Fit of Isolation by Distance Models: Curve Fitting

I.

Italo

Barrai

Istituto di Zoologia, Università di Ferrara, Italia

Dr. I. Barrai, Istituto di Zoologia, Università di Ferrara, Via Borsari 46, I-44100 Ferrara (Italy)

The recent work of Barbujani and Canella [1987] on the fit of an isolation by distance model to several gene frequencies was in many ways a laudable exercise in the description of population structures at large distances. However, it includes some considerations derived by parameters of isolation by distance which are not warranted by the estimation methods used. These deserve a word of clarification.

The function describing the variation of expected kinship $\Phi(x)$ at distance x is:

$$\Phi(x) = (1-L)a \exp(-bx) + L, \quad (1)$$

where a is an estimate of local kinship, b is the rate of exponential decay of kinship, and L is either an ancillary constant permitting fit of a logarithmic regression, or an estimate of kinship at infinite distance. If L is chosen as an arbitrary constant, the relation between the observed sample kinship at distance x , say $r(x)$, and kinship conditional of L , is $\Phi(x)$ such that:

$$\Phi(x) = [r(x)-L]/(1-L), \quad (2)$$

which permits to test the hypothesis of an exponential decay of kinship with distance through:

$$\alpha + \beta x = \ln\{[r(x)-L]/(1-L)\} > 0, \quad (3)$$

and L must be chosen in such a way that no logarithms of negative numbers are met in the calculations. It is convenient to choose L smaller than the smallest kinship value observed among negative values; in any case, the choice is arbitrary. This regression indicates the presence of an exponential decay of kinship with distance, but of course does not give indications as to the value of kinship at infinite distance. The expected values of a , b and L (as kinship at infinite distance) estimated from the nontransformed swarm kinship-distance, each with one degree of freedom, can be obtained through numerical methods, possibly through a numerical Newton-Raphson method, minimizing the quantity:

$$Q = \sum [r(x) - \Phi(x)]^2. \quad (4)$$

The goodness of fit of function 1 so obtained can be tested by analysis of variance, with the regression sum of squares having three degrees of freedom. Then a and b can be used to estimate demographic parameters from isolation by distance, or can be compared to other values obtained with similar methods in other populations, provided that the same unit of distance was used throughout.

382

Barrai

It must be noted that $\exp(\alpha)$ and β in equation 3 may be quite different from the values of a and b obtained from minimization of equation 4, since they depend on the arbitrary value chosen from L . Therefore, the considerations advanced by Barbujani and Canella [1987] with

regard to isolation by distance, insofar as they refer to a and b , and to their comparisons with the same parameters estimated simultaneously in other populations, are not substantiated – nor would any considerations be if the two aspects of L are confounded.

Reference

Barbujani, G.; Canella, R.: Diversity of gene frequencies in European and Asian populations. II. Fit of an isolation by distance model. *Hum. Hered.* 37: 265-272(1987).