The conceptual framework of Malécot's model of isolation by distance*

BY J. M. LALOUEL

Population Genetics Laboratory, University of Hawaii, Honolulu 96822

Strong criticism of Malécot's model of isolation by distance was expressed in a recent paper (Felsenstein, 1975), which concluded that 'this model is therefore biologically irrelevant'. This conclusion was arrived at after purportedly demonstrating that some of Malécot's assumptions are incompatible. However, we shall dismiss these claims, showing that this author's interpretation of Malécot's writing is incorrect, his demonstrations affected by mathematical errors, and his simulations irrelevant.

MALÉCOT'S MAIN ASSUMPTIONS

To Felsenstein (1975, p. 359), 'Malécot appears to be making three assumptions: (I) individuals are distributed randomly on the line with equal expected density everywhere; (II) each individual reproduces independently, the number of offspring being drawn from a Poisson distribution with a mean of one; and (III) each offspring migrates independently, the displacement being drawn from some distribution m(x), which we will take to be a normal distribution.' This author then claims to demonstrate 'that assumption I is incompatible with assumption II and III, and that a model embodying II and III will lead to the formation of larger and larger clumps of individuals separated by greater and greater distances'.

Let us examine what actually are Malécot's assumptions, following the 1969 translation of his 1948 book, although departing from his notations in one important respect: we initially consider the case of a population made of discrete subdivisions, as this conceptual formulation is shared by most models of population structure, the 'continuous model' simply being a limiting case. Consider a finite population, with discrete generations, distributed over a finite area A, and subdivided into elementary areas ΔS_P centred at P, with coordinates (x, y) such that the density of individuals in elementary area ΔS_P, at generation n, is given by δ_n(P). Let us assume that each individual from the time of birth to the reproductive stage, has a known probability f_n(P, Q) ΔS_Q, of migrating from P to an elementary area, ΔS_Q, centred at Q[Σf(P, Q) ΔS_Q = 1].

According to Bayes’s formula, each parent of an individual born at point Q will have the known probability

\[ g_n(P, Q) ΔS_P = \delta_n(P)f_n(P, Q) ΔS_P / Σδ_n(P)f_n(P, Q) ΔS_P \] (1)

of being born in an area ΔS_P centred at P[Σg_n(P, Q) ΔS_P = 1]. In a model where, as we shall see, births and deaths are not explicitly considered, we can assume that we are solely concerned with the population of reproducers, that is (1) applies, with δ_n(P) being the density of reproducers in ΔS_P at generation n. f_n(P, Q) ΔS_Q and g_n(P, Q) ΔS_P, forward and backward probabilities with respect to the prospective or retrospective consideration of the process under study, are conditional probabilities, and implicitly reflect the combined effects of migration, birth, death, and fecundity.

* PGL Paper No. 156.
Three critical assumptions common to all published models of population structure will be now introduced. The first one consists in assuming that densities in all subdivisions are constant in time; that is, \( v_n, \delta_n(P) = \delta(P) \), the number of individuals in each subdivision is not random. Following Malécot, we shall later assume densities to be constant in space as well as in time. Felsenstein clearly misinterpreted Malécot in writing his assumption (I), which departs fundamentally from Malécot in that densities are random quantities. That Malécot assumes constant densities is clear in page 69 of his book (but also obvious when considering his mathematical derivations). This is particularly well stated in Malécot (1959, p. 175), although referring to one local population: ‘[On regardera] \( N \) comme une constante, ou bien comme une quantité dont la loi de variation au cours des générations est connue (ce qui se justifie toutes les fois que la concurrence écologique impose à l’effectif un niveau sensiblement déterminé, mais exclut les cas de “peuplement des places vides” ou de “populations trop peu nombreuses”).’ Although discussed at times (for example, Feller, 1951), this assumption is always made for genetic models of population structure. Most authors simply state this assumption; Moran (1962), however, gives particular attention to this matter in his chapter 1, and justifies the present assumption in a way similar to Malécot’s (Moran, 1962, p. 3): ‘the fact that most animal and plant populations do survive and do not cover the whole earth shows that natural density-dependent factors are exceedingly strong. Thus apart from seasonal variations many populations will not vary greatly about a definite mean level. Small variations about such a level will not have much influence on the overall conclusion drawn from population genetic models and therefore in most of the models considered in this book \( N \) will be held fixed and little is lost by this assumption.’ Reference to Matras (1961), Rogers (1968) or Le Bras (1971a, b) may convince the reader of the difficulty of studying the demographic evolution of a subdivided population by a stochastic process involving births, deaths, and migration (particularly with regard to conditions for the existence of a stationary state), and to make birth and death rates density-dependent would further complicate the problem.

A second assumption shared by many genetic models (although generally only implicitly) concerns the sampling scheme of the gametes which contribute to the next generation. Malécot’s scheme of Bernoulli trials among an infinite gametic pool (Malécot, 1969a, pp. 52–3) was most clearly presented in another paper (Malécot, 1959, pp. 174–5): ‘Notre hypothèse revient donc à admettre que, aussi longtemps qu’il y a des tirages de gamètes, tous les individus de même génotype de la génération considérée, y participent également, quel que soit le nombre de gamètes utiles qu’ils ont déjà fourni; cela entraîne que le nombre de gamètes utiles produit par chacun de ces individus, est des variables aléatoires indépendantes, obéissant à des lois binomiales très voisines pratiquement de lois de Poisson.’ That this assumption turns out to be a realistic hypothesis is found in its agreement with the frequent assumption of a Poisson law of fecundity in many models. This is again an approximation of reality, but other schemes would no doubt considerably complicate the formulation of a model (see, for example, Rothman, Sing & Templeton, 1974). In view of Malécot’s assumption of constant density, however, this random aspect of natality appears only implicitly in his model, introducing stochastic independence between the ‘gamètes utiles’ produced by a same parent.

A third assumption is to consider the probabilities of migration constant in time. Together with the previous assumptions, as all individuals have the same probability of effectively reproducing, this implies that \( f_n(P, Q) = f(P, Q) \) for all \( n \).
Malecot's model of isolation by distance

Fundamental recurrence

Let us now present Malécot's fundamental recurrence, to which most other genetic models of population structure are related. Apart from the change of notation mentioned earlier, we follow Malécot (1969a, p. 67): two loci of two individuals in the \((n+1)\)th generation in \(C\) and \(D\) will have the probability \(g(E, C) g(F, D) \Delta S_E \Delta S_F\) of coming from parents in \(\Delta S_E\) and \(\Delta S_F\) and the probability \(g(E, C) g(E, D) (\Delta S_E)^2\) of coming from parents both reproducing in elementary area \(\Delta S_E\); in the latter case, they will have the conditional probability \(1/[2\delta(E) \Delta S_E]\) of coming from the same locus of the same parent and the probability, \(1 - 1/[2\delta(E) \Delta S_E]\) of coming from loci infinitely close but distinct. Mutation can be introduced through the concept of betamutation (Malécot, 1969b, p. 238), \((1 - k)^2\) denoting the probability of persistence of identity on a link of a kinship chain. Denoting by \(\phi_{n+1}(C, D)\) the probability of identity of two loci of two individuals in \(C\) and \(D\) at the \((n+1)\)th generation, we have, analogous to Malécot (1969a, p. 68, 3.3.1) for discrete subdivisions, the fundamental recurrence:

\[
\phi_{n+1}(C, D) = (1 - k)^p \sum \phi_a(E, F) g(E, C) g(F, D) \Delta S_E \Delta S_F + (1 - k)^p \left[1 - \phi_a(E, E)/2\delta(E) \Delta S_E\right] g(E, C) g(E, D) \Delta S_E.
\]

Noting that \(N(E) = \delta(E) \Delta S_E\), discontinuous migration will be treated by \(g^*(E, C) = g(E, C) \Delta S_E\) for any \(E, C\). Malécot’s migration matrix approach (1950, p. 48, formula (10)) is obtained from (2) by neglecting terms of order \(k^2\phi(\cdot)\), his formula (10) being itself an approximation of his formula (8) on the same page. Bodmer & Cavalli-Sforza’s (1968) or Maruyama’s (1970, 1971) main recurrences differ only by the way genetic similarity is studied, the degree of approximation involved, or assumptions concerning migration, but otherwise admit the same conceptual framework. Steppingstone migration (Kimura & Weiss, 1964; Maruyama, 1970; 1971) is a particular case of discontinuous migration that has been previously considered by Malécot (1950, p. 50). Models of isolation by distance derive from (2) through additional assumptions.

Isolation by distance, continuous approximation

Although Malécot referred to both discontinuous and continuous migration models, the distinction is more of a numerical nature than a fundamental one concerning the population distribution in space, as will be seen. First presented by Malécot (1948), we read in the English translation (Malécot, 1969a, p. 69): ‘let us suppose that the area occupied by the population can be considered unlimited, that the density \(\delta(E)\) is constant (in space and time), and that \(g(P, Q)\) depends only on the distance \(PQ = r\); then \(g(P, Q)\) is equal to \(f(P, Q)\)’. It follows that \(g(P, Q) = f(P, Q) = f(r)\); that is, migration is homogeneous and isotropic. For the unidimensional case (considered by Felsenstein, 1975), and with Malécot’s (1962) notations (to which Felsenstein’s bear some resemblance, see Fig. 1), equation (2) leads in the present case to the approximation:

\[
\phi_{n+1}(x, \beta) = \left[\left[1 - \phi_n(\lambda, \lambda)/2\delta\right] g(\lambda, x) g(\lambda, \beta) d\lambda \right. \\
+ \left. \int \phi_n(\mu, \lambda) g(\mu, x) g(\lambda, \beta) d\mu d\lambda, \right.
\]

or, when written in terms of distances (see Fig. 1):

\[
\phi_{n+1}(x) = \left[\left[1 - \phi_n(0)/2\delta\right] g(x + z) g(z) dz \right. \\
+ \left. \int \phi_n(x - y + z) g(y) g(z) dy dz. \right.
\]
That it is an approximation is clear: under our assumption of constant density $\delta$ (in time and space), the smallest elementary area (here unidimensional) that can be defined is such that $N(\lambda) = 1$, hence $\Delta S_\lambda = 1/\delta$, $\forall \lambda$, and the notation $d\lambda$ in (3) is an approximation in the sense that we cannot have $\Delta S_\lambda$ tending to zero because $\delta$ is assumed finite (as is made clear by Malecot, 1972, p. 404, note 2).

Results on the existence of a stationary state and solution, at the stationary state, of this recurrence under various assumptions (concerning mating, dimensionality, migration distribution, etc.) have been presented in various papers by Malecot, and his 1975 paper will provide the reader interested with the most complete, integrated treatment of this subject. Consideration of Malecot's conceptual framework only is necessary to answer Felsenstein's criticism, to which we shall turn now.

REFUTATION OF FELSENSTEIN'S CRITICISM

Felsenstein clearly misunderstood Malécot's assumption concerning density. As we have seen, Malecot assumes density to be constant in time and space, hence it belongs to the domain of certainty, and therefore cannot be evaluated by a probability: it is a parameter in the model. Felsenstein's consideration of a random, uniform distribution of individuals is enough of a departure from Malecot's assumption to allow dismissing Felsenstein's criticism as off target. Migration being only a function of distance, homogeneous and isotropic migration between a set of $m$ subdivisions determines the matrix of transition probabilities of a Markov chain of order one. This matrix being bistochastic and irreducible, the probabilities of occupancy of any subdivisions, at the stationary state, are independent of the initial conditions and all equal to $1/m$. This consideration would hold for the infinite case by use of a limiting argument (through the concept of finite additivity of de Finetti, 1974). Hence the assumption of constant density, homogeneous and isotropic migration, and a stationary state, is perfectly coherent. When applying this model to gene frequency covariances, selection is assumed constant and linear on the gene frequency. This is consistent with the uniform influence of environment implied by constant density in space and time as an approximation of the stationary state of a density dependent births and deaths process.

Further on, Felsenstein demonstrates 'that assumption I is incompatible with assumption II and III, and that a model embodying II and III will lead to the formation of larger and larger clumps of individuals separated by greater and greater distances'. Attempting to study 'the probability of joint occupancy of two locations at distance $x$', Felsenstein writes the recurrence (with the notations of Fig. 1):

$$f_{t+1}(x) = \int m(z) m(x-z) \rho \, dz + \int \int m(z) m(z-w) f_t(x-w) \, dz \, dw.$$  

(4)
It is not clear what $f_{t+1}(x)$ stands for; to disprove Felsenstein’s subsequent results, we simply show that (4) is at error. As can be seen in Fig. 1, infinitesimals of the double integral should be $dz$ and $d(z - w)$; the variables $z$ and $w$, on which integrations are carried out, are not independent. Writing $u = z - w$, this double integral becomes $\int m(z)m(u)f_{t}(x - z + u)dz
dw$; this is not a simple convolution, and Felsenstein’s subsequent derivation is erroneous. We can therefore dismiss his results, including his consideration of identity by descent (p. 364), affected by the same error.

As for the simulation on pages 363–4, it again does not correspond to Malécot’s model in assuming random density of individuals. For in Felsenstein’s model where all individuals are stochastically independent and where any individual has a constant, non-zero probability of not reproducing, each line will eventually become extinct, so that, as long as the population is initially of finite size, it will become extinct after an indefinite number of generations. The result presented is a transient state of the population, not a stationary state.

Other confusions occur on page 360, in connexion with forward and backward probabilities of migration and in relation to the continuous approximation, as the reader will find. Our last remark will be that it is not true that ‘steppingstone’ models are exempt from the criticism directed at Malécot’s model of isolation by distance (p. 367), as they both share the assumption of constant density.

**CONCLUSION**

Felsenstein’s criticism stems from a misinterpretation of what are Malécot’s assumptions, and we have shown that his contentions were not supported by his subsequent demonstrations and simulations. It should be unnecessary to say that the present discussion centred on only a particular aspect of Malécot’s work on isolation by distance, various situations and numerous extensions having been considered by this author (see Malécot, 1975). We hope we have brought to light the ingenuity and coherence of this work of one of the most original contributors to the theory of population genetics.

**SUMMARY**

The conceptual framework of Malécot’s model of isolation by distance is presented, and recent criticisms of this model are dismissed, as they originate in misinterpretation of Malécot’s work and are supported by mathematical derivations in error.

We are grateful to Gustave Malécot for interesting comments made on an earlier version of this paper. This work was supported by grants GM 17173, HD 06003 from the U.S. National Institutes of Health.

**REFERENCES**


