude that the gender distinction for survival probabilities is not justified in this case.

Based on prior biological knowledge we have to minimize the number of potential models fitting to our data (Lebreton et al. 1992, Lebreton et al. 2009). If we still have to choose one from many different models, or if the potential models are non-hierarchical, then we may choose the one fitting best to our empirical results on the basis of some information criterion like the *Akaike information criterion*, constructed from the likelihood values (L) and the number of estimated parameters (n):

$$\mathrm{AIC} = -2\ln L + 2n \qquad (2.3.3)$$

(Johnson and Omland 2004). In our example the more complicated model gives AIC = 21.31, whereas for the simpler one AIC = 19.75. As the smaller AIC is the better, this again suggests that the model assuming equal mortality for the two genders is the better choice.

If the difference between AIC values is not too high (according to a rule of thumb, not higher than two), we can combine the model's predictions by calculating their weighted mean (Johnson and Omland 2004). The Akaike weights are calculated from the AIC values and they are low for models with high AIC. Model with AIC much higher than the most parsimonious model has so low Akaike weight that its prediction would have negligible effect on the weighted mean of predictions. Therefore, such models can be regarded as unsupported by the data. Low Akaike weights of the best fitting models of SADs (species area distributions) for pooled data in the vast majority of subplots provided the basis of the refutation of a putative evidence for the neutral theory of biodiversity (TBE, p. 249).

References

https://en.wikipedia.org/wiki/Likelihood-ratio_test

Burnham, K.P. and Anderson, D.R. (2013). *Model Selection and Inference: A Practical Information-Theoretic Approach*. New York, Springer New York.

Johnson, J.B. and Omland, K.S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19(2): 101-08.

Lebreton, J.-D., Burnham, K.P., Clobert, J. and Anderson, D.R. (1992). Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies. *Ecological Monographs*, 62(1): 67-118.

Lebreton, J.-D., Nichols, J.D., Barker, R.J., Pradel, R. and Spendelow, J.A. (2009). Modeling Individual Animal Histories with Multistate Capture–Recapture Models. In H. Caswell, ed. *Advances in Ecological Research*, pp. 87-173. Burlington, Academic Press.