The Ising Model for Population Biology

Hirotugu MATSUDA

Department of Biology, Faculty of Science, Kyushu University
Fukuoka 812

(Received May 8, 1981)

The Ising model is introduced in population biology. The exact stationary distribution is obtained for the one-dimensional nearest-neighbor dispersal case in the presence of mutation and selection. Discussion is given for the feature of genetic correlation in general cases.

In statistical physics macroscopic properties of matter consisting of many particles are deduced from assumptions on microscopic properties of constituent particles, and sometimes the underlying microscopic properties are induced by comparison between theory and experiment on macroscopic properties. In this respect theoretical population biology is similar to statistical physics, if we regard appropriate constituents of population as something like particles. In population genetics genes or chromosomes are usually regarded as constituents, and we are concerned in how genetic structure of the population depends on relevant parameters characterizing properties of each gene in a certain genetic state. In population ecology individual organisms are usually taken as appropriate constituents. Both in population genetics and in ecology the constituent is an entity capable of replicating itself and of suffering death. Therefore, in order to study population biology from a unified point of view as an extension of statistical physics, let us broadly call the appropriate constituent of the population ‘replicon’.

For geographically structured population we ask how the structure will depend on relevant parameters such as range and rate of dispersal, mutation rate and fitness or Malthusian parameter characterizing each replicon. Although various models have already been proposed, no comprehensive answer can yet been obtained. In this short note we propose a lattice model to study the above problem. We show that the canonical distribution of the corresponding Ising model will give an insight into the stationary geographical structure.

Consider the lattice space consisting of \( M \) lattice sites, which represent habitable places of replicons. The habitable places in nature may form a continuous space. But as in a lattice gas model the artificial introduction of the lattice allows us to simplify the description of such situation that the habitat must have a limited carrying capacity for replicons in the following way: The state of \( i \)-th site \((i = 1, 2, \ldots, M)\) is specified by the variable \( \sigma_i \in \{+, -, 0\} = S \), where the state \( +, - \) represents the \( i \)-th site being occupied by the replicon of + and \(-\) genetic state, and by no replicon, respectively. Thus, the state of the whole space (total state) is specified by \( \vec{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_M) \subseteq S^M \).

We assume that \( \vec{\sigma} = \vec{\sigma}(t) \) is a certain Markov process. The process is characterized by the transition probability from state \( \vec{\sigma} \) to \( \vec{\sigma}' = (\sigma_1', \sigma_2', \ldots, \sigma_M') \subseteq S^M \) due to extinction, invasion and mutation process constituting the whole process. They are respectively given in the following form:

\[
P^{\text{ext}}(\vec{\sigma}' | \vec{\sigma}) = \prod_{i=1}^{M} [\epsilon(\sigma_i | \vec{\sigma})] \delta_{\sigma_i; 0} + 1 - \epsilon(\sigma_i | \vec{\sigma})] \delta_{\sigma_i; \sigma_i'},
\]

\[
P^{\text{inv}}(\vec{\sigma}' | \vec{\sigma}) = \prod_{i=1}^{M} [\eta(\sigma_i' | \vec{\sigma})] \delta_{\sigma_i; \sigma_i'},
\]

(1)
Here \( \varepsilon_i(\alpha | \bar{\alpha}) \) is the probability that the replicon of state \( \alpha \) occupying \( i \)th site is extinguished: \( \alpha' = 0 \). When there is no replicon at \( i \)th site there can be no extinction: \( \varepsilon_i(0 | \bar{\alpha}) = 0 \). For no extinction of \( i \)th site occurring with probability \( 1 - \varepsilon_i(\alpha | \bar{\alpha}) \) we must have \( \alpha' = \alpha \). The invasion to the \( i \)th site by a replicon is possible only when \( i \)th site is vacant, so that we must have \( \alpha'_i = \alpha_i \) for \( \alpha_i \in \mathbb{S} = \{+, -\} \) in the invasion process. \( q_i(\alpha | \bar{\alpha}) \) is the probability that vacant \( i \)th site is invaded by a replicon of state \( \alpha \in \mathbb{S} \), or that no invasion occurs \( (\alpha'_i = 0) \). From definition we have

\[
\sum_{\alpha \in \mathbb{S}} q_i(\alpha | \bar{\alpha}) = 1. 
\]

Finally \( \mu \) is the mutation rate from state \( \alpha_i \) to \( -\alpha_i \).

From these elementary processes we may construct the whole process in any way. For instance, assuming that the above three processes occur cyclically, we may get the sample process of \( \alpha(t) \) by a computer simulation, as was done for the kinetic Ising model.\(^7\)

As a first step for analytical study, we consider the simple case in which the invasion occurs as soon as the extinction has taken place, so that the total state space is reduced from \( \mathbb{S}^M \) to \( \mathbb{S}^M \). We assume the following equation for the time development of the probability \( P(\bar{\alpha}) = P(\bar{\alpha}, t) \) that the total state is \( \bar{\alpha} \) at time \( t \):

\[
\frac{dP(\bar{\alpha})}{dt} = \sum_{\bar{\alpha}' \in \mathbb{S}^M} (\varepsilon_{\bar{\alpha}')(\bar{\alpha})' P((\bar{\alpha}')')) - P((\bar{\alpha}')|\bar{\alpha}) P(\bar{\alpha}),
\]

where

\[
(\bar{\alpha}') = (\alpha_1, \alpha_2, ..., \alpha_{i-1}, -\alpha_i, \alpha_{i+1}, ..., \alpha_M), \quad (\bar{\alpha})' = (\varepsilon_{\bar{\alpha}')(\bar{\alpha})') + \mu_{-\alpha_1},
\]

\( q_i(\alpha | \bar{\alpha}) = \sum_{\bar{\alpha} \in \mathbb{S}^M} m_{i\bar{\alpha}}(1 + \alpha_i \bar{\alpha})/2. \)

\( \varepsilon_{\bar{\alpha}')(\bar{\alpha})' \) is the extinction probability per unit time of replicon of \( i \)th site in state \( \alpha \), and \( m_{i\bar{\alpha}}(= m_{\bar{\alpha}i}) \) is the probability that the \( i \)th site is invaded from \( \bar{\alpha} \)th site \( (\sum_{\alpha \in \mathbb{S}} m_{i\bar{\alpha}} = 1) \). The Malthusian parameter\(^\ast\) of the replicon of \( i \)th site is given by

\[
m_i(\bar{\alpha}) = -\varepsilon_{\bar{\alpha}'}(\bar{\alpha})' + \sum_{\bar{\alpha} \in \mathbb{S}^M} m_{\bar{\alpha}i} \varepsilon_{\bar{\alpha}'}(\bar{\alpha})'.
\]

Let us consider the stationary state of (5). In the particular case where \( \mu \) is proportional to \( \varepsilon_{\bar{\alpha}}(\bar{\alpha})' \), \( (i = 1, 2, ..., M) \), such that

\[
\mu = \bar{\mu} \varepsilon_{\bar{\alpha}}(\bar{\alpha})' \quad (\sigma \in \mathbb{S}, \bar{\mu}: \text{positive constant})
\]

the stationary state is represented by the following canonical distribution \( P^{eq}(\bar{\alpha}) \) of the ferromagnetic Ising model:

\[
P^{eq}(\bar{\alpha}) \propto \exp\left[ J \sum_{\alpha, \bar{\alpha}} m_{\alpha \bar{\alpha}} \alpha \bar{\alpha} + \sum_{i} H_i \bar{\alpha}_i \right],
\]

if a constant \( J \) satisfies

\[
z = (1 + 2\bar{\mu}) \tanh(2J) \quad \text{for any } z \text{ given by}
\]

\[
z = \sum_{\alpha, \bar{\alpha}} m_{\alpha \bar{\alpha}}. \quad (\bar{\alpha} \in \mathbb{S}^M)
\]

In (11), \( \sum_{\alpha, \bar{\alpha}} \) denotes the summation over all different pairs of sites \( i \) and \( j \), and

\[
H_i = \log(\varepsilon_{\bar{\alpha}'}(\bar{\alpha})'/\varepsilon_{\bar{\alpha}'}(\bar{\alpha}))
\]

the stationary state, denoting by \( \langle \cdot \rangle \) the average of \( \cdot \), we can define the genetic correlation between \( i \)th and \( j \)th sites as

\[
\bar{f}_{ij} = \langle \alpha_i \bar{\alpha}_j \rangle - \langle \alpha_i \rangle \langle \alpha_j \rangle /
\]

\[
1 - \langle \alpha_i \rangle \langle \alpha_j \rangle .
\]

Generally there is no constant \( J \) satisfying (12). However, in the one-dimensional torus lattice with nearest neighbor invasion; \( m_{i\bar{\alpha}} = 1/2(\delta_{i,j+1} + \delta_{i,j-1}) \), we have \( z \in \{ \pm 1, 0 \} \), so that (12) is satisfied by

\[
J = \frac{1}{2} \log(1 + \bar{\mu}/\bar{\mu}).
\]

\(^\ast\) The instantaneous rate of increase of the expected number of the replicon plus its progenies per replicon per unit time.
and \(P^{eq}(\bar{\sigma})\) represents the exact stationary state. Assuming the uniformity of the lattice: \(H_i = H_o\) \((i = 1, 2, \ldots, M)\), we obtain from (11) and (15) the genetic correlation:

\[
f_{ij} = \gamma^{r_{ij}} - \beta,
\]

(17)

where

\[
\gamma = \frac{\cosh H - \sqrt{\sinh^2 H + e^{-2H}}}{\cosh H + \sqrt{\sinh^2 H + e^{-2H}}}.
\]

(18)

Denote by \(<\bar{\sigma}\>_t^0\) the average of \(\bar{\sigma}\) under the condition \(\sigma_i = \sigma\), and the selective advantage \(s\) of the + state is given by

\[
s = \langle m_i(\bar{\sigma})\rangle^0 - \langle m_i(\bar{\sigma})\rangle^0,
\]

(19)

while the root mean square dispersal distance \(d\) per unit time is given by

\[
d^2 = a^2 \sum_{\bar{\sigma}, \bar{\sigma}'} m_{i,j} \langle \bar{\sigma}\rangle^0 \langle \bar{\sigma}'\rangle^0 |j - i|^2.
\]

(20)

where \(a\) is a lattice spacing.

By (11) \(\rightarrow (20)\) we obtain for \(|s| < 1\) and \(\bar{\mu} < 1\) the genetic correlation \(f(r)\) between sites separated by distance \(r\) as

\[
f(r) = e^{-r_{e} r},
\]

(21)

where

\[
(r_{e}/d)^2 = \frac{1}{2} \frac{1}{\mu + \sqrt{\mu^2 + (s/2)^2}}.
\]

(22)

Therefore, the genetic correlation distance \(r_{e}\) is roughly equal to the distance traversed in the time interval \(\text{Min}(s^{-1}, \mu^{-1})\) by a random walker whose root mean square dispersal distance is \(d\) per unit time. This corresponds to the extreme case of short range dispersal.

In the long range dispersal limit without selection, we shall have \(s = 0\) in (13) for most cases, so that from (12) we may effectively put

\[
J = (1 + 2\bar{\mu})^{-1}.
\]

(23)

In this case we have \(J < 1\), so that the correlation distance will be at most of the order of the dispersal range. For the intermediate range dispersal case in the uniform lattice, an effective \(z\) may be given by \(z_{e} = \sum_{i=1}^{i} m_{ij} \langle \bar{\sigma}\rangle^0\) and the effective \(J\) may be determined from (12) for \(z = z_{e}\) in a self-consistent way. Although at a moment it is not clear to what extent such an approximation is valid, we hope that the lattice model is useful for further study of population biology.

The author is grateful to Dr. J. Felsenstein for sending reprints of his publications, which have stimulated him to this study. He thanks Professor M. Toda and Dr. K. Ishii for valuable comments.

1) For the review of papers published before 1976:
5) For the review:
C. Domb and M. S. Green, eds., Phase Transitions and Critical Phenomena (Academic press), vol. 1 (1972); 2 (1972); 3 (1974); 5A (1976); 5B (1976); 6 (1976).