Journal of Statistical Physics, Vol. 30, No. 3, 1983

Statistical Mechanics and Population Biology

Lloyd Demetrius[1]

Received February 14, 1980; revised May 14, 1982

This paper exploits the connection between statistical mechanics and stochastic processes in order to derive a class of macroscopic observables for populations. This review treats the dynamics of populations in both constant and variable environments and derives in each case the thermodynamic analogs of the population parameters.

KEY WORDS: Entropy; Gibbs measures; phase transition; Malthusian parameter; adaptive value; density dependence; effective size.

1. INTRODUCTION

A large class of biological systems, like systems which statistical mechanics studies, consists of a large number of interacting units. Determining the future behavior of these systems generates difficulties analogous to those raised in the analysis of the microscopic behavior of a gas. A population, for example, consists of a large number of individuals of different ages and different genotypes. These individuals are also subject to different kinds of environmental interactions. In order, for example, to predict the time of death of a particular individual, we need to know the individual's age, his genetic constitution, and the environmental circumstances he will confront. These characteristics are hard to determine precisely. In addition, as the number of individuals in the population increases, the characterization of the state space becomes less tractable. The methods of classical mechanics are not adequate to deal with population phenomena in which so many different variables interact. Statistical methods abandon the attempt to

---

[1] Max–Planck–Institut für biophysikalische Chemie, D-3400 Göttingen, Federal Republic of Germany.

0022-4715/83/0300-0709$03.00/0

determine the precise demographic and genetic trajectory of each individual of any age in any area who can be expected to survive to a certain period. It seeks in its stead macroscopic measures which describe the average behavior of the system. Thus the mean life expectancy of individuals in a given environment emerges as a macroscopic parameter which provides, along with other parameters, a correct basis for the prediction of the behavior of the population.

The aim of this paper is to use the methods of statistical physics to provide a systematic basis for the derivation of macroscopic observables for population models. This paper revolves around two statistical notions, entropy and adaptive value, both of which have precise interpretations in a population context.

The entropy concept we introduce is based on the Kolmogorov–Sinai invariant. This invariant has played an important role in recent studies on the asymptotic statistical properties of models in statistical mechanics. Dobrushin(1) discovered that the equilibrium states of a class of models in statistical mechanics can be described as the equilibrium states of certain time evolutions of a Markov process. Bowen and Ruelle,(2) exploiting ideas due to Sinai,(3) have shown that the equilibrium states of certain dynamical systems can be described by a variational principle. The work of these authors have brought statistical mechanics in close contact with the theory of stationary stochastic processes and suggests the possibility of finding analogues of thermodynamic concepts in other probabilistic and dynamic contexts. In Ref. 4, I exploited the connection between statistical mechanics, stochastic processes, and dynamical systems to provide a new model of an evolutionary process. On the basis of this model, one could make precise sense of the notion that certain populations have a more complex life-history than others. One was able to assign a number to these populations; this number is called entropy since its derivation is based on the Kolmogorov–Sinai invariant.

The notion adaptive value of a population generalizes the entropy concept. This new notion has its origins in information theory. Shannon,(5) in his efforts to determine at what rate one can transmit messages through a noisy channel, extended his information measure to a measure that described the correlation between the variability of the information source and the randomness of the channel. In Ref. 6, I recognized an analogy between the engineering problem of coding messages for efficient communication in a channel that causes errors in transmission, and the evolutionary problem of a population adapting its birth and death rates in order to survive in an environment that randomly affects its demography. On the basis of this analogy, one could make precise sense of the notion that certain populations have a fecundity and mortality schedule that is correlated with the environmental variation. A number was assigned to these populations. This new parameter is called adaptive value(6,7) on account of the fact that it measures the capacity of a population to survive and reproduce under random environmental conditions, that is, the degree to which the population is adapted to the environment.

The ideas of entropy and adaptive value have been developed in a series of papers addressed to population biologists.(8–11) This paper is addressed to physicists. Here I emphasize the connection between statistical mechanics, dynamical systems, and information theory in order to bring out the ideas that underlie the entropy and adaptive value concepts.

This paper is organized as follows. Section 2 describes and contrasts two types of population models, the classical demographic model due to Leslie(12) and Lotka(13) and a new model based on ideas of statistical mechanics, Ref. 4. We distinguish between density-independent and density-dependent models. Density-independent models are discussed in Section 4. In this section we make explicit the connection between thermodynamic parameters and the population variables. The principal analogs are: the free energy corresponds to the Malthusian parameter or population growth rate; the temperature corresponds to the reciprocal of the generation time. In Section 5 we analyze a class of density-dependent models and we discuss the relationship between multiple equilibrium states which these models exhibit and the phenomena of phase transitions in statistical physics. Section 6 distinguishes between various measures of population size and introduces the concept of effective size. We relate this parameter to the population entropy. The effect of environmental interactions on the population is analyzed in Section 7. This section revolves around the notion of adaptive value, a natural generalization of the entropy concept.

2. POPULATION MODELS

2.1. Representation as Age Distributions

We describe two demographic models, the first due to Leslie,(12) the second more recently introduced in Ref. 4. Both models only consider the female population and assume that there are enough males not to alter the birth or death rates as a function of age of the females we are studying. In the Leslie model, the phase space is the set of all age distributions. Equilibrium states are described by age distributions which remain constant in time. In the second model, the phase space is the set of all genealogies. Equilibrium states are described by shift-invariant measures on the space of genealogies. These probability measures are described by a variational principle. We shall now discuss the relationship between these two demographic models. The relation between the two notions of equilibrium states and the derivation of the thermodynamic analogs of the population parameters will be developed in Sections 3 and 4.

In the Leslie model, we treat time and age as discrete. The population is partitioned into age groups corresponding to the unit intervals of time. We shall assume that there are n age groups. We then define $x_i(t)$ to be the number of females in age group (i) at time (t). The proportion of females in age group (i) at time (t) surviving to be in age group $(i + 1)$ at time $(t + 1)$ is $b_i(t)$, which for $i \leqslant n - 1$ is strictly positive. Furthermore, $m_i(t)$ denotes the average number of daughters born per female to females in age group (i) at time (t), these daughters surviving to be in age group (1) at time $(t + 1)$.

Using these definitions, we have the following matrix equation:

$$\bar{x}(t+1) = A(t)\bar{x}(t) \tag{2.0}$$

where

$$A(t) = \begin{bmatrix} m_1(t) & m_2(t) & & & m_n(t) \\ b_1(t) & 0 & & \cdots & 0 \\ 0 & b_2(t) & & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & b_{n-1}(t) & 0 \end{bmatrix}$$

We shall assume that for each (t), $m_{n-1}(t) > 0$, $m_n(t) > 0$. A consequence of this assumption is that the matrix $A(t)$ for each (t) is irreducible and primitive. These restrictions are not the weakest condition for primitivity to hold. The important property of these conditions is that they are satisfied by a large class of real populations whose age structure is truncated after the last age with positive fecundity.

We now describe a notion of equilibrium for the dynamical system given by the Leslie model. This notion of equilibrium was first introduced by Lotka in connection with the continuous analog of the matrix model.

The population is said to be at *Lotka equilibrium* if for some positive real number λ and for all $t > t_0$,

$$\bar{x}(t+1) = \lambda\bar{x}(t) \tag{2.1}$$

Thus, at Lotka equilibrium, the relative number of individuals in each age group remains constant in time.

The case of a constant fecundity m_i and a constant mortality b_i is of particular interest. When this condition holds, an equilibrium state exists and is unique.

This fact follows immediately from the Perron–Frobenius theorem. Since the matrix $A \geqslant 0$ is irreducible, there exist vectors $\bar{u} = (u_i) > 0$, $\bar{v} = (v_i) > 0$ associated with the dominant eigenvalue λ such that

$$A\bar{u} = \lambda\bar{u} \qquad \bar{v}\mathrm{A} = \lambda\bar{v} \qquad (\bar{u}, \bar{v}) = 1 \tag{2.2}$$

The eigenvector $\bar{u}$ is the equilibrium state for the dynamical system given by (2.0).

There exist explicit expressions for the vectors $\bar{u}, \bar{v}$.

Let

$$l_j = \begin{cases} 1, & j = 1 \\ \prod_{r=1}^{j-1} b_r, & j > 1 \end{cases}$$

The vectors $\bar{u} = (u_i)$ and $\bar{v} = (v_i)$ are given by

$$u_i = l_i / \lambda^i \tag{2.3}$$

$$v_i = \left(\sum_{j=i}^{n} m_j u_j \right) \Big/ T u_i \tag{2.4}$$

where

$$T = \sum_{j=1}^{n} j m_j u_j \tag{2.5}$$

The element v_i is equal to the discounted future births to a female aged i. This element is called the reproductive value of a woman in age group i.

The expression T describes the mean age of parents of all newborn females when the equilibrium age-distribution is attained. This number T is called the generation time.

The dominant eigenvalue λ is the unique positive real root of the equation

$$1 - \sum_{j=1}^{n} \frac{l_j m_j}{\lambda^j} = 0 \tag{2.6}$$

The matrix A is also primitive. This implies that

$$\lim_{k \to \infty} \frac{A^k \bar{x}}{\lambda^k} = c\bar{u} \tag{2.7}$$

where $\bar{x} = (x_i) \geqslant 0$ is any initial age-distribution. This result means that the population converges towards the equilibrium state given by the vector $\bar{u}$.

Now consider the expression $N_t = \sum_{i=1}^{n} x_i(t)$, where $\bar{x}(t) = \{x_i(t)\}$ denotes the age-distribution at time t. The number N_t describes the population size at time t. Using (2.7), we can show that

$$\lim_{t\to\infty} \frac{1}{t} \log N_t = \log\lambda \tag{2.8}$$

Thus

$$N_t \sim e^{rt} \tag{2.9}$$

where

$$r = \log\lambda$$

The number r, called the Malthusian parameter, thus represents the asymptotic growth rate of the population.

2.2. Representation as Genealogies

We shall now describe a model which focuses on what we call the genealogies of the individuals in the population, Ref. 4. In this model, time and age are also discrete and we partition the population into n age groups. We assume as before that $m_i(t)$ is the fecundity of individuals in the ith age group, the offspring of these individuals surviving to be in the first age group at time $(t+1)$. We also assume that the proportion of individuals surviving from the (i)th to $(i+1)$th age group at time (t) is $b_i(t)$.

The evolution of the population starting from a single individual age (i) can be represented by the process

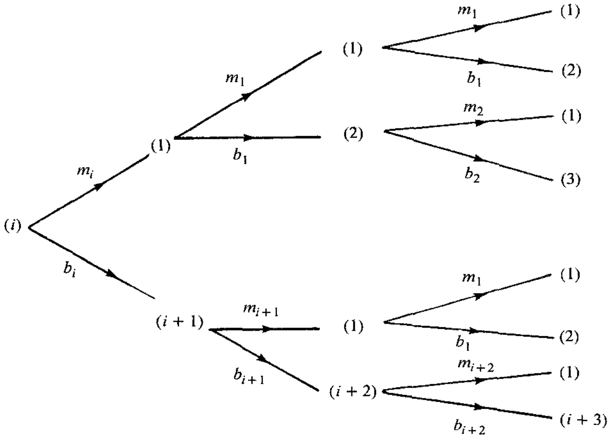


Fig. 2.0.

This process can be summarized by the graph G

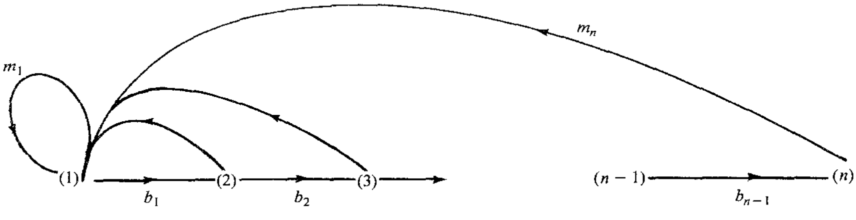


Fig. 2.1.

where

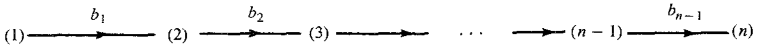


Fig. 2.2.

denotes the aging process and

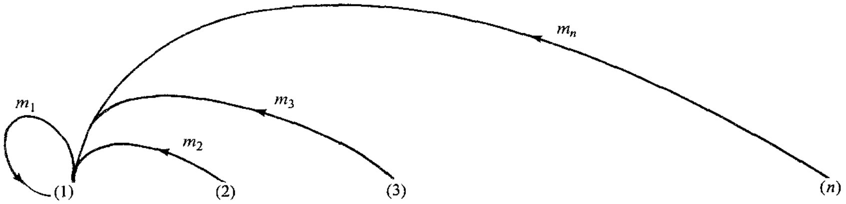


Fig. 2.3.

denotes the reproductive process.

Let Ω denote the set of all doubly infinite paths of the graph G. The set Ω is the phase space or configuration space of our system

An element $x \in \Omega$ is a doubly infinite sequence

$$(\ldots x_{-1}, x_0, x_1 \ldots)$$

where the x_i can assume values between 1 and n. We shall call each element $x \in \Omega$ a genealogy.

Let

$$T\colon (x_k) \to (x'_k)$$

where

$$x'_k = x_{k+1}$$

The equilibrium measures we shall consider will be probability measures that are invariant under the transformation T.

Example 2.0. Before we formalize the notion of equilibrium state for this model, we shall illustrate the distinction between the Leslie model and the new model we have described by discussing a simple example.

We consider the model described by the following graph:

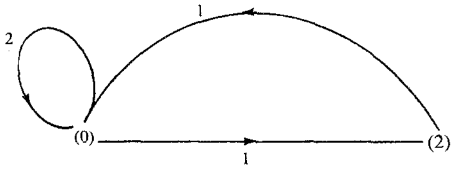


Fig. 2.4.

In this model, individuals in age group (1) produce two offspring. All individuals in this age group survive to age group (2). Individuals in the second age group produce a single offspring and die.

The set of genealogies Ω are described by the set of all paths of the above graph. For example, the element

$$\dots 1\ 1\ 1\ 1 \dots$$

belongs to Ω.

This genealogy describes an individual age (1) at time (t), the individual's daughter at time ($t+1$), the individual's granddaughter at time ($t+2$), and so on.

The sequence

$$\dots 1\ 2\ 1\ 1 \dots$$

also belongs to Ω. This sequence describes an individual age (1) at time (t), the same individual aged (2) at time ($t+1$), the daughter of this individual at time ($t+2$) and so on.

Now the genealogies generated by a single individual aged (1) at time (t) are given by

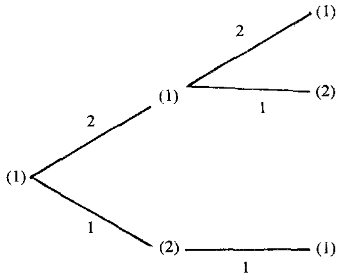


Fig. 2.5.

We observe that the age distribution at time t can be obtained from the cross sections of the genealogical tree at each instant t. The change in age distribution at time $t = 0$, $t = 1$, $t = 2$, is given by

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \to \begin{bmatrix} 2 \\ 1 \end{bmatrix} \to \begin{bmatrix} 5 \\ 2 \end{bmatrix} \to$$

This change can be immediately derived by inspecting the genealogies in Fig. 2.5.

3. THERMODYNAMIC FORMALISM AND POPULATION MODELS

We now return to the general model described in Fig. 2.0. In order to analyze our new model from the point of view of statistical mechanics, we will consider a new configuration space obtained by reversing the arrows in the graph G in Fig. 2.1 and considering the set of all paths of this new graph. Our analysis will be in terms of this new configuration space which we denote Ω despite the possible ambiguity. We now introduce a suitable topology on the space Ω.

Let $S = (1, 2, \ldots, n)$. We assign S the discrete topology. Let

$$X = \prod_{i \in Z} S_i$$

where $S_i = S$, that is X is the set of all functions from the integers Z to the set S.

We give X the product topology. The configuration space Ω, which is a subspace of X, is given the same topology as X. It is important to note that a metric on Ω compatible with the topology is given by

$$d(\{x_n\}, \{y_n\}) = \sum_{n=-\infty}^{\infty} \frac{|x_n - y_n|}{2^{|n|}} \tag{3.0}$$

Consider a continuous function $\Phi : \Omega \to R$. This function Φ assigns to each genealogy $x \in \Omega$ a real number which depends on the fecundity and mortality of the individuals that describe the genealogy. We shall call such functions Φ potentials.

Let M denote the set of all probability measures invariant under the shift T. For each $\mu \in M$, let $H_\mu(T)$ denote the Kolmogorov–Sinai entropy and $\Phi_\mu = \int \Phi \, d\mu$ the mean energy. The Free energy F_μ is given by

$$F_\mu = H_\mu(T) + \int \Phi \, d\mu \tag{3.1}$$

We shall call a state $\hat{\mu}$ an equilibrium state and we shall say the population is in *statistical equilibrium* if

$$F_{\hat{\mu}} = \sup_{\mu \in M} \left[H_\mu(T) + \int \Phi \, d\mu \right] \tag{3.2}$$

The precise expression for Φ will depend on whether or not the fecundity and mortality vary with population density. To describe the class of potentials we will consider, we recall the formalism used in the study of lattice models in statistical mechanics.

Consider a finite subset J of Z. Let X_J denote the set of mappings of the set J into the set S. This set X_J can be described as

$$X_J = \prod_{i \in J} S_i$$

The elements in the set X_J can be considered as restricted configurations of the set X.

Thus if $J = (i_1, i_2, \ldots, i_{|J|})$, where $|J|$ denotes the cardinality of the set J, an element of the set X_J has the form

$$(x_{i_1}, x_{i_2}, \ldots, x_{i_{|J|}}) \qquad \text{with} \quad x_{i_k} \in S.$$

We will use the symbol x to denote the elements both of the space X and X_J. The instances in which x denotes an element of the restricted subspace X_J will be clear from the context.

Now, the phase space of our population model is Ω, the set of all genealogies and not the set X. We take this into account by choosing for any $J \subset Z$ a subset Ω_J of the space X_J, the elements in the set Ω_J being the restricted configuration of the set Ω.

An interaction on the phase space Ω is defined as any real-valued continuous function φ on the space

$$\bigcup_{J \subset Z} \Omega_J$$

of all configuration spaces over the finite subsets J which satisfies the following conditions:

(i) $$\varphi(\Omega_{\emptyset}) = 0$$

where $\emptyset$ denotes the empty set in Z.

(ii) For all $k \in Z$, the quantity

$$\|\varphi\|_k = \sum_{J, J \ni k} \frac{1}{|J|} \sup_{x \in \Omega_J} |\varphi(x)|$$

is finite.

We introduce some terminology which will facilitate our discussion of interactions that arise in population models; see Refs. 14, 15. An interaction φ is said to be an *n-body* interaction if

$$\varphi(x) = 0 \qquad \text{for all} \quad x \in \Omega_J \quad \text{with} \quad |J| > n$$

An interaction is said to be *finite range* is there exists a real number $k \in N$

with

$$\varphi(x) = 0 \qquad \text{for} \quad \operatorname{diam} J = \max_{i,j \in J} |i - j| > k$$

Translational invariant interactions are functions which satisfy the relation

$$\varphi(x) = \varphi(Tx)$$

where T is the translation operator.

We will invoke the preceding classification to distinguish between density-independent and density-dependent systems in the context of our genealogical models. In this classification we study potentials $\Phi : \Omega^+ \to R$.

We note that in the case of the Leslie dynamical model given by (2.0), density independence describes the condition where fecundity and mortality are independent of population size. This implies that: (a) fecundity and mortality are time independent, (b) the age distribution at time (t) is uniquely determined by the age distribution at time ($t - 1$). In the context of our statistical representation, condition (a) implies that the corresponding potential on the one-sided sequence space Ω^+ assumes the form $\sum_{j=1}\varphi(x_0, x_j)$, where each two-body interaction $\varphi(x_0, x_j)$ extends by translation invariance over the whole space Ω^+. Condition (b) implies that the interaction is finite range, that is, $\varphi(x_0, x_j) = 0$ for $j > 1$. On account of this characterization we shall use the term "density independent" to describe potentials that have the form $\varphi(x_0, x_1)$.

In the density-dependent dynamical models, age-specific fecundity and mortality are no longer constants, but depend on the different age classes. The population size at each generation thus depends on the whole past history of the population. The potential, in this case, arises from interactions which depend on the whole genealogical history. The interaction will depend now on the sequences $(x_0), (x_0, x_1), (x_0, x_1, x_2) \ldots$ and so on. The corresponding potential has the form

$$\varphi_0(x_0) + \varphi_1(x_0, x_1) + \varphi_2(x_0, x_1, x_2) + \cdots$$

We will use the term "density dependent" to describe this class of many-body potentials.

The analysis of the existence and uniqueness of equilibrium states is based on properties of an operator defined on the space of continuous functions on the one-sided sequence space Ω^+. The shift operator $T : \Omega^+ \to \Omega^+$ is a finite-to-one continuous map of Ω^+ onto itself. For $\Phi : \Omega^+ \to R$, define

$$L_\Phi : f(x) \to \sum_{y \in T^{-1}(x)} \exp \Phi(y) f(y) \tag{3.3}$$

The map L_Φ defines a positive operator on $C(\Omega^+)$. Let

$$S_m\Phi(x) = \sum_{i=0}^{m-1} \Phi[T^i(x)] \tag{3.4}$$

Then

$$L_\Phi^m f(x) = \sum_{y \in T^{-m}(x)} \exp[S_m\Phi(y)]f(y) \tag{3.5}$$

Let $M(\Omega^+)$ denote the set of probability measures on Ω^+. The function Φ is said to satisfy the Perron–Frobenius condition if there are $\lambda > 0$, $h \in C(\Omega^+)$ with $h > 0$ and $\nu \in M(\Omega^+)$ for which

$$\begin{array}{ll} \text{(i)} & L_\Phi(h) = \lambda h \\ \text{(ii)} & L_\Phi^*(\nu) = \lambda\nu \quad \text{with} \quad \nu(h) = 1 \\ \text{(iii)} & \left\| \dfrac{L_\Phi^m(f)}{\lambda^m} - \nu(f)h \right\| \to 0 \end{array} \tag{3.6}$$

for all $f \in C(\Omega^+)$

The following result, Refs. 16 and 17, will be central to our studies.

Theorem 3.0. If the function Φ satisfies the Perron–Frobenius condition, then the measure $\mu(f) = \nu(hf)$ is translation invariant and is the unique equilibrium state for the interaction Φ.

We remark, however, that μ is not necessarily a Gibbs state. We recall that μ is a Gibbs state if there exist constants $c_1, c_2 > 0$ and a constant P such that

$$c_1 \leqslant \frac{\mu[x]}{\exp\left[-Pm + \sum_{k=0}^{m-1} \Phi(T^k(x))\right]} \leqslant c_2 \tag{3.7}$$

for every $x \in \Omega^+$ and every $m > 0$ where

$$[x] = \{z : z \in \Omega^+, z_i = x_i, 0 \leqslant i \leqslant m-1\} \tag{3.8}$$

This means that up to constants $[c_1, c_2]$, the relative probability of the $x_0 x_1 \ldots x_m$ are given by $\exp[\sum_0^m \Phi(T^k(x))]$.

Showing that μ is a Gibbs measure for Φ reduces to determining the constants c_1, c_2 verifying (3.7). This verification exploits the two relations (3.9) and (3.10) given below. These relations will be used in the characterization of Gibbs states given in Remarks 4.2, 5.1, and 5.2.

First we write

$$\mathrm{Var}_k(\Phi) = \sup|\Phi(x) - \Phi(y)| : x_i = y_i, \qquad 0 \leqslant i \leqslant k \tag{3.9}$$

It is easily shown that if

$$K = \sum_{k=0}^{\infty} \mathrm{Var}_k \Phi < \infty$$

for $y, z \in [x]$, then

$$|S_m\Phi(y) - S_M\Phi(z)| \leqslant K \tag{3.10}$$

Secondly, using (3.5), one has for $g \in C(\Omega^+)$

$$[(L_\Phi^m f) \cdot g](x) = \sum_{y \in T^{-m}(x)} \exp S_m\Phi(x) f(y) g(T^m(y)) \tag{3.11}$$

We shall exploit this circle of ideas and in particular Theorem 3.0 in our analysis of the equilibrium state of density-independent and density-dependent models.

4. DENSITY-INDEPENDENT INTERACTIONS AND GIBBS STATES

Consider the dynamical system $\bar{x}(t+1) = A\bar{x}(t)$ where $A = (a_{ij}) \geqslant 0$ denote the $n \times n$ population matrix given by

$$a_{ij} = \begin{cases} m_j & \text{for} \quad i = 1 \\ b_j & \text{for} \quad i = j + 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.0}$$

where $m_j \geqslant 0$, $m_{n-1} > 0$, $m_n > 0$, $0 < b_j < 1$.

The condition on the fecundity elements m_j are sufficient to ensure the primitivity of the matrix A.

The phase space Ω is given by

$$\Omega = \left\{ x \in \prod_{i=0}^{\infty} (1, 2, \ldots, n),\ a_{x_i, x_{i+1}} > 0 \text{ for all } i \right\}$$

The dynamical system described by the model (4.0) is density independent. Hence the potential for this model is described by a translation-invariant finite-range two-body potential. The interaction is given by

$$\varphi(x_0, x_1) = \log a_{x_0 x_1} \tag{4.1}$$

This interaction defined on the restricted subspace (x_0, x_1) extends by translation invariance to the one-sided sequence space Ω^+.

In order to analyze the equilibrium states of this model, we give another characterization of (4.1): Write

$$X_k = \{x \in \Omega^+, x_0 = k\}, \qquad 1 \leqslant k \leqslant n \tag{4.2}$$

The sets $\{X_k\}$ form a partition of Ω^+. Write

$$Y_{1k} = \{x \in X_1, x_1 = k\}, \qquad 1 \leqslant k \leqslant n$$

The interaction given by (4.1) can be defined by

$$\varphi(x) = \begin{cases} \log m_k & \text{if} \quad x \in Y_{1k}, \quad 1 \leqslant k \leqslant n \\ \log b_{k-1} & \text{if} \quad x \in X_k, \quad k > 1 \end{cases} \tag{4.3}$$

Let I_k denote the characteristic function of the set X_k, that is

$$I_k(x) = \begin{cases} 1, & x \in X_k \\ 0, & x \notin X_k \end{cases}$$

Write L for L_Φ where Φ is the translation-invariant potential derived from the interaction φ given by (4.1). The restriction of L to the finite-dimensional subspace S generated by the characteristic functions I_k has a matrix representation whose transpose is precisely the population matrix A. We will exploit this to prove Theorem 4.0. We refer to Rohlin(18) for notions of natural extension and Billingsley(19) for notions of Bernoulli shift.

Theorem 4.0. The system described by (4.1) has a unique equilibrium state which is a Gibbs state. The natural extension to a measure μ on the phase space Ω is translation invariant and (Ω, μ, T) is equivalent to a Bernoulli shift.

Proof. The transfer matrix of the density-independent interaction given by (4.1) is the population matrix A. Since A is irreducible and primitive, it follows that the operator L on the restricted subspace S satisfies the Perron–Frobenius conditions given by (3.6). Hence by Theorem 3.0, the measure $\mu = h\nu$ is the unique equilibrium state which is a Gibbs state (Remark 4.2). The natural extension of this measure, which is also denoted μ despite the possible ambiguity, is also translation invariant (Remark 4.1). The Bernoulli property follows from the fact that the natural extension of the partition $\{X_k\}$ with respect to Ω is a weak Bernoulli partition for T. Hence by the Friedman–Ornstein theorem, (Ω, μ, T) is equivalent to a Bernoulli shift. ■

Remark 4.1. To make μ into a measure on Ω, we use the fact that μ defined on Ω^+ is T-invariant, that is $\mu(f) = \mu(f \circ T)$ for $f \in C(\Omega^+)$. Now for $f \in C(\Omega)$, define $f^* \in C(\Omega^+)$ by

$$f^*\{x_i\}_{i=0}^{\infty} = \min\{f(y) : y \in \Omega, y_i = x_i \text{ for all } i \geqslant 0\}$$

An argument using the Cauchy criterion shows that

$$\tilde{\mu}(f) = \lim_{n\to\infty} \mu((f \circ T^n)^*) \tag{4.4}$$

exists.

Now $\tilde{\mu} \in C(\Omega)^*$, the dual of $C(\Omega)$. Hence by the Riesz representation theorem, $\tilde{\mu}$ defines a probability measure on Ω.

Furthermore, we observe that

$$\tilde{\mu}(f \circ T) = \lim_{n\to\infty} \mu\big((f \circ T^{n+1})^*\big) = \tilde{\mu}(f) \tag{4.5}$$

Hence $\tilde{\mu}$ is also T-invariant. Also $\tilde{\mu}(f) = \mu(f)$ for $f \in C(\Omega^+)$. We shall use the same symbol μ to denote the extension $\tilde{\mu}$ to Ω. The meaning will be clear from the context.

Remark 4.2. We shall show that the measure $\mu = h\nu$ satisfies the Gibbs condition with respect to the potential Φ derived from (4.1).

We denote the characteristic function of the set $[x]$ by I, where the set $[x]$ for a given $x \in \Omega^+$ is defined as in (3.8). We have therefore

$$\mu[x] = \nu(hI) \tag{4.6}$$

Let us write

$$W = \frac{\lambda^m \mu[x]}{\exp[S_m\Phi(x)]} \tag{4.7}$$

Now for the interaction (4.1), it can be shown, see (3.9), that

$$K = \operatorname{Var}_k(\Phi) = \sup_i \log\left(\frac{a_{ij}}{a_{ij'}}\right) \qquad \text{for all } j, j'$$

Now, let d denote the index of primitivity of the matrix A; that is, d is the minimum positive integer for which A^d is a strictly positive matrix.

We shall show that the Gibbs condition is satisifed for $c_1 = \lambda^{-d}\exp(-d\|\Phi\| - K)$, $c_2 = \|h\|\exp K$. The verification is in two steps.

(i) For any $z \in \Omega^+$, there is at most one $w \in T^{-m}(z)$ with $w \in [x]$, hence using (3.10) and (3.11) we have

$$\begin{aligned} L^m(hI)(z) &= \sum_{y \in T^{-m}z} \exp[S_m\Phi(y)]h(y)I(y) \\ &\leqslant \|h\|\exp[S_m\Phi(x)]\exp K \end{aligned} \tag{4.8}$$

From (4.6)

$$\lambda^m\mu[x] = \nu(L^m(hI)) \leqslant \|h\|\exp[S_m\Phi(x)]\exp K$$

Hence

$$W \leqslant \|h\|\exp K \tag{4.9}$$

(ii) Now given any $z \in \Omega^+$, there is at least one $w \in T^{-m-d}(z)$ with $w \in [x]$.

We conclude using (3.10) and (3.11) that

$$L^{m+d}(hI)(z) \geqslant \exp[S_{m+d}\Phi(w)]h(w) \tag{4.10}$$

Now from (4.6) and using (4.10), we have

$$\lambda^{m+d}\mu[x] = \nu\big(L^{m+d}(hI)\big) \geqslant \exp(-d\|\Phi\| - K)\exp\big[S_m\Phi(x)\big]$$

Hence

$$W \geqslant \lambda^{-d}\exp\big[-d\|\Phi\| - K\big] \tag{4.11}$$

From (4.9) and (4.11), we see that the condition (3.7) is verified for $P = \log\lambda$.

The Gibbs condition in effect means that up to constants in $[c_1, c_2]$, the relative probabilities of the genealogies $x_0x_1 \cdots x_{m-1}$ are given precisely by the number of individuals at time m with ancestry $x_0x_1 \cdots x_{m-1}$. This interpretation rests on the fact, shown in Remark (6.0), that

$$\exp\big[S_m\Phi(x)\big] = a_{x_0x_1}a_{x_1x_2} \cdots a_{x_{m-2}x_{m-1}}$$

The above represents at time m, the total number of individuals generated by a single individual at time (0) and possessing ancestry $x_0x_1 \cdots x_{m-1}$.

To compute the measure μ and its extension on Ω, we need to calculate $h \in C(\Omega^+)$ and $\nu \in M(\Omega^+)$. These elements can be explicitly expressed in terms of the eigenvectors $\bar{u} = (u_k)$ and $\bar{v} = (v_k)$ of the population matrix A. This correspondence is given in Theorem 4.1 below. The proof exploits the relation between the operator L and its matrix representation; this representation, we recall, is the transpose of the population matrix A.

Theorem 4.1. The elements $h \in C(\Omega^+)$ and $\nu \in M(\Omega^+)$ are given by

$$\text{(i)}\quad h(x) = av_k,\quad x \in X_k \quad\text{and}\quad \text{(ii)}\quad \nu(X_k) = (1/a)u_k$$

with $a > 0$ and where X_k is the partition given by (4.2).

Proof. Since the operator L satisfies the Perron–Frobenius condition, (3.6), we have,

$$\frac{L^n(f)}{\lambda^n} \to \nu(f)h \qquad \text{for all}\quad f \in C(\Omega^+)$$

Putting $f \equiv 1$, we get

$$\frac{L^n 1(x)}{\lambda^n} \to h(x) \qquad \text{for}\quad x \in \Omega^+$$

Hence $h(x)$ only depends on x_0.

Define $w_k = h(k, x_1, x_2, \ldots)$ for any choice of $x_1, x_2, x_2 \ldots$. Then since $Lh = \lambda h$, we have

$$\sum_k w_k a_{kj} = \lambda w_j$$

This means the $\bar{w} = (w_k)$ is a left eigenvector of A corresponding to the dominant eigenvalue λ. However, A is primitive, hence λ has geometric multiplicity 1. Since $\bar{v}A = \lambda\bar{v}$, it follows that $\bar{w} = a\bar{v}$ for some scalar a. We conclude that

$$h(x) = av_k \qquad \text{for} \quad x \in X_k \tag{4.12}$$

To prove (ii), we note, using the relation between L and its representation A transpose,

$$\frac{1}{\lambda^n}\left(a_{kj}^{(n)}\right) = \frac{1}{\lambda^n} L^n f_k(j, x_1, x_2, \ldots)$$

where

$$f_k(x_0, x_1, \ldots) = \begin{cases} 1 & \text{if} \quad x_0 = k \\ 0 & \text{if} \quad x_0 \neq k \end{cases}$$

However, using the primitivity of the matrix A, we have for each k,

$$\frac{1}{\lambda^n}\left(a_{kj}^{(n)}\right) \to u_k v_j \tag{4.13}$$

Since L satisfies condition (3.6), we have for each k

$$\frac{1}{\lambda^n} L^n\left(f_k(j, x_1, x_2 \ldots)\right) \to h(j, x_1, x_2 \ldots)\nu(f_k) \tag{4.14}$$

From (4.13) and (4.14) we conclude that $h(j, x_1, x_2, \ldots)\nu(f_k) = u_k v_j$. Using (4.12), we have $\nu(f_k) = (1/a)u_k$, which is (ii). ■

Remark 4.3. In (2.3) and (2.4) the eigenvectors of A are given explicitly in terms of the fecundity, mortality elements. Using these values and Theorem 4.1, we have the measure $\mu = h\nu$ on Ω^+,

$$h(x) = a\left(\frac{\lambda^k}{l_k}\right)\frac{\sum_{j=k}^{n} p_j}{\sum_{j=1}^{n} j p_j} \qquad \text{for} \quad x \in X_k$$

where

$$p_j = \frac{l_j m_j}{\lambda^j} \quad \text{and} \quad \nu(X_k) = \left(\frac{1}{a}\right)\frac{l_k}{\lambda^k}$$

The extension of this measure to the space of genealogies Ω is a probability measure μ given by

$$\mu(x_0, x_1, x_2, \ldots x_l) = q_{x_0} P_{x_0 x_1} P_{x_1 x_2} \cdots P_{x_{l-1} x_l}$$

where

$$q_j = u_j v_j = \frac{\sum_{k=j}^{n} p_k}{\sum_{k=1}^{n} k p_k} \tag{4.15}$$

and

$$P_{ij} = a_{ij}u_j/\lambda u_i$$

Using (2.3) and (4.0), the probability matrix P_{ij} is given by

$$P_{ij} = \begin{cases} p_j, & i = 1 \\ 1, & i > 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.16}$$

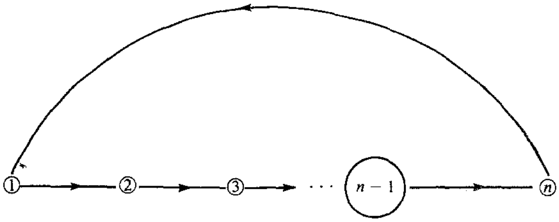
Remark 4.4. We can derive explicit expressions for the entropy $H_\mu(T)$ and the $\int \Phi \, d\mu$. Since (Ω, μ, T) is a Markov shift, the entropy which we denote H is given by

$$H = -\sum_i \sum_j q_i P_{ij} \log P_{ij}$$

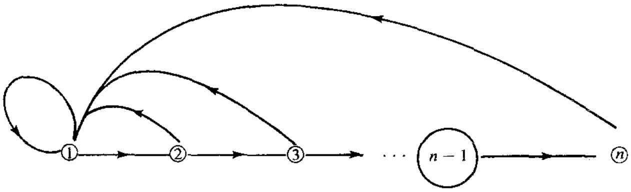
Using (4.15) and (4.16) this reduces to

$$H = -\frac{\sum p_j \log p_j}{\sum j p_j} \tag{4.17}$$

Now the expression p_j, which can be written $p_j = u_j m_j$, where (u_j) is the age-distribution at equilibrium, represents the expected future contribution of individuals of age class (j) to the ancestry of future generation. Thus H measures the spread or variability of this contribution as a function of age. Semelparous populations, that is populations that reproduce only once in their lifetime, have zero entropy. The corresponding graph is given by



Interparous populations, that is populations that breed at various intervals over their life-cycle, can be represented by the following graph:



These populations have positive entropy.

The mean energy $\int \Phi \, d\mu$, which we denote $\overline{\Phi}$, is given by

$$\overline{\Phi} = \frac{\sum p_j \log l_j m_j}{\sum j p_j} \tag{4.18}$$

The above expression, which is called the reproductive potential, describes the means of the contribution of the different age classes to the equilibrium age distribution.

The expression in the denominator of H and $\overline{\Phi}$ is the generation time T.

We have, from (4.17) and (4.18),

$$\log \lambda = H + \overline{\Phi} \tag{4.19}$$

Writing $H^* = -\sum p_j \log p_j$, the expression (4.19) becomes

$$\log \lambda = H^*/T + \overline{\Phi} \tag{4.20}$$

Remark 4.5. The population parameters we have described have precise analogs in thermodynamic theory. To illustrate these analogies we consider the thermodynamic formalism in the context of finite systems.

Let X denote a nonempty finite set. Given a probability measure μ on S, we define its entropy

$$S(\mu) = -\sum_{x \in X} \mu(x) \log \mu(x)$$

Let U denote a real-valued function on X and write

$$\mu(U) = \sum_{x \in X} \mu(x) U(x)$$

and

$$Z = \sum_{x \in X} \exp[-U(x)]$$

It can easily be shown that the maximum of the expression

$$S(\mu) - \mu(U)$$

over all probability measures μ on X is $\log Z$ and is attained precisely for $\mu = \sigma$ where

$$\sigma = \frac{\exp[-U(x)]}{Z}$$

In physical applications, we can interpret X as the space of configurations. Let us write $U = \beta E$, where E denotes the energy of the configuration x and $\beta = 1/k\overline{T}$, where $\overline{T}$ is the absolute temperature and k is a Boltzmann's constant. Then we have the statement that the measure σ

minimizes the expression $E - kS\overline{T}$ or maximizes the expression

$$kS\overline{T} - E \tag{4.21}$$

The thermodynamic analogs of our population parameters emerge when we compare the right-hand side of (4.20) with the classical thermodynamic expression (4.21).

We observe that the temperature corresponds to the reciprocal of the generation time, the mean energy corresponds to the reproductive potential, and the free energy is analogous to the growth rate.

5. DENSITY-DEPENDENT MODELS AND PHASE TRANSITIONS

Phase transition in one-dimensional models of statistical mechanics can arise either in the case of pair interaction potentials that violate a strong short-range condition, Dyson,(20) or in the case of many-body interactions of indefinitely high order, Fisher and Felderhof.(21) Now, density-independent models are characterized by a class of translation-invariant finite-range potentials. These potentials give rise to unique equilibrium states. Density-dependent models are characterized by many-body potentials of indefinitely high order. These potentials will in general yield multiple equilibrium states. We will now consider a class of potentials that describe the density-dependent models and study their equilibrium properties. This analysis is influenced by the model of Fisher and Felderhof,(21) who studied a class of many-body interactions, and the work of Hofbauer(22) and Ledrappier,(23) who studied the corresponding situation in terms of models on two-shift spaces. Our motivation and methods are quite different. In our detailed choice of interactions we are guided by our insight into the behavior of density-dependent interactions in demographic models. The associated operator for the potential function we consider is an infinite-dimensional Leslie matrix. The biological interpretation of this matrix suggests the conditions we must impose on the interaction in order to have multiple equilibrium states. In studying the spectral properties of the Perron–Frobenius operator we apply results concerning the classification of infinite matrices into positive recurrent, null recurrent, and transient types, Ref. 24.

Consider a population divided into age classes denoted (0) and (1). The dynamics is given by the matrix $A(t)$, where

$$A(t) = \begin{bmatrix} m_1(t) & m_2(t) \\ b_1(t) & 0 \end{bmatrix}$$

We assume that the fecundity is density dependent, while the mortality $b_1(t)$ is constant. The condition on the fecundity implies that the potential

Φ on Ω is a many-body potential. The one-sided sequence space Ω^+ is given by

$$\Omega^+ = \left\{ x \in \prod_0^\infty (0,1),\ a_{x_i, x_{i+1}} > 0\,,\ i \geqslant 0 \right\}$$

The sets

$$Y_0 = \{x : x \in \Omega^+, x_0 = 0\}$$
$$X_0 = \{x : x \in \Omega^+, x_0 = 1\}$$

are a partition of Ω^+.

The constant mortality is described by a potential

$$\Phi(x) = \log b_0\,, \qquad x \in X_0$$

To describe the time-dependent fecundity we partition Y_0 into sets of the form

$$X_k = \{x : x \in \Omega^+, x_n = 0, \text{ for } 0 \leqslant n \leqslant k-1, x_k = 1\}, \qquad k = 1, 2, 3, \ldots \tag{5.0}$$

We describe the potential due to the fecundity elements by

$$\Phi(x) = \log b_k\,, \qquad x \in X_k \quad k \geqslant 1$$

Hence the potential for the population model is given by

$$\Phi(x) = \log b_k\,, \qquad x \in X_k\,, \quad k \geqslant 0 \tag{5.1}$$

This potential can be expressed in terms of sum of interactions:

$$\varphi_1(x_0) + \varphi_2(x_0, x_1) + \varphi_3(x_0, x_1, x_2) + \cdots$$

where

$$\varphi_1(x_0) = \begin{cases} \log b_0\,, & x_0 = 1 \\ 0, & x_0 \neq 1 \end{cases}$$

$$\varphi_2(x_0, x_1) = \begin{cases} \log b_1\,, & x_0 x_1 = 01 \\ 0, & x_0 x_1 \neq 01 \end{cases}$$

$$\vdots$$

$$\varphi_k(x_0, x_1, \ldots, x_k) = \begin{cases} \log b_k\,, & x_0 x_1, \ldots, x_k = 000\ldots01 \\ 0, & x_0 x_1, \ldots, x_k \neq 000\ldots01 \end{cases}$$

Write L for the operator L_Φ induced by the potential (5.1) and let I_k denote the characteristic function of the set X_k. The transpose B of the matrix representation of the operator L with respect to the basis $\{I_k\}$ is given by

$$\begin{bmatrix} 0 & b_0 & b_0 & b_0 & \cdots \\ b_1 & 0 & 0 & 0 & \cdots \\ 0 & b_2 & 0 & 0 & \cdots \\ 0 & 0 & b_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

This matrix which is of Leslie type can be considered as describing a population with an infinite number of age classes, each age class having a fecundity b_0 independent of age. The survivorship b_i is age dependent.

This description will be useful in interpreting biologically the analytic conditions we will impose on the elements b_k. Write

$$l_k = \prod_1^k b_r$$

This describes the expected survivorship to age $(k + 1)$. Write

$$w_k = b_0 l_k$$

The dominant eigenvalue of the matrix B is the unique positive real root of the equation

$$\sum_{k=1}^{\infty} \frac{w_k}{\lambda^{k+1}} = 1 \tag{5.2}$$

The number $r = \log\lambda$ corresponds to the Malthusian parameter or the intrinsic rate of increase of the population. We observe from (5.2) that when $\sum w_k > 1$, the growth rate is positive, whereas $\sum w_k = 1$ corresponds to zero growth rate. These conditions are the key to distinguishing between unique and multiple equilibrium states for the potential Φ given by (5.1) and analyzed in Theorems 5.0 and 5.1. The proofs of these two theorems are based on the relationship between the spectral radius of the positive operator L and the convergence parameter associated with the matrix representation B; see Ref. 24.

We first analyze the case $\sum w_k > 1$.

We have the following theorem:

Theorem 5.0. Suppose $\sum w_k > 1$, then the potential Φ given by (5.1) has a unique equilibrium state.

Proof. The matrix B is clearly irreducible with period $d = 1$. The elements $\{b_{ij}^{(n)}\}$, the (i, j) element of the matrix B^n, all exist and are finite. Hence by Ref. 24, $\lim[b_{ij}^{(n)}] = 1/R$ exists. $1/R$ is called the convergence parameter and R the convergence norm.

Now a simple computation using the finite truncations of B shows that the convergence norm $R > 1$ and hence that B is positive recurrent. Hence by Ref. 24, there exist unique left and right eigenvectors, say (u_k), (v_k) for the eigenvalue $1/R$, with the property that

(i) $$\sum_{k=1}^{\infty} u_k v_k < \infty$$

and

(ii) $$\lim_{n\to\infty} b_{ij}^{(n)} R^n = \frac{u_j v_i}{\sum u_k v_k}$$

A simple computation shows that the spectral radius of the operator L_Φ and the convergence parameter of the matrix B coincide, therefore by Ref. 24 L_Φ satisfies the Perron–Frobenius condition. Hence by Theorem 3.10 Φ has a unique equilibrium state. ■

Remark 5.0. The equilibrium state μ is given by $\mu = h\nu$, where

$$L_\Phi h = \nu h$$

$$L_\Phi^* h = \lambda \nu$$

with $h \in C(\Omega^+)$ and $\nu \in M(\Omega^+)$.

To compute h and ν, we determine the left and right eigenvectors of B. We have by considering finite truncations of B and taking the limit

$$u_k = \frac{w_k}{\lambda^{k+1}}$$

$$v_k = \frac{\sum_{j=k}^{\infty} u_j}{T u_k} \qquad \text{for} \quad k = 0, 1, 2, \ldots$$

where

$$T = \sum_{j=1}^{\infty} j u_{j-1}$$

Remark 5.1. Theorem 4.1 given in Section 4 can be extended to treat the case where the matrix representation is infinite dimensional. We omit the details; the argument yields ν and h in terms of u_k and v_k, respectively. We have

$$\nu(X_k) = u_k$$

and

$$h(x) = v_k \qquad \text{for} \quad x \in X_k$$

where $\{X_k\}$ denotes the partition defined by (5.0). As in Remark 4.1 the measure $\mu = h\nu$ on Ω^+ extends in a natural way to a measure denoted also μ on the configuration space Ω.

Remark 5.2. Convergence of $\log w_k$ implies that $\mu = h\nu$ is a Gibbs measure for Φ. This can be shown using the same argument invoked in Remark 4.2. In effect the Gibbs condition is satisfied for $c_2 = \|h\| \exp K$ and $c_1 = \lambda^{-d}\exp(-d\|\Phi\| - K)$, where $K = \lim w_k$ and d is the index of primitivity of the 2×2 population matrix.

The convergence of $\log w_k$ means that all individuals in the population have finite life expectancy. Hence at each instant m the population size is finite and the relative probabilities of the genealogies $x_0 x_1 \cdots x_m$ are given by the number of individuals with this ancestry.

Remark 5.3. Divergence of the sequence $\log w_k$ implies that μ is not a homogeneous measure. To see this, we note that if (3.7) holds, then for y, $z \in [x]$, we have

$$|S_m(y) - S_m(z)| \leqslant K, \qquad \text{where} \quad k = \log(c_2/c_1)$$

We now let $x_i = 0$, $0 \leqslant i \leqslant m-1$, and $y_i = 1$, $i \geqslant m$, $z_i = 0$, $i \geqslant m$. Then clearly $|\log w_k| \leqslant K$ and $\log w_k$ converges.

The divergence of $\log w_k$ means that some individuals in the population have infinite life-expectancy. Hence at some instants, with m finite, the population will have infinite population size. This means that the relative probabilities of the genealogies $x_0 x_1 \cdots x_m$ cannot be described by the number of individuals with the prescribed ancestry. The homogeneous condition thus fails.

We now consider the case $\sum w_k = 1$. In analyzing this case we consider the induced system (Ω^*, T^*) defined as follows. Let

$$X_0 = \{x \in \Omega^+, x_0 = 1\}$$

$$\Omega^* = X_0 \setminus \bigcup_{n=0}^{\infty} T^{-n}\{0\}$$

where $\{0\} = (0, 0, 0, \ldots)$. Consider the set Y_k defined by

$$Y_k = \{x \in X_0, x_0 = 1, x_i = 0, 1 \leqslant i \leqslant k, x_{k+1} = 1\}$$

Let

$$Y_k^* = Y_k \cap \Omega^*$$

The elements $\{Y_k^*\}$ are a partition of Ω^*. The transformation $T: \Omega^+ \to \Omega^+$ induces a transformation

$$T^* : \Omega^* \to \Omega^*$$

defined by $T^*(x) = T^{k+1}(x)$ if $x \in Y_k^*$.

In studying the relation between the two systems (Ω^+, T) and (Ω^*, T^*) the properties we invoke are the following.

1. Every T-invariant measure μ on Ω^+ with $\mu(0) = 0$ determines a T^*-invariant measure μ^* on Ω^* by

$$\mu^*(A) = \mu(A \cap X_0)/\mu(X_0) \tag{5.3}$$

Moreover by Abramov's theorem

$$H_\mu(T) = \mu(\Omega^*)H_{\mu^*}(T^*) \tag{5.4}$$

2. Every T^*-invariant measure σ^* on Ω^* is of the form μ^* for a unique T-invariant μ with $\mu(0) = 0$ provided

$$(k+1)\,\mu^*(Y_k^*) < \infty$$

Now consider the potential Φ^* defined on Ω^* by

$$\Phi^*(x) = \log w_k \qquad \text{if} \quad x \in Y_k^* \tag{5.5}$$

The adjoint of the operator $L_\Phi^*: C(\Omega^*) \to C(\Omega^*)$ has the matrix representation $B^* = (b_{ij}^*)$, where $b_{ij}^* = w_i$. Let R denote the convergence parameter of the matrix B. Since $\sum w_k = 1$, then $R = 1$ and B^* is recurrent. Moreover, the spectral radius of L_Φ^* and the convergence parameter of B^* coincide. We can now prove the following.

Theorem 5.1. Assume Φ is given by (5.1) and suppose $\sum w_k = 1$.

(1) If $\sum(k+1)w_k < \infty$, then Φ has two equilibrium states, both ergodic.

(2) If $\sum(k+1)w_k = \infty$, then Φ has a unique equilibrium state.

Proof. We consider the induced system (Ω^*, Φ^*, T^*) with operator L_Φ^* and matrix B^*.

(i) Since $\sum(k+1)w_k < \infty$, then B^* is positive recurrent. Since the spectral radius of L_Φ^* and the convergence parameter of B^* coincide, and since B^* is positive recurrent, then L_Φ^* satisfies the Perron–Frobenius condition. The product measure μ^* on Ω^* with weights

$$w_k \Big/ \sum_{k=1}^{\infty} w_k$$

is the unique equilibrium state for Φ^*. Clearly μ^* is ergodic, and a direct computation shows

$$H_{\mu^*}(T^*) + \int \Phi^*\, d\mu^* = 0$$

The measure μ^* determines a unique ergodic T-invariant measure μ on Ω^+. A direct computation shows

$$\int \Phi\, d\mu = \mu(\Omega^*) \int \Phi^*\, d\mu^*$$

Hence, using (5.4), we have $H_\mu(T) + \int \Phi\, d\mu = 0$.

Now let ω be a T-invariant measure on Ω with $\omega(0) = 0$, $\omega \neq \mu$, and let ω^* denote the corresponding measure on T^*. The partition $\{Y_k^*\}$ is a one-sided generator with finite entropy for ω^*. Since μ^* is the unique equilibrium state for Φ^*, we conclude that

$$H_{\omega^*}(T^*) + \int \Phi^* \, d\omega^* < 0$$

Hence, we have

$$H_{\omega}(T) + \int \Phi \, d\omega < 0$$

and μ is an equilibrium state for Φ.

We note however, that the point measure δ_0 concentrated at (0) also satisfies

$$H_{\delta_0}(T) + \int \Phi \, d\delta_0 = 0 \tag{5.6}$$

Hence δ_0 is also an equilibrium state for Φ. The equilibrium states for Φ are convex combinations of the ergodic measures μ and δ_0.

To complete the proof of the theorem, we note that if $\sum(k+1)w_k = \infty$, then B^* is null recurrent and hence Φ^* has no equilibrium states. This implies that the Dirac measure δ_0, an equilibrium state for Φ, is the unique equilibrium state. ■

Remark 5.3. The condition $\sum w_k = 1$ as we have observed corresponds to a population with zero growth rate.

The restriction $\sum(k+1)w_k < 0$ describes a population with finite mean generation time. Under these conditions, we have two ergodic equilibrium states. The Dirac measure δ_0 concentrated at age group (0), corresponds to a semelparous population, that is, a population in which reproduction is concentrated at a single instant in the individual's life-time. The other equilibrium state corresponds to an iterparous population, that is, one in which both age groups are represented at equilibrium and reproduction occurs in both age classes.

The restriction $\sum(k+1)w_k = \infty$ corresponds to a population with infinite generation time. In this case, one has a unique equilibrium state, the Dirac measure. This example describes a population with zero growth rate, infinite generation time, and zero entropy.

The conclusions drawn from these two cases are compatible with the analogy we have discovered between generation time and the reciprocal of the temperature. For example, perfect crystals have a very orderly structure, and at very low temperature, the lattice vibrations will all be in their lowest state, which corresponds to zero entropy. Thus a crystal will have a very low entropy at temperatures approaching the absolute zero. Organisms in the dormant phase have no metabolic activity. These organisms can be considered as having infinite generation time and zero entropy. Organisms in the dormant phase are the biological analogs of perfect crystals.

6. POPULATION SIZE AND GROWTH RATES

6.1. Census Size and the Malthusian Parameter

The census size N_t describes the total number of individuals in the population at time t. In models in which this number has been characterized, one considers the density-independent case where the dynamics is given by the time-independent Leslie model. The quantity N_t is the norm of the age-distribution vector $\bar{u}(t)$. At equilibrium, each age class increases at the rate λ which is the dominant eigenvalue of the Leslie matrix A. Thus N_t increases asymptotically at the rate $r = \log\lambda$, the Malthusian parameter. The aim of this section is to apply the thermodynamic formalism to generalize this result to nonlinear models. We will therefore consider a population (Ω, Φ) where the only hypothesis we place on the potential Φ is that it gives rise to a unique equilibrium state μ. We shall assume that the measure μ is mixing.

Consider an individual aged k. This individual generates a set of genealogies which we describe as follows.

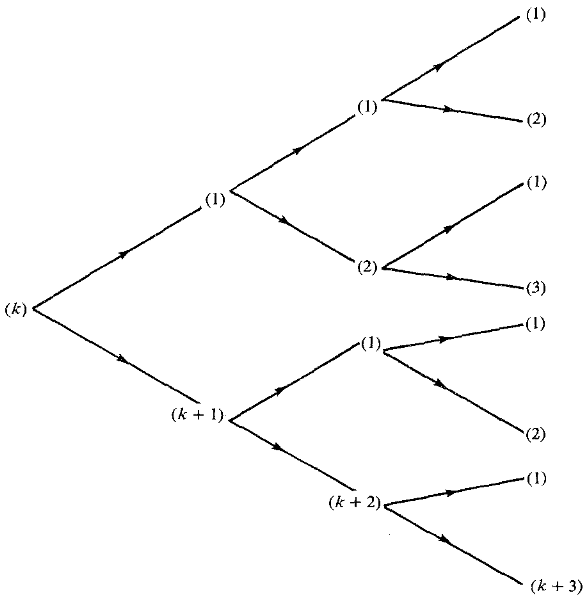


Fig. 5.0.

Consider a particular genealogy

$$kx_1x_2 \cdots x_{t-1}$$

of length t and let

$$[x] = \{y \in \Omega, y_0 = k, y_i = x_i\,;\ 1 \leqslant i \leqslant t\}$$

For $z \in [x]$, write

$$S_t[\Phi(z)] = \sum_{i=1}^{t-1} \Phi[T^i(z)] \tag{6.0}$$

$$N_{k,t} = \sum_{k, x_1, x_2 \ldots x_{t-1}} \sup_{z \in [x]} \exp S_t[\Phi(z)] \tag{6.1}$$

The quantity $N_{k,t}$ describes the total number of individuals in the population at time t, generated by an individual aged (k) at time $t = 0$. See Remark 6.0.

Write

$$N_t = \sum_k N_{k,t} = \sum_{x_0 x_1 \ldots x_{t-1}} \sup_{z \in [x]} \exp S_t[\Phi(z)] \tag{6.2}$$

This number represents the total number of individuals in the population at time (t), assuming that the initial population consists of a single individual in each age class.

Now, let L_Φ denote the positive operator associated with the potential Φ and let λ denote the spectral radius of L_Φ. We shall show the following.

Theorem 6.0. $\lim_{t\to\infty}(1/t)\log N_t = r(\Phi)$, where $r(\Phi) = \log\lambda$.

Remark 6.0. It is instructive to consider the relationship between this result and the result for density-independent models. In the case of density-independent models, the adjoint of the operator L_Φ is the matrix A, hence the $r(\Phi)$ in Theorem 6.0 coincides with the Malthusian parameter. We will now interpret the expression $S_t[\Phi(z)]$ given by (6.0), when Φ is described by the density-independent interaction $\Phi = \log a_{x_0x_1}$. We will show that (6.2) reduces, in this case, to the norm of the age-distribution vector $\bar{u}(t)$.

Consider as before the genealogy

$$kx_1x_2 \cdots x_{t-1}$$

The expression $S_t[\Phi(z)]$ becomes

$$S_t[\Phi(z)] = \log a_{kx_1} a_{x_1x_2} \cdots a_{t-2\,t-1}$$

and

$$\exp S_t[\Phi(z)] = a_{kx_1} a_{x_1x_2} \cdots a_{t-2\,t-1}$$

This expression given at time t represents the total number of individuals with ancestry

$$kx_1x_2 \cdots x_{t-1}$$

Assuming that the initial population consists of an individual from each age class, then the total number of individuals at time t is

$$\sum_{x_0x_1 \cdots x_{t-1}} a_{x_0x_1}a_{x_1x_2} \cdots a_{t-2\,t-1}$$

which we denote N_t. The expression for N_t can also be written as $N_t = \sum u_i(t)$, where $\bar{u}(t) = u_i(t)$ denote the age-distribution at time t. We know that for density-independent models, N_t increases asymptotically at the rate r, the Malthusian parameter. Thus Theorem 6.0 is a natural generalization of the density-independent case.

Proof of Theorem 6.0. The definition we have given, see (6.0), (6.1), (6.2), for census size is based on ideas invoked in defining the pressure function.$^{(17)}$ Our proof imitates the arguments given in Ref. 17 to characterize this concept. We sketch the main argument, which depends on the following properties:

(i) $$L_\Phi^t f(x) = \sum_{y \in T^{-t}(x)} \exp[S_t\Phi(x)]f(y)$$

(ii) $$\lim_{t\to\infty} \frac{1}{t} \log \|L_\Phi^t\| = \log \lambda$$

(iii) $$\lim_{t\to\infty} \frac{1}{t} \log N_t = \inf_t \log N_t$$

(i) follows from the definition of L_Φ; see (3.5); (ii) is based on the fact that λ is the spectral radius of L_Φ. To derive (iii), we use the result that if $\{a_n\}$ is a sequence of positive real numbers with $a_{n+m} \leqslant a_n + a_m$ for all n, m then $\lim(1/n)a_n$ exists and equals $\inf(1/n)a_n$. Thus to obtain (iii), we observe that

$$\sup_{z\in[x]} S_{t+s}[\Phi(z)] \leqslant \sup_{z\in[x]} S_t[\Phi(z)] + \sup_{z\in[x]} S_s[\Phi(z)]$$

From this one gets

$$N_{t+s} \leqslant N_tN_s$$

Now, by writing $a_t = \log N_t$, (iii) follows.

Remark 6.1. In the case of the density-independent model defined by the interaction (4.1), we have the relation

$$r = H + \overline{\Phi} \tag{6.3}$$

where H and $\overline{\Phi}$ are given by (4.17) and (4.18). The above relation is obtained using the explicit expressions for $H_\mu(T)$ and $\int \Phi \, d\mu$ at equilibrium.

For arbitrary interactions Φ, we have no explicit expression for the entropy and the mean energy at equilibrium. However, if $\hat{\mu}$ denote the equilibrium state, an analog of the relation for density-independent models holds, namely,

$$r(\Phi) = H_{\hat{\mu}}(T) + \int \Phi \, d\hat{\mu} \tag{6.4}$$

To prove this relation, it is sufficient to show (Ref. 25) that

$$r(\Phi) = \sup_{\mu} \left[H_{\mu}(T) + \int \Phi \, d\mu \right] \tag{6.5}$$

The function $r(\Phi)$ is analogous to the pressure function in statistical mechanics and the proof of (6.4) uses ideas given in Ref. 17.

6.2. Effective Size and Entropy

The census size N_t has been characterized for populations described by arbitrary interactions Φ. The asymptotic behavior of N_t is described by the dominant eigenvalue of the operator L_{Φ}. We now introduce the notion of effective size and show that its asymptotic behavior is given by the population entropy H.

Consider a population on a phase space Ω with equilibrium measure μ. We assume μ is ergodic.

Let Ω^t denote the set of genealogies generated by a single individual during the interval $(0, t)$. Let E denote a subset of $\Omega^{(t)}$. We associate with the set E a measure as follows:

Consider an element $z \in E$

$$z = (z_0, z_1, \ldots, z_{t-1})$$

and write

$$[z] = \{x \in \Omega, x_i = z_i, 0 \leqslant i \leqslant t-1\}$$

We now write

$$\mu_t(E) = \sum_{z \in E} \mu[z]$$

We say that $E \subset \Omega^{(t)}$ is ϵ-observable if

$$\mu_t(E) > 1 - \epsilon$$

The effective size N_t^* is defined by

$$N_t^*(\epsilon) = \min\{\operatorname{card} E : E \subset \Omega^{(t)}, \mu_t(E) > 1 - \epsilon\} \tag{6.6}$$

Thus the effective size describes the minimal number of genealogies whose total probability exceeds $1 - \epsilon$.

We now have the following theorem.

Theorem 6.1. $\lim[(1/t)\log N_t^*(\epsilon)] = H$.

Proof. Consider $z \in \Omega^t$. Since μ is ergodic, we have by the Shannon–McMillan theorem that $-(1/t)\log \mu[z]$ converges in probability to H.

Hence, given any $\delta > 0$,

$$\lim_{t\to\infty} \mu\left\{ z : z \in \Omega^{(t)} : \left| -\frac{1}{t}\log \mu[z] - H \right| \geqslant \delta \right\} = 0 \tag{6.7}$$

This implies that for $\epsilon_1 > 0$, $\epsilon + \epsilon_1 > 1$, $\epsilon_1 \geqslant \epsilon$, there exists t_0 such that for any $t > t_0$

$$\text{(i)} \qquad \mu\left\{ z : z \in \Omega^{(t)}, \left| -\frac{1}{t}\log \mu[z] - H_\mu(T) \right| \geqslant \delta \right\} < \epsilon_1$$

$$\text{(ii)} \qquad 1 - \epsilon - \epsilon_1 < e^{-\delta t}$$

Write

$$E_t = \left\{ z : z \in \Omega^{(t)}, e^{-t(H+\delta)} < \mu[z] < e^{-t(H-\delta)} \right\} \qquad \text{for} \quad t = 1, 2, 3 \ldots \tag{6.8}$$

According to the above relation (i), we have $\mu(E_t) \geqslant 1 - \epsilon$. Now, by the definition of $N_t^*(\epsilon)$, we have that the number of elements in E_t is greater than $N_t^*(\epsilon)$. Since $z \in E_t$ implies that

$$\mu[z] > e^{-t(H+\delta)} \tag{6.9}$$

We conclude that

$$1 \geqslant \sum_{z \in E_t} \mu[z] > N_t^*(\epsilon) e^{-t(H+\delta)}$$

and hence

$$\frac{1}{t}\log N_t^*(\epsilon) < H + \delta \tag{6.10}$$

Now let us order the elements $\{z^{(i)}\} \in \Omega^t$ such that $\mu[z^{(1)}] \geqslant \mu[z^{(2)}] \geqslant \cdots \geqslant \mu[z^{(m)}]$, where m denotes the number of elements in $\Omega^{(t)}$.

To simplify notation, write $N_t^*(\epsilon) = n^*$. Then, clearly

$$\sum_{k=1}^{n^*} \mu[z^{(k)}] > 1 - \epsilon$$

Put $X_t = \{k : 1 \leqslant k \leqslant n^*,\ z^{(k)} \in E_t\}$. Since $k \notin X_t$ together with $1 \leqslant k$

$\leqslant n^*$ implies that $\mu[z^{(k)}] \in \Omega^t - E_t$, we have

$$\sum_{k=1}^{n} \mu[z^{(k)}] - \sum_{k \in X_t} \mu[z^{(k)}] \leqslant \mu(\Omega^{(t)} - E_t)$$

However $\mu(\Omega^{(t)} - E_t) < \epsilon_1$. Hence

$$\sum_{k \in X_t} \mu[z^{(k)}] > \sum_{k=1}^{n^*} \mu[z^{(k)}] - \epsilon_1$$

Now $k \in X_t$ implies that $\mu[z^{(k)}] < e^{-t(H-\delta)}$. Hence

$$\sum_{k \in X_t} \mu[z^{(k)}] < n^* e^{-t(H-\delta)}$$

Thus

$$1 - \epsilon - \epsilon_1 < n^* e^{-t(H-\delta)}$$

which now yields

$$H - \delta - \frac{1}{t} \log \frac{1}{1 - \epsilon - \epsilon_1} < \frac{1}{t} \log n^*$$

which gives

$$H - 2\delta < \frac{1}{t} \log n^* \tag{6.11}$$

From (6.10) and (6.11), we have $(1/t)\log N_t^*(\epsilon) \rightarrow H$ and the theorem is proved. ■

Remark 6.2. A heuristic interpretation of the effective size concept can be derived as follows. We note that a single individual generates a certain number of genealogies. By the Shannon–McMillan theorem, we can assert that for arbitrary small $\epsilon > 0$ and $\delta > 0$, and for sufficiently large t, all genealogies can be separated into two classes S_1 and S_2 such that

(a) for every genealogy x in the class S_1

$$\left| \frac{\log \mu(x)}{t} + H \right| < \epsilon$$

(b) the sum of the probabilities of genealogies belonging to the class S_2 is less than δ.

For all genealogies $x \in S_1$, $-(1/t)\log \mu(x)$ is close to H. Therefore all genealogies of the class S_1 have approximately the same probability, namely, e^{-tH}. This means that the number of genealogies in this class is approximately e^{tH}. This number characterizes our effective size N_t^*. It describes the number of typical genealogies and corresponds roughly speaking to the total number of reproductives in the population.

By contrast, the census size N_t describes the mean number of genealogies and corresponds to the total number of individuals in the population.

7. ENVIRONMENTAL INTERACTIONS

7.1. Adaptive Value

The main thrust of the analysis we have described so far is based on the contrast between the dynamical model $\bar{x}(t+1) = A(t)\bar{x}(t)$ and its statistical representation. The macroscopic parameters derived from the statistical representation completely characterize the dynamics of populations in fixed or constant environments. We will now extend this analysis to the case where random environmental forces affect the population dynamics. In this context, the dynamical description which generalizes the Leslie model is given by $\bar{x}(t+1) = A(t)\bar{x}(t) + \bar{\eta}(t)$, where $\bar{\eta}(t)$ denotes a random vector which represents the influence of the environment on the age distribution. In the statistical representation of this model, the phase space is again described by the set of all genealogies. The environment is modeled statistically by giving a probability distribution over the set of all possible observed genealogies for each genealogy generated by the population. The central concepts derived from this model are the adaptive value, the environmental capacity, and the environmental effective size. The adaptive value measures the correlation between the variability of the fecundity–mortality schedule and the environmental variability. It is a natural generalization of population entropy; in effect, it reduces to population entropy when the environment is constant. The environmental capacity is a function only of the environment and measures the maximum possible correlation that exists between a population and its environment. The environmental effective size represents the number of individuals that survive the random mortality induced by the environmental process. Our main result asserts that the asymptotic rate of the increase of this effective size is the environmental capacity. This is done in Section 7.3. Section 7.2 discusses a new notion of statistical equilibrium analogous to (3.2) and describes the structure of the equilibrium states. The material which follows develops the principal features of our statistical model.

Let (Ω, μ) denote a population, Ω the set of genealogies, and μ a stationary probability distribution on Ω. Each individual in the population generates a sequence of genealogies. An element in the sequence has the form

$$x = (\ldots x_0, x_1, x_2, \ldots, x_k, \ldots)$$

where the values x_k assume represent the age class at generation k. The environment acts instantaneously at each generation on each genealogy in the sequence. The action of the environment, which we assume to be the same at each generation, is described as follows.

At generation (k), say, an individual in age class (i) either (a) survives with probability α_i, subsequently reproducing at generation $(k + 1)$ according to the fixed age-specific fecundity, or (b) dies with probability $(1 - \alpha_i)$ and is assigned to the class of nonsurvivors which we denote (0). The environmental action thus generates new genealogies

$$x' = (\ldots x'_0, x'_1, x'_2, \ldots, x'_k, \ldots)$$

where x'_k will assume the value (i) with probability α_i and (0) with probability $1 - \alpha_i$. When $x'_k = 0$, the genealogy terminates.

Now, in general, the probability α_i will depend on the age of the individual, the age of his parent at his birth, the age of his grandparent at his parent's birth, and so on. If α_i depends only on the individual's age, we say that the environment has zero memory. If the dependence on ancestors ceases after m generations, we say that the environment has finite memory of order m, otherwise it has infinite memory. For zero-memory environments, the statistical properties of the environment can be completely described by the $(n) \times (n + 1)$ matrix $M = (m_{ij})$, where m_{ii} is the probability that an individual age (i) survives and $m_{i,n+1}$ the probability that he dies. The matrix M is given by

$$M = \begin{bmatrix} \alpha_1 & 0 & 0 & \cdots & & 1 - \alpha_1 \\ 0 & \alpha_2 & 0 & \cdots & & 1 - \alpha_2 \\ 0 & 0 & \alpha_3 & \cdots & & 1 - \alpha_3 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & \alpha_n & 1 - \alpha_n \end{bmatrix} \tag{7.0}$$

To formalize this model, we let Ω' denote the set of non-terminating genealogies resulting from the action of the environment. Let F and F' denote the Borel field generated by the cylinder sets in Ω and Ω', respectively. The environment is characterized by a probability distribution $\nu(x, \ )$ defined over F', one for each $x \in \Omega$. For $S \subset F'$, we interpret $\nu(x, S)$ as the probability that the induced genealogy x' belongs to the set S, given that the genealogy generated by the population is x.

The ideas of stationarity and finite memory can be formally described in the context of our model. Stationarity is described by the condition

$$\nu(T(x), T(S)) = \nu(x, S) \tag{7.1}$$

for every $x \in \Omega$ and $S \subset F'$, where T is the shift operator.

To represent the finite memory condition, we describe using the doubly infinite sequences Ω', the set of genealogies of length t, which we denote $\Omega'^{(t)}$. Thus the sequence $z = (z_0, z_1, \ldots, z_{t-1}) \in \Omega'^{(t)}$. Write

$$[z] = \{w : w \in \Omega', z_i = w_i, 0 \leqslant i \leqslant t-1\}$$

The memory is defined as the least nonnegative integer m, denoted m^*, which satisfies the condition

$$\begin{aligned} &\nu(x, [z]) = \nu(y, [z]) \qquad \text{for} \quad x, y \in \Omega, \quad z \in \Omega'^{(t)} \quad \text{if} \\ &x_i = y_i \qquad \text{for} \quad -m \leqslant i \leqslant t \quad (t = 1, 2, \ldots) \end{aligned} \tag{7.2}$$

In this model, randomness intervenes in two ways: (a) the set of genealogies $\{x\}$ generated by the population is random—this randomness is governed by the distribution μ; (b) the set of observed genealogies $\{x'\}$ is random—this randomness is determined by the environmental action. We are interested in the correlation between the generated distribution μ on Ω and the observed distribution μ' on Ω'. We require that this correlation reduces to population entropy when the environmental noise has no effect on mortality, and reduces to zero when no individuals survive. To describe this correlation we need to determine in terms of ν and μ the probability distributions on the space $\Omega \times \Omega'$ and on Ω'.

We let $A \subset \Omega$, $B \subset \Omega'$. Write $C = A \times B$. We think of the probability $\omega(C)$ of this subset of the space $\Omega \times \Omega'$ as the probability of the joint event $x \in A$, $x' \in B$. Now the distribution on Ω is determined by μ and for a given genealogy x, the distribution on Ω' is determined by ν. Therefore

$$\omega(C) = \omega(A \times B) = \int_A \nu(x, B)\, d\mu(x) \tag{7.3}$$

The distribution μ' on Ω' is given by

$$\mu'(B) = \omega(\Omega \times B) = \int_\Omega \nu(x, B)\, d\mu(x) \tag{7.4}$$

Since μ and ν are stationary, it is easily shown that the processes $(\Omega \times \Omega', \omega)$ and (Ω', μ') are both stationary, hence to each of the processes (Ω, μ), $(\Omega \times \Omega', \omega)$, and (Ω', μ') we can associate entropy functions H_μ, H_ω, and $H_{\mu'}$ respectively.

Let us write

$$\psi_\mu = H_\mu + H_{\mu'} - H_\omega \tag{7.5}$$

We call ψ_μ the adaptive value of the population.

The term is justified by the following facts. The quantity H_μ describes the *generated* variability and $H_{\mu'}$, the *observed* variability in the fecundity–mortality distribution; H_ω represents the joint variability in these two distributions. Thus the expression $H_\omega - H_{\mu'}$, which is clearly nonnegative, represents the discrepancy between the environmental variation and the age-specific variation in reproduction and mortality. If the population produces offspring at instances at which the environment has minimal effects on survivorship, then this discrepancy is small. The observed and generated variability will be small and the population will show little fluctuation in growth rate. In the language of the biologists such a population is adapted to its milieu. On the other hand, if reproduction coincides with periods during which the environment severely decreases survivorship, the discrepancy is large. The growth rate will now undergo fluctuations and extinction will result. Such a population is considered unadapted to its milieu.

We shall now distinguish between three types of environmental conditions which are easy to analyze. These conditions have a precise biological interpretation.

(1) We say that the environmental is of Type I if $H_\omega - H_{\mu'} = 0$ for all μ. Type I environments correspond to the case where the discrepancy between the environmental and the age-specific variations in fecundity and mortality is zero. In this situation, there is no environmentally induced mortality. For this reason we shall call Type I models constant environments.

Example 7.1. The interaction matrix due to the constant environment with zero memory is given by the matrix M, see (7.0), with $\alpha_i = 1$ for all i. The adaptive value $\psi = H$, where H is the population entropy given by (4.17).

(2) Type II environments have the property $H_\mu - (H_\omega - H_{\mu'}) = 0$ for all μ. This is characterized by the condition that the stochastic processes (Ω, μ) and (Ω', μ') are independent. This corresponds to the case in which there is high environmentally induced mortality.

Example 7.2. The interaction matrix due to a Type II environment with zero-memory is given by the matrix M with identical rows. This matrix has entries $\alpha_i = 0$ for all i. The adaptive value $\psi = 0$.

(3) Type III environments are described by the condition $H_\omega - H_\mu$ independent of the distribution μ. This class of environments, which we call *uniform*, occurs when mortality in each age class is affected to the same extent by the external environment.

Example 7.3. For a uniform environment with zero-memory, the rows of the matrix M are permutations of the same two numbers α, $1 - \alpha$. Hence $H_\omega - H_\mu = \alpha \log \alpha + (1 - \alpha)\log(1 - \alpha)$. The adaptive value is given by

$$\psi = H_{\mu'} - \left[\alpha \log \alpha + (1 - \alpha)\log(1 - \alpha)\right]$$

Now the adaptive value depends both on the population and the environment. We now introduce a quantity which depends only on the environment. We note that the function ψ is bounded above by $\log n$, where n denotes the number of age classes. Hence ψ_μ has a least upper bound K as μ varies over all stationary measures. We denote this least upper bound K_ν to indicate its dependence on the environment and we call it the environmental capacity. We shall show in Section 7.2 that K_ν measures the asymptotic rate of increase of a population in the given environment.

In Example 7.1, the capacity

$$K_\nu = \sup_{\{p_j\}} - \frac{\sum_{j=1}^{n} p_j \log p_j}{\sum_{j=1}^{n} j p_j} = \log 2$$

In Example 7.2

$$K_\nu = 0$$

In Example 7.3

$$K_\nu = \alpha \log 2 + \alpha \log \alpha + (1 - \alpha)\log(1 - \alpha)$$

7.2. Capacity and Equilibrium States

The structure of the equilibrium states for constant environment models, discussed in Sections 4 and 5, depends on the interaction Φ. In analyzing this structure the distinction between density-independent and density-dependent models was crucial. The aim of this section is to extend the notion of equilibrium state in constant environment models to the case of the population growth in variable environments.

The structure of the equilibrium state is determined by the kernel ν. We now introduce our new notion of equilibrium as follows.

Consider a population with phase space Ω. We call a distribution $\hat{\mu}$ on the set of genealogies Ω an equilibrium distribution for the environment if

$$\psi_{\hat{\mu}} = \sup_{\mu} \left(H_\mu + H_{\mu'} - H_\omega\right) \tag{7.6}$$

where the supremum is taken over all stationary distributions on Ω. Thus a population is in equilibrium with the environment if its adaptive value is equal to the environmental capacity K_ν.

Let E_ν denote the set of equilibrium states. To characterize the structure of these states, we draw on certain ideas from convex analysis.

We first note that the phase space Ω endowed with the topology given by the metric (3.0) is a compact space. Let $M(\Omega)$ denote the set of T-invariant probability measures on $M(\Omega)$. From the Riesz representation theorem, we may think of $M(\Omega)$ as a subset of the dual space $C(\Omega^*)$, where $C(\Omega)$ is the space of continuous real-valued functions on Ω, and we set $\mu(f) = \int f d\mu$ for $f \in C(\Omega)$. This gives a topology on $M(\Omega)$, the so-called weak topology. With this topology, $M(\Omega)$ becomes a compact metrizable space. It is also a convex subset of $C(\Omega)^*$. Moreover, the extreme points of $M(\Omega)$ are precisely the ergodic invariant measures on Ω.

We can now state the following theorem:

Theorem 7.0. For zero-memory environments, the set of equilibrium states E_ν for a given environment ν is a nonempty set of probability measures which is convex and closed, hence compact in the weak topology. The extremal points of this set are exactly the invariant equilibrium states which are the extremal points of the set of all invariant probability measures, that is, the ergodic equilibrium states.

The proof of this theorem rests on showing that ψ_μ is upper semicontinuous on $M(\Omega)$. To show this we need a characterization of ψ_μ in terms of finite probability spaces.

Write

$$[x] = \{z : z \in \Omega : z_i = x_i, 0 \leqslant i \leqslant n\}$$

$$[x'] = \{z' : z' \in \Omega' \, z'_i = x'_i, 0 \leqslant i \leqslant n\}$$

Define

$$H_\mu^{(n)} = -\frac{1}{n} \sum_{x_0, x_1, \ldots, x_n} \mu[x] \log \mu[x] \tag{7.7}$$

$$H_{\mu'}^{(n)} = -\frac{1}{n} \sum_{x'_0, x'_1, \ldots, x'_n} \mu[x'] \log \mu[x'] \tag{7.8}$$

$$H_\omega^{(n)} = -\frac{1}{n} \sum_{\substack{x_0 x_1, \ldots, x_n \\ x'_0 x'_1, \ldots, x'_n}} \omega([x],[x']) \log \omega([x],[x']) \tag{7.9}$$

It is known that

$$H_\mu^{(n)} \to H_\mu, \quad H_{\mu'}^{(n)} \to H_{\mu'}, \quad \text{and} \quad H_\omega^{(n)} \to H_\omega$$

Hence

$$\psi_\mu = \lim_{n \to \infty} \left[H_\mu^{(n)} + H_{\mu'}^{(n)} - H_\omega^{(n)} \right]$$

We now have the following proposition:

Proposition. ψ_μ is an upper semicontinuous function on $M(\Omega)$.

Proof. We shall show that $H_{\mu'}$ and $H_\mu - H_\omega$ are each upper semicontinuous.

Consider $H_{\mu'}^{(n)}$. Using (7.8) and the zero-memory condition, it is clear that for each n, $H_{\mu'}^{(n)}$ is continuous. Moreover from (7.8) $H_{\mu'}^{(n+1)} \leqslant H_{\mu'}^{(n)}$. Since $H_{\mu'}$ is the limit of a decreasing sequence of continuous functions, it is upper semicontinuous. The same argument applies to $H_\mu - H_\omega$. Using the fact that the sum of two upper semicontinuous functions is upper semicontinuous, the proof is complete. ■

The above proposition and our remarks concerning the structure of $M(\Omega)$ immediately yield the proof of Theorem 7.0.

Proof of Theorem 7.0. It is sufficient to observe that since ψ_μ is upper semicontinuous and is bounded, the ψ_μ assumes its supremum on at least one of the extremal points of the set $M(\Omega)$. ■

Remark 7.0. If the environment is constant or uniform, then $|E_\nu| = 1$. This follows from the fact that the mapping $H_\mu : M(\Omega) \to R$ assumes its supremum on a unique point of the set $M(\Omega)$. The equilibrium state in this case corresponds to pure thermodynamic phases.

Remark 7.1. We should emphasize that Theorem 7.0 assumes that the environment has zero memory. The extension to finite or infinite memory environment is not fully understood. However, preliminary studies indicate that the condition $|E_\nu| > 1$, corresponding to phase transition, seems possible even for zero-memory systems.

Remarks 7.0 and 7.1 should be contrasted with our results concerning the structure of the set of equilibrium states E_Φ for a given potential Φ on the phase space. In these cases we recall that for *all* density-independent models, we have $|E_\Phi| = 1$ and the unique $\mu \in E_\Phi$ is a Gibbs state. The case $|E_\Phi| > 1$, corresponding to mixed phases only occurs for a certain class of density-dependent models.

7.3. Environmental Effective Size

We will first formally define the notion of effective size and then interpret the mathematical operations invoked. We denote by $\Omega^{(m+1)}$ and $\Omega'^{(t)}$, the set of genealogies of the form $(x_0, x_1, \ldots, x_{m+t})$ and $(x_0', x_1', \ldots, x_t')$, respectively.

For each $x \in \Omega^{m+t}$, write

$$[x] = \{z : z \in \Omega, z_i = x_i \ 0 \leqslant i \leqslant m+t\}$$

and for $x' \in \Omega'^{(t)}$, write

$$[x'] = \{z' : z' \in \Omega', z'_i = x'_i, 0 \leqslant i \leqslant t\}$$

Let us assume the finite memory condition given by (7.2), and that the environment has memory m^*.

Then for each $x' \in \Omega'^{(t)}$, $x \in \Omega^{(m+t)}$ with $y, \tilde{y} \in T^m[x]$

$$\nu(y, [x']) = \nu(\tilde{y}, [x']) \qquad \text{if} \quad m \geqslant m^*$$

We can therefore write for each $x \in \Omega^{m+t}$ and $E \subset \Omega'^{(t)}$

$$\nu_t(x, E) = \sum_{x' \in E} \nu(y, [x']) \qquad \text{if} \quad y \in T^m[x] \tag{7.10}$$

Now let L denote the set of mappings $f : \Omega'^{(t)} \rightarrow \Omega^{(m+t)}$.

Let X denote the set of elements $x \in \Omega^{(m+t)}$ such that

$$\nu_t\big(x, f^{-1}(x)\big) > 1 - \epsilon \tag{7.11}$$

Write

$$N_t^{**}(f, \epsilon) = \operatorname{card} X$$

and

$$N_t^{**}(\epsilon) = \max N_t^{**}(f, \epsilon) \tag{7.12}$$

the maximum being taken over all mappings f. We call $N_t^{**}(\epsilon)$ the effective size.

We now explain the ideas behind the above operations and our reason for calling $N_t^{**}(\epsilon)$ the effective size. In order to make the explanation more intuitive, we shall assume in the discussion that follows that the environment has zero memory, that is $m = 0$.

To understand the connection between genealogies and population size, a connection which is crucial to our interpretation of $N_t^{**}(\epsilon)$, we note that each individual at generation (t) has a unique "backward" genealogy which we write $x_t, x_{t-1}, x_{t-2}, \ldots, x_0$. Thus to each genealogy generated up to time (t) corresponds a single individual. This implies that the number of genealogies generated up to time (t) describes the total number of individuals, that is, the population size at time (t).

Now consider the set of genealogies generated up to time (t). At generation (k) say, each individual age (i) survives with probability (α_i) and dies with probability $(1 - \alpha_i)$. Thus the generated genealogy

$$x = (x_0, x_1, \ldots, x_k, \ldots)$$

with $x_k = (i)$, is transformed into a genealogy

$$x' = (x'_0, x'_1, \ldots, x'_k, \ldots)$$

where $x'_k = (i)$ with probability (α_i) and (0) with probability $(1 - \alpha_i)$.

The genealogies with $x'_k = 0$ are the terminating genealogies since the symbol (0) represents the nonsurvivors. The terminating genealogies make no further contribution to the population size. This contribution is completely determined by the nonterminating sequences. Our problem is to characterize the number of nonterminating sequences.

Now given a mapping $f: \Omega'^{(t)} \to \Omega^{(t)}$ we observe that if $x, y \in \Omega^{(t)}$ then $f^{-1}(x) \cap f^{-1}(y) = \emptyset$.

Thus to each f we can associate a partition of $\Omega'^{(t)}$ into disjoint sets. Furthermore for each f, we can consider the expression $\nu_t(x, f^{-1}(x))$ as describing the probability that for a given genealogy x generated, the genealogy x' observed belongs to the set $f^{-1}(x)$. The condition (7.11) means that the genealogy x' observed when x is generated belongs to the set $f^{-1}(x)$ with arbitrary high probability.

Now each element $x' \in \Omega'^{(t)}$ is a finite sequence derived from the set of nonterminating genealogies Ω'. Hence for each f, the set of elements $x \in \Omega$ satisfying (7.11) represents those generated genealogies which result in a nonterminating genealogy with arbitrary high probability. The set X thus describes the set of genealogies that are not perturbed by the environmental noise. X clearly depends on the partition induced by f. The maximum of the number of elements in X, this maximum being taken over all mappings f, thus gives the number of genealogies that have not undergone any distortion as a result of the environmental noise. Owing to the correspondence between genealogies and individuals, this number represents the total number of individuals whose mortality is not affected by the environmental noise. This is the effective size given by (7.12).

We can now state the main result of this section.

Theorem 7.1. $\lim_{t\to\infty}(1/t)\log N_t^{**}(\epsilon) = K_\nu$

The proof of this theorem, Ref. 25, exploits arguments in coding theory$^{(26,27)}$ and is based on two inequalities valid for large t:

(i) $N_t^{**}(\epsilon) < e^{t(K_\nu+\epsilon)}$

(ii) $N_t^{**}(\epsilon) > e^{t(K_\nu-\epsilon)}$

Statement (i) is based on Feinstein's theorem.$^{(26)}$ Statement (ii) is derived using ideas due to Winkelbauer.$^{(28)}$

Theorem 7.1 should be constrasted with (2.9) and (6.12) which relates other measures of population size with the Malthusian parameter and entropy.

Furthermore we note that for a population (Ω, μ) in equilibrium with the environment, the adaptive value ψ_μ coincides with the capacity K_ν. The effective size is therefore given asymptotically by

$$N_t^{**} \sim e^{t\psi_\mu} \tag{7.13}$$

Remark. We give a heuristic account of (7.13). Consider a set of genealogies generated by an individual aged (1), say. After t generations, for t large, this set falls into two classes, a class consisting of about e^{tH_μ} elements, corresponding to the typical genealogies, and a class of remaining genealogies of low probability. Each genealogy generated up to time t corresponds to an individual. The typical genealogies will generate as a result of the environmental disturbance, approximately $e^{t(H_{\mu'}-H_\omega)}$ genealogies. Hence the total number of genealogies resulting, that is, the effective size, is given by

$$e^{t(H_\mu)} \cdot e^{t(H_{\mu'}-H_\omega)} = e^{t\psi_\mu}$$

8. CONCLUSION

We shall make a resume of the main results of this paper from a point of view different from that which was developed in the preceding. We will describe the biological problems that we have analyzed and then indicate the role statistical mechanics and information theory have played in the analysis of these problems. This treatment should make more explicit the connection between population biology, statistical physics, and information theory, which is the main theme of the paper.

Classical models of population dynamics are based on two main concepts, the Malthusian parameter r and the census population size N_t. At equilibrium the relation between these two parameters is given by

$$N_t \sim e^{tr} \tag{8.0}$$

This result is the cornerstone for most work in population genetics and ecology, and the fundamental theorem of natural selection as formulated by Fisher(29) is based explicitly on this relation. However, the models on which this result is based essentially neglect two factors:

(i) Populations are demographically heterogeneous, that is, their birth and death rates are age dependent.

(ii) The age-dependent birth and death rates are subject to fluctuations due to the external environment.

This paper has dealt with these two factors by introducing a new class of population models that incorporates demographic heterogeneity and environmental variation.

In constant environment models, the two main parameters introduced are population entropy H and the effective population size. The entropy H measures the spread of the fecundity–mortality distribution. The effective population size N_t^* describes, roughly speaking, the total number of reproductives in the population. The connection between these two parameters is given by the relation

$$N_t^* \sim e^{tH} \tag{8.1}$$

In variable environments, the two main parameters introduced are the adaptive value ψ and the effective size N_t^{**}. The measure ψ represents the degree to which the fecundity–mortality distributions of the individuals in the population are matched to the environmental variation. The parameter N_t^{**} describes the number of reproductives whose mortality is unaffected by the environmental disturbance. At equilibrium, we have the relation

$$N_t^{**} \sim e^{t\psi} \tag{8.2}$$

The connection between population biology, statistical physics, and information theory lies in the methods that we have used to obtain (8.1) and (8.2). These methods are distinct from the techniques that have been applied in classical population models.

The analyses of classical age-structured models are based on the dynamical equation

$$\bar{x}(t+1) = A(t)\bar{x}(t) + \bar{\eta}(t) \tag{8.3}$$

where $\bar{x}(t)$ denote the age distribution at time t, $A(t)$ the population matrix, and $\bar{\eta}(t)$ an environmental factor. The critical parameter in this class of models is the age distribution. One investigates how this age distribution evolves in time.

In the formulation we have introduced, we consider the statistical representation of (8.3). This is given at equilibrium by the system (Ω, μ, T), where Ω denote the set of genealogies, T the shift operator, and μ a T-invariant probability measure on Ω. The critical parameter in our models is the genealogy of an individual and we study the distribution of genealogies under different kinds of interactions between individuals.

In the constant environment models, $\bar{\eta}(t) = 0$. The matrix $A(t)$ is characterized by a potential function Φ on the space of genealogies. Equilibrium states are described by measures $\hat{\mu}$ which satisfy (3.2)

The potential function Φ is chosen in such a way that the equilibrium states in the dynamical model and its statistical representation coincide. The expression (3.2) has its origins in statistical physics and constitutes the basis for the thermodynamic analogs we have derived. The population entropy H is precisely the Kolmogorov–Sinai entropy $H_{\hat{\mu}}(T)$, in the equilibrium state $\hat{\mu}$. The analogy with phase transitions is based on the fact that the set of equilibrium states E_Φ for a certain class of interactions Φ has the property $|E_\Phi] > 1$. The case $|E_\Phi| = 1$ which arises in density-independent models corresponds to "pure thermodynamic phases."

In the variable environmental model, $\bar{\eta}(t) \neq 0$. In the statistical representation of these models, three stochastic processes are involved: namely, (Ω, μ), (Ω, ν, Ω'), and (Ω', μ'). The population process generates genealogies $x \in \Omega$, the probability distribution on the genealogies being described by μ. The environmental action which is described by the measure ν perturbs genealogies $x \in \Omega$. The measure μ' is the measure induced on Ω'.

Equilibrium states for this class of models are described by measure $\hat{\mu}$ which satisfies (7.6). This expression has its roots in information theory and provides the basis for the analogs between the information theoretic concepts we have noted.

In this class of population–environment interaction, the analogy with phase transition arises from the fact that for a certain class of interaction ν, the set E_ν of equilibrium states has the property $|E_\nu| > 1$. By contrast with the constant environment models, we should point out that the structure of the equilibrium states for this class of models is incompletely understood.

The asymptotic relations (8.1), (8.2) should be contrasted with the classical asymptotic relation (8.0). In effect, population entropy, adaptive value, and the Malthusian parameter are all related.

For constant environment models, we have

$$r = H + \bar{\Phi} \tag{8.4}$$

where $\bar{\Phi}$ denotes the reproductive potential. The expression $\bar{\Phi}$ is a measure of demographic heterogeneity. Thus when $\bar{\Phi} = 0$, the entropy and the Malthusian parameter coincide. In this case, the effective population size reduces to the census size.

For variable environments

$$\psi = H - H^* \tag{8.5}$$

where H^*, which is equal to $H_\omega - H_{\mu'}$, measures the degree of variability of the environment. Thus when $H^* = 0$, a constant environment, the adaptive value is precisely the entropy and the two measures of effective size coincide.

As a final remark, we should point out that although we have restricted our discussion to demographic models, the analysis we have described applies to a wide class of models, in particular, models of cellular development and the evolution of macromolecules(30). Similar analogs to thermodynamic relations exist, and the new concepts which emerge from the application of statistical physics ideas play an important role in the study of these phenomena.

ACKNOWLEDGMENTS

I would like to thank the referee for his careful reading of the manuscript. This work was begun at the University of Paris with the partial support of the Guggenheim Foundation and completed in Göttingen with the support of the von Humboldt Foundation.

REFERENCES

1. R. Dobrushin, *Functional Anal. Appl.* **2**:302 (1968).
2. R. Bowen and D. Ruelle, *Invent. Math.* **29**:181 (1975).
3. Y. Sinai, *Functional Anal. Appl.* **2**:64 (1968).
4. L. Demetrius, *Proc. Natl. Acad. Sci.* **71**:4645 (1974).
5. C. Shannon, *Bell System Tech. J.* **27**:379 (1948).
6. L. Demetrius, *Nature* **274**:213 (1978).
7. L. Demetrius, *C. R. Acad. Sci.* **382**:471 (1980).
8. L. Demetrius, *Proc. Natl. Acad. Sci. USA* **74**:384 (1977).
9. L. Demetrius, *Genetics* **79**:535 (1975).
10. L. Demetrius, *Demography* **16**:328 (1979).
11. L. Demetrius, *Population* **4**:869 (1979).
12. P. Leslie, *Biometrika* **33**:183 (1945).
13. A. Lotka, *Proc. Natl. Acad. Sci.* **8**:339 (1922).
14. D. Ruelle, *Thermodynamic Formalism* (Addison-Wesley, Reading, Massachusetts, 1979).
15. D. Mayer, *The Ruelle–Araki Transfer Operator in Statistical Mechanics, Lecture Notes in Physics* 123 (Springer-Verlag, Berlin, 1980).
16. R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Differomorphisms, Lecture Notes in Mathematics* 470 (Springer-Verlag, Berlin, 1975).
17. P. Walters, *Am. J. Math.* **97**:937 (1975).
18. V. Rohlin, *Russian Math. Surveys* **22**:52 (1967).
19. P. Billingsley, *Ergodic Theory and Information* (Wiley, New York, 1965).
20. F. Dyson, *Commun. Math. Phys.* **21**:269 (1971).
21. M. Fisher and B. Felderhof, *Ann. Phys.* (*N.Y.*) **58** (1977).
22. F. Hofbauer, *Trans. Am. Math. Soc.* **228**:223 (1977).
23. F. Ledrappier, *Monatsh. Math.* **83**:147 (1977).
24. D. Vere-Jones, *Pacific J. Math.* **22**:361 (1967).
25. L. Demetrius, preprint.
26. A. Feinstein, *Foundations of Information Theory* (McGraw-Hill, New York, 1958).
27. J. Wolfowitz, *Coding Theorems in Information Theory* (Springer-Verlag, New York, 1964).
28. K. Winklebauer, Trans. of the Second Prague Conf. on Inf. Theory, Prague, p. 685 (1960).
29. R. Fisher, *The Genetical Theory of Natural Selection* (The Clarendon Press, Oxford, 1930).
30. L. Demetrius, *Jour. Theor. Biol.* (in press).