

Article

Principle of Minimum Discrimination Information and Replica Dynamics

Georgiy P. Karev

Lockheed Martin MSD, National Institutes of Health, Bldg. 38A, Rm. 5N511N, 8600 Rockville Pike, Bethesda, MD 20894, USA; E-Mail: karev@ncbi.nlm.nih.gov

Received: 4 March 2010; in revised form: 15 June 2010 / Accepted: 18 June 2010 /

Published: 28 June 2010

Abstract: Dynamics of many complex systems can be described by replicator equations (RE). Here we present an effective method for solving a wide class of RE based on reduction theorems for models of inhomogeneous communities. The solutions of the RE minimize the discrimination information of the initial and current distributions at each point of the system trajectory, not only at the equilibrium, under time-dependent constraints. Applications to inhomogeneous versions of some conceptual models of mathematical biology (logistic and Ricker models of populations and Volterra' models of communities) are given.

Keywords: replicator equation; selection system; model reduction; discrimination information; cross-entropy

AMS subject classification: 92B05; 34G99

1. Introduction

Replicator equations describe the dynamics of distributions in heterogeneous populations and communities under selective forces, when the heterogeneity implies existence of selective differences between individuals. One of the first replicator equations was used by Fisher, Haldane, and Wright to study the evolution of multi-allelic one-locus gene frequencies under the force of natural selection (for more information see [1]). Another well-known source of replicator equations comes from the evolutionary game theory [2,3].

A very high or even infinite system dimensionality is one of the most fundamental difficulties in the study of replicator equations. Another approach to inference of unknown distribution p subject to some given testable information (*constraints*) about the system can be based on the Principle of minimum discrimination information, MinxEnt (in equivalent terms, the Principle of maximum information entropy, MaxEnt). The divergence between the distribution p and reference distribution m can be measured with information discrimination $I[p:m]$ (known also as KL-divergence between p and m , see s.2 for definitions). The MinxEnt principle [4] states that, given new information, a new distribution p should be chosen in such a way as to minimize $I[p:m]$; see also [5,6] for rationalization and applications of the MaxEnt principle.

A grave objection against this approach is that the MaxEnt principle does not follow from the basic laws and fundamental theories and hence may or may be not postulated as an independent assertion. The problem can be eliminated for some particular systems if one can derive the MaxEnt principle from the system dynamics.

Generally, in applications to dynamical systems the MinxEnt principle is used to estimate the unknown distribution p at given constraints when the system is *in equilibrium*. Recently it was shown that the MinxEnt principle is valid as an exact theorem for a wide class of *selection systems* not only in equilibrium states but also at every point along the system trajectory. More precisely, it was proven in [7], that: 1) some complex models of selection systems can be reduced to an escort system of ordinary differential equations; 2) the solutions of corresponding RE have the form of time-dependent Boltzmann distributions (in other terms, they belong to the exponential family of distributions), and conversely, every time-dependent Boltzmann distribution satisfies a replicator equation; for a simplified model of the selection system it was shown in [8] that 3) the Dynamical principle of minimum of discrimination information (MinxEnt) is valid: the solution to the RE minimizes the KL-divergence of the initial and current distributions under some natural constraints *at every instant*; these constraints can in turn be computed explicitly at every moment from the system dynamics. The obtained results were illustrated on some simple models of free growing Malthusian inhomogeneous populations.

In this paper we consider a more general version of selection systems with self-regulation than the systems studied in [7]; it allows us accounting for a possible dependence of reproduction rate on different statistical characteristics of the set of traits given in the model, such as mean values, covariance and higher moments of the system distribution.

We formulate the reduction theorem (Section 3) and Dynamical MinxEnt principle for such systems (Section 4); the main results of these sections are similar to those obtained in [7,8], but they are relevant for more general selection systems. In Section 5 the results are applied to inhomogeneous versions of classical logistic and Ricker population models. In Section 6 we give the “conjugate” description of the solution to RE based on the Dynamical MinxEnt principle; we show that the KL-divergence is the Legendre transform of the logarithm of the partition function for corresponding time-dependent Boltzmann distributions; we also show that the solution to the escort system and the current mean values of the “traits” accounted for by the selection system are conjugate variables. In Section 7 we extend the reduction theorem and Dynamical MinxEnt principle to models of inhomogeneous communities and apply them to some classical Volterra’ type models of mathematical biology. Overall, three fundamental mathematical objects are under consideration in this paper:

MinxEnt principle, Exponential families of distributions, and Replica dynamics (selection system). There exist close interconnections between these objects (loosely speaking, they are in some sense equivalent). In particular,

(A) The Kullback’ (or Jaynes’, for MaxEnt) theorem states that the MinxEnt distribution belongs to the Exponential family, implying that MinxEnt => Exponential family;

(B) It is almost evident (and was mentioned in the literature [8,9]), that if the distribution belongs to the exponential family, then it satisfies the MinxEnt, suggesting the inverse implication: Exponential family =>MinxEnt.

(C) Any time-dependent exponential distribution solves corresponding replicator equation, so the implication: time-dependent Exponential family => Replicator equation becomes trivial;

(D) The major focus of this paper is on the inverse implication: Replica dynamics (selection system) => time-dependent Exponential family. It is composed of the reduction theorem and formulas for the system distribution and current constraints. Not only the Dynamical MinxEnt principle but also the reduction theorem follow from the fact that the solution to the replicator equation belongs to the exponential family. In this case all *current* statistical characteristics of the system can be computed with the help of the moment generating function (or, more generally, by generating functional) for the *initial* distribution. This implies that one can construct a closed escort system for auxiliary variables, whose time derivatives are equal to the “weights” of the traits, defining the reproduction rate. These variables coincide with the Lagrange multipliers for time-dependent MinxEnt distribution. The dimensionality of the escort system does not depend on the dimensionality of the initial model and is equal to the number of traits. Exact formulations are given in s.3 and Mathematical Appendix.

2. The KL-Divergence and MinxEnt Principle

In the case of continuous distributions the discrimination information, or the KL - divergence of the distribution *p* from *m*, is:

$$I[p : m] = \int_A p(x) \ln \frac{p(x)}{m(x)} dx = E_p [\ln \frac{p}{m}] \tag{2.1}$$

The value $S[p : m] = -I[p : m]$ is known also as the relative entropy. The probability density function (pdf) *m* is assumed to be given. The inference of *p* by minimizing $I[p : m]$ (equivalently, by maximizing $S[p : m]$) is known as the Principle of minimum discrimination information, MinxEnt (accordingly, the Principal of Maximum relative or cross-entropy, MaxEnt).

Assume that expected values A_s of some variables $\varphi_s, s = 1, \dots, n$ over the pdf *p* are known: $E_p[\varphi_s] = A_s$. Then the “MinxEnt distribution” *p** that minimizes the discrimination information $I[p : m]$ subject to these constraints is:

$$p^*(\mathbf{a}) = \frac{1}{Z} m(\mathbf{a}) \exp\left(\sum_{s=1}^n \lambda_s \varphi_s(\mathbf{a})\right) \tag{2.2}$$

where $Z(\boldsymbol{\lambda}) = E_m[\exp(\sum_{s=1}^n \lambda_s \varphi_s)]$ is known as a partition function and the Lagrange multipliers λ_s solve the system:

$$\partial \ln Z / \partial \lambda_s = A_s \quad (2.3)$$

Distributions of the form (2.2) belong to the exponential family of distribution [10]. The minimized value of the discrimination information is:

$$I[p^* : m] = -\ln Z(\lambda) + \sum_{s=1}^n \lambda_s A_s \quad (2.4)$$

In the MinxEnt distribution (2.2), the information concerning the constraints A_s is encoded in the set of Lagrange multipliers λ_s via equation (2.3), given the reference pdf m ; conversely, if λ_s in (2.2) are known, then the MinxEnt distribution is also known and hence the mean values A_s can be computed. In other words, the MinxEnt principle implies the equivalence of description of the distribution by the set of constraints and by the set of Lagrange multipliers. Below (see Section 6) we show that for replica dynamics this equivalence is universal and does not depend on the MinxEnt principle.

3. Selection Systems with Self-regulation and the Reduction Theorem

The selection system is a mathematical model of an inhomogeneous population, in which every individual is characterized by a vector-parameter $(a_1, \dots, a_n) = \mathbf{a}$ that takes on values from set A . The parameter \mathbf{a} specifies an individual's inherited invariant properties and does not change with time; the set of all individuals with a given value of the vector-parameter \mathbf{a} in the population is called \mathbf{a} -clone. Let $l(t, \mathbf{a})$ be the density of the population at the moment t over the parameter \mathbf{a} , so that the total population size:

$$N(t) = \int_A l(t, \mathbf{a}) d\mathbf{a} \quad (3.1)$$

and the current population distribution $P(t, \mathbf{a}) = l(t, \mathbf{a}) / N(t)$. Denote $F(t, \mathbf{a})$ to be the per capita reproduction rate (Malthusian fitness) at the moment t . We suppose that the reproduction rate of every \mathbf{a} -clone does not depend on other clones but can depend on \mathbf{a} and on some general population characteristics such as the total population size. These quantities evolve with time providing some self-regulation of the system dynamics. For example, the reproduction rate for the logistic model is proportional to $(1 - N(t)/B)$ where B is the upper boundary of the population size; the reproduction rate of the *Ricker*' model is proportional to $\exp(-\beta N(t))$.

An abstract selection system (or, in the author's terms, a system with inheritance) was studied in [11] (see also references to earlier work therein) where a general selection theorem was proven.

In [7] a class of selection systems with self-regulation was studied and a reduction theorem was proved; the theorem gives an effective algorithm for investigation of the selection systems and corresponding replicator equations. Below we formulate a more general version of this theorem.

It was assumed in [7] that the individual reproduction rate can depend on two types of integral characteristics of the system ("regulators"): the extensive characteristics, which depend on the total size of the system (as in most population models) and intensive characteristics, which do not depend on the total size but only on the population frequencies (as in most genetic models). The intensive characteristics are of the form:

$$H_k(t) = \int_A h_k(\mathbf{a})P(t, \mathbf{a})d\mathbf{a} = E_t[h_k] \tag{3.2}$$

and the extensive characteristics:

$$G_i(t) = \int_A g_i(\mathbf{a})l(t, \mathbf{a})d\mathbf{a} = N(t)E_t[g_i] \tag{3.3}$$

where g_i, h_k are appropriate weight functions.

Overall, we specify for each model a finite set of the extensive regulators $\mathbf{G}(t) = G_1(t), \dots, G_m(t)$, which contains the total population size; we assume that the individual reproduction rate can depend on this set of regulators at each time moment.

If we assume the overlapping generations and smoothness of $l(t, \mathbf{a})$ in t for each $\mathbf{a} \in A$, then the population dynamics can be described by the following master model:

$$dl(t, \mathbf{a}) / dt = l(t, \mathbf{a})F(t, \mathbf{a}) \tag{3.4}$$

$$F(t, \mathbf{a}) = \sum_{i=1}^n u_i(t, \mathbf{G})\varphi_i(\mathbf{a}) \tag{3.5}$$

here $u_i(t, \mathbf{G})$ are continuous functions .

The initial distribution $P(0, \mathbf{a})$ and the initial population size $N(0)$ are assumed to be given. The current system distribution $P(t, \mathbf{a}) = l(t, \mathbf{a}) / N(t)$ solves the replicator equation:

$$dP(t, \mathbf{a}) / dt = P(t, \mathbf{a})(F(t, \mathbf{a}) - E_t[F]) \tag{3.6}$$

The mathematical form of the fitness (3.5) suggests (from a biological point of view) that the individual fitness depends on a given finite set of traits.

The function $\varphi_i(\mathbf{a})$ in (3.5) may describe quantitative contribution of a particular i -th trait to the total fitness and then $u_i(t, \mathbf{G})$ describes the relative importance (weight) of the trait contributions, which at every time moment can depend on the state of the environment, population size, the mean, variance, covariance, and other statistical characteristics of the traits. We emphasize that the model accounts for the interactions between the traits only with the help of a given set of regulators. For example, if one needs to account for all moments up to the 2nd order, the following set of regulators should be used:

$$N(t) = \int_A l(t, \mathbf{a})d\mathbf{a}, G_i(t) = \int_A \varphi_i(\mathbf{a})l(t, \mathbf{a})d\mathbf{a}, G_{ik}(t) = \int_A \varphi_i(\mathbf{a})\varphi_k(\mathbf{a})l(t, \mathbf{a})d\mathbf{a} \tag{3.7}$$

Then, the covariance between the traits φ_i, φ_k at the moment t is the function of these regulators:

$$u_i(t, \mathbf{G}) = Cov[\varphi_i, \varphi_k](t) = G_{ik}(t) / N(t) - G_i(t)G_k(t) / N^2(t) \tag{3.8}$$

Clearly, this way one can account for the dependence of the fitness on mixed moments of any order; however, the approach that is described below is truly useful only when considers just a few regulators.

In model (3.5)-(3.6) the regulators and hence the reproduction rate $F(t, \mathbf{a})$ are not given explicitly but should be computed using the current pdf $P(t, \mathbf{a})$ at each time moment, so in the general case, the

model is a nonlinear equation of infinite dimensionality. Nevertheless, it can be reduced to a Cauchy problem for the escort system of ODE. For a less general version of the model (which allows dependence of the functions u_i on a single regulator only) the reduction theorem was proven in [7]. Below we formulate a more general version of this theorem, which gives an effective algorithm for investigation of the selection systems and the corresponding replicator equations.

Introduce the *generating functional*:

$$\Phi(r; \lambda) = \int_A r(\mathbf{a}) \exp\left(\sum_{i=1}^n \lambda_i \varphi_i(\mathbf{a})\right) P(0, \mathbf{a}) d\mathbf{a} \tag{3.9}$$

where $\lambda = (\lambda_1, \dots, \lambda_n)$, and $r(\mathbf{a})$ is a measurable function on A .

Define *auxiliary variables* as a solution to the *escort system* of differential equations:

$$dq_i / dt = u_i(t, \mathbf{G}^*(t)), q_i(0) = 0, i = 1, \dots, n \tag{3.10}$$

where $\mathbf{G}^*(t) = G_1^*(t), \dots, G_m^*(t)$, and:

$$G_k^*(t) = N(0)\Phi(g_k, \mathbf{q}(t)), \mathbf{q}(t) = (q_1(t), \dots, q_n(t)) \tag{3.11}$$

Denote:

$$K_t(\mathbf{a}) = \exp\left(\sum_{i=1}^n q_i(t)\varphi_i(\mathbf{a})\right) \tag{3.12}$$

Theorem 1. *Let $0 < T \leq \infty$ be the maximal value of t such that Cauchy problem (3.10) has a unique global solution $\{\mathbf{q}(t)\}$ at $t \in [0, T)$. Then the functions:*

$$\begin{aligned} l(t, \mathbf{a}) &= l(0, \mathbf{a})K_t(\mathbf{a}) \\ G_k(t) &= G_k^*(t) = N(0)\Phi(g_k; \mathbf{q}(t), \mathbf{s}(t)) \end{aligned} \tag{3.13}$$

satisfy system (3.4)-(3.5) at $t \in [0, T)$.

In particular the total size of the population:

$$N(t) = N(0)\Phi(1; \mathbf{q}(t)) = N(0)E_0[K_t] \tag{3.14}$$

As a corollary, we obtain the central formula for the current distribution of the system:

$$P(t, \mathbf{a}) = P(0, \mathbf{a})K_t(\mathbf{a}) / E_0[K_t] \tag{3.15}$$

In particular, $E_t[f] = E_0[fK_t] / E_0[K_t]$ for any (measurable) function.

Equality (3.12) shows that pdf (3.15) belongs to the exponential family of distributions [10]. In more “physical” terms, the pdf is the time-dependent Boltzmann distribution of the form $P(t, \mathbf{a}) = P(0, \mathbf{a}) \exp(B) / Z$ with the Boltzmann factor $\exp(B)$ where:

$$B(\mathbf{q}(t); \mathbf{a}) = \sum_{i=1}^n q_i(t)\varphi_i(\mathbf{a}) \tag{3.16}$$

and the partition function:

$$Z(\mathbf{q}(t)) = E_0[\exp(B(\mathbf{q}(t); \mathbf{a}))] = E_0[K_t] \tag{3.17}$$

Remark that in our case the partition function is completely known, given the initial pdf $P(0, \mathbf{a})$ and the solution to the Cauchy problem (3.10). Within the frameworks of selection system (3.5)-(3.6) the partition function has a clear biological meaning: $Z(\mathbf{q}(t)) = N(t)/N(0)$ is proportional to the current population sizes, which follows from formula (3.14).

4. Dynamical MinxEnt Principle

Comparing the distribution (3.15) with the MinxEnt distribution (2.2) one can conclude that under time-dependent constraints $A_i = E_t[\varphi_i]$ the solution to replicator equation (3.6) minimizes the KL-divergence of the initial and current distributions not only at the equilibrium but also at each point of the system trajectory. These constrains in turn can be computed explicitly at every instant depending on the system dynamics. The following theorems collect together corresponding mathematical assertions.

Theorem 2.

1) Let $P_t = P(t, \mathbf{a})$ be the solution (3.15) of replicator equation (3.6). Then at every moment t the distribution P_t provides minimum of $I[P_t : P_0]$ over all probability distributions compatible with the constraints $A_i(t) = E_t[\varphi_i]$, $i = 1, \dots, n$;

2) The values of constraints evolve due to escort system (3.10) and at each time moment can be computed using the formula:

$$A_i(t) = E_0[\varphi_i K_t] / E_0[K_t] = \Phi(\varphi_i; \mathbf{q}(t)) / \Phi(1; \mathbf{q}(t)) \tag{4.1}$$

3) Dynamics of the constraints are determined by the covariance equation:

$$dA_i(t) / dt = Cov_t[F, \varphi_i] \tag{4.2}$$

Theorem 3. The discrimination information $I[P_t : P_0]$ solves the covariance equation:

$$\frac{dI[P_t : P_0]}{dt} = Cov_t[B, F] \tag{4.3}$$

and can be computed using the following formulas:

$$\begin{aligned} I[P_t : P_0] &= E_0[BK_t] / E_0[K_t] - \ln E_0[K_t] \\ I[P_t : P_0] &= E_t[B] - \ln(N(t) / N(0)) \end{aligned} \tag{4.4}$$

where $F = F(t, \mathbf{a})$, $B = B(\mathbf{q}(t); \mathbf{a})$.

In the following section we demonstrate how Theorems 1-3 can be applied to some classical population models. A similar theory can be developed for replicator equations with discrete time and corresponding selection systems (maps) [13].

5. The Ricker’ and Logistic Inhomogeneous Models and the Dynamical MinxEnt Principle

5.1. Inhomogeneous Logistic Model

Many particular models of selection systems have the form of inhomogeneous logistic equation:

$$dl(t; \beta, \mu) / dt = l(t; \beta, \mu)[(\beta f_1(N(t)) - \mu f_2(N(t)))] \tag{5.1}$$

The general solution to this equation with distributed parameters was obtained in [7], example 5. Let $M(\lambda_1, \lambda_2) = E_0[\exp(\lambda_1\beta + \lambda_2\mu)]$ be the mgf of the joint initial distribution of β and μ . Then $l(t; \beta, \mu) = l(0; \beta, \mu)\exp(q_1(t)\beta + q_2(t)\mu)$ where q_1, q_2 solve the escort system:

$$\begin{aligned} dq_1 / dt &= f_1(N(0)M(q_1, q_2)) \\ dq_2 / dt &= -f_2(N(0)M(q_1, q_2)) \\ q_1(0) &= 0, q_2(0) = 0 \end{aligned} \tag{5.2}$$

The total population size $N(t) = N(0)M(q_1(t), q_2(t))$ and the current distribution:

$$P(t; \beta, \mu) = P(0; \beta, \mu)\exp(q_1(t)\beta + q_2(t)\mu) / M(q_1(t), q_2(t)) \tag{5.3}$$

Now we are able to apply the results of s.4. The discrimination information at moment t :

$$I[P_t : P_0] = \frac{E_0[(q_1(t)\beta + q_2(t)\mu)e^{q_1(t)\beta + q_2(t)\mu}]}{E_0[e^{q_1(t)\beta + q_2(t)\mu}]} - \ln E_0[e^{q_1(t)\beta + q_2(t)\mu}] \tag{5.4}$$

The distribution (5.3) provides the minimum of discrimination information, which is equal to (5.4) at each time moment among all distributions subject to the given mean values of birth and death rates at this moment:

$$E_t[\beta] = \frac{E_0[\beta \exp(q_1(t)\beta + q_2(t)\mu)]}{E_0[\exp(q_1(t)\beta + q_2(t)\mu)]}, \quad E_t[\mu] = \frac{E_0[\mu \exp(q_1(t)\beta + q_2(t)\mu)]}{E_0[\exp(q_1(t)\beta + q_2(t)\mu)]} \tag{5.5}$$

A particular case of equation (5.1):

$$dl(t; \beta, \mu) / dt = l(t; \beta, \mu)(\beta - \mu N(t)) \tag{5.6}$$

was studied in [12] for independent uniformly distributed parameters β and μ , $\beta \in [a_1, b_1]$, $\mu \in [a_2, b_2]$ and $P(0; \beta, \mu) = 1/((b_1 - a_1)(b_2 - a_2))$. It was proven in [7] that then $P(t; \beta, \mu) = P^1(t; \beta)P^2(t; \mu)$ where:

$$\begin{aligned} P^1(t; \beta) &= \frac{\exp(t\beta)}{(b_1 - a_1)E_0[\exp(t\beta)]} = \frac{t \exp(t\beta)}{\exp(tb_1) - \exp(ta_1)} \\ P^2(t; \mu) &= \frac{\exp(q_2(t)\mu)}{(b_2 - a_2)E_0[\exp(q_2(t)\mu)]} = \frac{q_2(t) \exp(q_2(t)\mu)}{\exp(q_2(t)b_2) - \exp(q_2(t)a_2)} \end{aligned} \tag{5.7}$$

The current mean values of the birth and death rates:

$$\begin{aligned}
 E_t[\beta] &= b_1 - \frac{b_1 - a_1}{1 - \exp((b_1 - a_1)t)} - \frac{1}{t} \rightarrow b_1 \\
 E_t[\mu] &= b_2 - \frac{b_2 - a_2}{1 - \exp((b_2 - a_2)q_2(t))} - \frac{1}{q_2(t)} \rightarrow a_2
 \end{aligned}
 \tag{5.8}$$

Distribution (5.7) provides the minimum of discrimination information at each time moment among all distributions concentrated in the rectangle $[a_1, b_1] \times [a_2, b_2]$ subject to the mean values of the birth and death rates (5.8).

In a general case, the solution to (5.6) is given by (5.3) at $q_1(t) = t$ and $q_2(t)$ that solves the equation $dq_2/dt = -N(0)M_0(t, q_2)$, $q_2(0) = 0$. The asymptotical behaviors of the solution to equation (5.6) vary dramatically depending on the initial distribution. Let the positive parameters β, μ be independent again, and the initial distributions of both parameters be exponential, $P_i(x) = s_i \exp(-xs_i)$. Let $s_1 = T, s_2 = 1$, and $N(0) = 1$ for simplicity. Then $q_1(t) = t, q_2(t) = 1 - \sqrt{1 - 2T \ln(1 - t/T)}$. The current system distribution $P(t; \beta, \mu) = P^1(t; \beta)P^2(t; \mu)$ where both marginal distributions are again exponential:

$$\begin{aligned}
 P^1(t; \beta) &= (T - t) \exp(-\beta(T - t)) \\
 P^2(t; \mu) &= \sqrt{1 - 2T \ln(1 - t/T)} \exp(-\mu \sqrt{1 - 2T \ln(1 - t/T)})
 \end{aligned}
 \tag{5.9}$$

$$\begin{aligned}
 E_t[\beta] &= \frac{1}{T - q_1(t)} = \frac{1}{T - t} \\
 E_t[\mu] &= \frac{1}{1 - q_2(t)} = \frac{1}{\sqrt{1 - 2T \ln(1 - t/T)}}
 \end{aligned}
 \tag{5.10}$$

Hence, the solution to equation (5.6) with the initial exponential distribution exists only up to the moment $t = T$. The total population size $N(t) = N(0) / \{(1 - t/T) \sqrt{1 - 2T \ln(1 - t/T)}\}$ tends to infinity as $t \rightarrow T$ and the population vanishes in any finite interval of values of both parameters, β and μ .

The discrimination information for $P_t = P^1(t; \beta)P^2(t; \mu)$

$$I[P_t : P_0] = \frac{t}{T - t} + \frac{1}{\sqrt{1 - 2T \ln(1 - t/T)}} - 1 + \ln(1 - t/T) + \ln(\sqrt{1 - 2T \ln(1 - t/T)})
 \tag{5.11}$$

The exponential distribution $P(t; \beta, \mu) = P^1(t; \beta)P^2(t; \mu)$ provides the minimum of the discrimination information over all the distributions subject to the mean values of the birth and death rates (5.10), and this minimum is equal to (5.11).

5.2. Inhomogeneous Ricker' Model

Let us consider the inhomogeneous version of the well known Ricker' equation:

$$dl(t; \beta, \mu) / dt = l(t; \beta, \mu) [(\beta \exp(-cN(t)) - \mu)]
 \tag{5.12}$$

The general solution to this equation with distributed parameters was obtained in [7], example 6. Let $M_0(\lambda_1, \lambda_2)$ be the mgf of the joint initial distribution of β and μ . Then:

$$l(t; \beta, \mu) = l(0; \beta, \mu) \exp(q(t)\beta - t\mu) \tag{5.13}$$

where the auxiliary variable $q(t)$ solve the Cauchy problem:

$$dq / dt = \exp(-cN(0)M_0(q, -t)), q(0) = 0 \tag{5.14}$$

The total population size $N(t) = N(0)M(q(t), -t)$, and the system distribution:

$$P(t; \beta, \mu) = P(0; \beta, \mu) \exp(q(t)\beta - t\mu) / M(q(t), -t) \tag{5.15}$$

Applying the results of s.4 we can compute the discrimination information at moment t :

$$I[P_t : P_0] = E_0[(q(t)\beta - t\mu) \exp(q(t)\beta - t\mu) / M(q(t), -t) - \ln M(q(t), -t)] \tag{5.16}$$

The current mean values of the parameters are given by:

$$E_t[\beta] = \frac{E_0[\beta \exp(q(t)\beta - t\mu)]}{E_0[\exp(q(t)\beta - t\mu)]}, E_t[\mu] = \frac{E_0[\mu \exp(q(t)\beta - t\mu)]}{E_0[\exp(q(t)\beta - t\mu)]} \tag{5.17}$$

The pdf (5.15) provides minimum of the discrimination information at every time moment subject to the constraints (5.17), and this minimum is equal to (5.16).

For example, let the parameters β and μ be independent and exponentially distributed in $[0, \infty)$ with the means s_1 and s_2 at the initial instant. Then $M(q, -t) = s_1 s_2 / ((s_1 - q)(s_2 + t))$, and:

$$dq / dt = \exp(-cN(0)s_1 s_2 / ((s_1 - q)(s_2 + t))), q(0) = 0 \tag{5.18}$$

This equation has a stable state $q = s_1$. As $t \rightarrow \infty$, $q(t) \rightarrow s_1$, the total population size tends to infinity and the population density concentrates at the value $\mu = 0$ of the parameter μ and vanishes in any finite interval of values of the parameter β . The distribution:

$$P(t; \beta, \mu) = P(0; \beta, \mu) \exp(q(t)\beta - t\mu) (s_1 - q)(s_2 + t) / s_1 s_2 \tag{5.19}$$

provides minimum of the discrimination information subject to the constraints:

$$E_t[\beta] = \frac{1}{s_1 - q(t)}, E_t[\mu] = \frac{1}{s_2 + t} \tag{5.20}$$

and this minimum is equal to:

$$I[P_t : P_0] = \frac{q(t)(s_2 + t)}{s_1 - q(t)} - \frac{t(s_1 - q(t))}{s_2 + t} + \ln(s_2 + t) + \ln(s_1 - q(t)) - \ln(s_1 s_2) \tag{5.21}$$

6. “Conjugative” Approach to the Selection System Dynamics

In the previous section we presented solutions to inhomogeneous logistic and Ricker models using the corresponding auxiliary variables and escort systems. These solutions minimize the information discrimination under certain constraints and hence can be found throw solving a conditioned optimization problem. Similar results can be obtained for other models of inhomogeneous populations. Let us clarify the interconnections between these two approaches.

Let us first come back to logistic model (5.1). Formally, the values $q_1(t), q_2(t)$ at t moment can be found independently on the system (5.2) by minimization of discrimination information (5.4) subject to the mean values $E_t[\beta]$, $E_t[\mu]$. We should emphasize that these constraints cannot be assigned arbitrarily but are completely defined by the system dynamics at the given initial distribution. Hence, if $E_t[\beta]$, $E_t[\mu]$ can be estimated independently, e.g., by data processing, then they must coincide with (5.5).

It implies that if a distribution P_t is such that $I[P_t : P_0]$ is equal to (5.4) and the mean values of the traits β, μ over P_t are equal to the estimated ones then P_t coincides with (5.3) and hence the values $q_1(t), q_2(t)$ are equal to the solution to system (5.2). Practically this means that we can solve equations (5.5) for $q_1(t), q_2(t)$ and this solution must coincide with the solution to system (5.2). To be more specific, let us consider a simple model (5.6) with exponentially distributed birth and death rates at the initial moment. For this model:

$$\begin{aligned} q_1(t) &= t = T - 1/E_t[\beta] \\ q_2(t) &= 1 - \sqrt{1 - 2T \ln(1 - t/T)} = 1 - 1/E_t[\mu] \end{aligned} \quad (6.1)$$

We see that the mean values $E_t[\beta]$, $E_t[\mu]$, if known, completely determine the auxiliary variables $q_1(t), q_2(t)$ and hence the solution to selection system (5.6) and the corresponding replicator equation given the initial distribution.

These arguments can be made rigorous for any selection system (3.5) and replicator equation (3.6). The following theorem states that the information discrimination $I[P_t : P_0]$ is the Legendre transform of the logarithm of partition function (3.17); the auxiliary variables $q_i(t)$ and the constraints $A_i(t) = E_t[\varphi_i]$ are conjugate under this transform (see MA for some definitions and proof).

Theorem 4.

- i) *The information discrimination $I[P_t : P_0]$ as a function of constraints $A_i(t)$ is the Legendre transform of $W = \ln E_0[K_t]$ as a function of variables $q_i(t)$, and conversely;*
- ii) *the variables $\mathbf{q}(t)$ are conjugate to the constraints $\mathbf{A}(t)$;*
- iii) *for given constraints $\mathbf{A}(t)$, the values of the variables $\mathbf{q}(t)$ and $I[P_t : P_0]$ can be found as a solution to optimization problem:*

$$I[P_t : P_0] = \sup \left\{ \boldsymbol{\alpha} : \sum_{i=1}^n \alpha_i A_i(t) - W(\boldsymbol{\alpha}) \right\} \quad (6.2)$$

and $q_i(t) = \alpha_i^*$ where α_i^* is the value at which the right hand side of (6.2) reaches its supremum.

It follows from the theorem that dynamics of the selection system and the corresponding replicator equation can be equally described either in terms of the auxiliary variables $q_i(t)$ or in terms of the constraints, i.e., the current mean values of the traits, $E_t[\varphi_i]$, and this equivalence *does not depend* on the MinxEnt principle. Technically, the former approach is more appropriate as the auxiliary variables can be found from the escort system. The latter approach is of principal importance, because it shows that for in order to completely determine the dynamics of system (3.5) and its distribution at any time moment it is enough to know only the mean values of the traits at this moment together with the initial distribution.

7. Inhomogeneous Models of Communities

7.1. Reduction Theorem and Dynamical MinxEnt

Consider the model of a community consisting from r interacting populations. We suppose again that every individual is characterized by their own value of vector-parameter \mathbf{a} . Let $l^j(t, \mathbf{a})$ be the density of j -th population at moment t . In this section we consider the model of an inhomogeneous community where the reproduction rates can depend on current characteristics of every population in the community composing a “regulator”. Formally, we consider the set of m regulators, each of which is the r -dimension vector-function $G_i(t) = (G_i^1(t), \dots, G_i^r(t))$, $i = 1, \dots, n$ where:

$$G_i^j(t) = \int_A g_i(\mathbf{a}) l^j(t, \mathbf{a}) d\mathbf{a} \tag{7.1}$$

Each regulator corresponds to appropriate weight function g_i . A finite set of the regulators corresponds to each specific model; we denote this set as $\mathbf{G}(t) = (G_1(t), \dots, G_n(t))$.

The current population sizes $N^j(t) = \int_A l^j(t, \mathbf{a}) d\mathbf{a}$ compose the regulator of a special importance, $N(t) = (N^1(t), \dots, N^r(t))$. We assume that $N(t)$ is included in the set of the model’s regulators. The distribution of j -th population in the community is by definition $P^j(t, \mathbf{a}) = l^j(t, \mathbf{a}) / N^j(t)$.

The model of inhomogeneous community considered here is of the form:

$$dl^j(t, \mathbf{a}) / dt = l^j(t, \mathbf{a}) F^j(t, \mathbf{a}) \tag{7.2}$$

$$F^j(t, \mathbf{a}) = \sum_{i=1}^n u_i^j(t, \mathbf{G}) \varphi_i(\mathbf{a}) \tag{7.3}$$

where the functions u_i^j can be specific for each trait and each population. The initial pdf-s $P^j(0, \mathbf{a})$ and the initial population sizes $N^j(0)$ are assumed to be given. The current pdf $P^j(t, \mathbf{a})$ solves the replicator equation:

$$dP^j(t, \mathbf{a}) / dt = P^j(t, \mathbf{a}) (F^j(t, \mathbf{a}) - E_i^j[F^j]) \tag{7.4}$$

The theory for inhomogeneous community model (7.1)-(7.4) is similar to the theory presented in Sections 3 and 4 for inhomogeneous populations up to more complex technical details. Theorem 5 (see MA for complete formulation) reduces complex model (7.1)-(7.3) to an escort system of ordinary non-autonomic equations of dimension $r \times n$ and gives the solution to replicator equation (7.4). Theorem 6 (MA) establishes the Dynamical MinxEnt Principle for the inhomogeneous community model and gives explicit formulas for discrimination information and constraints at each time moment. Let us apply the general theory to some classical models of biological communities consisting of interacting inhomogeneous populations.

7.2. Inhomogeneous Prey-predator Volterra' Model

The prey-predator Volterra' model in its simplest form reads:

$$\begin{aligned} dx/dt &= a_1x - a_2xy \\ dy/dt &= -a_3y + a_4xy \end{aligned} \tag{7.5}$$

where $x(t)$ and $y(t)$ denote prey and predator densities, a_1 is the reproduction rate of the prey population, a_2 is the per capita rate of the consumption of prey by the predators, a_3 is the death rate of the predator, and a_4/a_2 is the fraction of prey biomass, which is converted into predator biomass.

Let us consider the inhomogeneous version of this classical model supposing that parameters a_1 , a_2 , and a_3 are distributed and the ratio a_4/a_2 is fixed (and hence could be chosen equal to 1). We also suppose that the reproduction and death processes are specific for each subpopulation, while the consumption is driven by the interaction of the prey (predator) subpopulation with the entire predator (prey) population. Let $x(t; a_1, a_2)$, $y(t; a_3)$ be the densities of the prey and predator populations over parameters a_1, a_2 and a_3 correspondingly, and:

$$X(t) = \int_A x(t; a_1, a_2) da_1 da_2, Y(t) = \int_A y(t; a_3) da_3 \tag{7.6}$$

be the total sizes of the populations. The initial population sizes and initial distributions $P^1(0; a_1, a_2), P^2(0; a_2)$ are assumed to be given. The total rate of consumption is equal to:

$$Y(t) \int_A a_2 x(t; a_1, a_2) da_1 da_2 = Y(t) X(t) E_t^1[a_2] \tag{7.7}$$

Assuming the ‘‘proportional distribution’’ of prey among the predators we can write the inhomogeneous version of Volterra' model in the form:

$$\begin{aligned} dx(t; a_1, a_2) / dt &= x(t; a_1, a_2)(a_1 - a_2 Y(t)) \\ dy(t; a_3) / dt &= y(t; a_3)(G(t) - a_3) \end{aligned} \tag{7.8}$$

where $G(t) = \int_A a_2 x(t; a_1, a_2) da_1 da_2 = X(t) E_t^1[a_2]$.

Theorem 5 gives a method for studying this model and a more general model (7.2); the principal step is a reduction of the model to the escort system of ODE. It is instructive to deduce the escort system and the main results informally to clarify the main idea of the method in application to community models.

It is natural to suppose that the parameter a_3 is stochastically independent on the parameters a_1, a_2 . Let $M^1(\lambda_1, \lambda_2)$ be the mgf of the initial joint distribution of the parameters a_1, a_2 , and $M^2(\lambda_3)$ be the mgf of the initial distribution of the parameter a_3 .

Introduce the auxiliary variables $q_1(t)$, $q_2(t)$ as a solution to the Cauchy problem:

$$\begin{aligned} dq_1 / dt &= Y(t) \\ dq_2 / dt &= G(t) = X(t) E_t^1[a_2] \\ q_1(0) &= q_2(0) = 0 \end{aligned} \tag{7.9}$$

Then system (7.8) can be written formally as:

$$\begin{aligned} dx(t; a_1, a_2) / dt &= x(t; a_1, a_2)(a_1 - a_2 dq_1 / dt) \\ dy(t; a_3) / dt &= y(t; a_3)(dq_2 / dt - a_3) \end{aligned} \tag{7.10}$$

Its solution is:

$$\begin{aligned} x(t; a_1, a_2) &= x(0; a_1, a_2) \exp(a_1 t - a_2 q_1(t)) \\ y(t; a_3) &= y(0; a_3) \exp((q_2(t) - a_3 t)) \end{aligned} \tag{7.11}$$

Now we can express all values of interest with the help of the mgf-s of the initial distributions and the auxiliary variables:

$$\begin{aligned} X(t) &= X(0) \int_A \exp(a_1 t - a_2 q_1(t)) P^1(0; a_1, a_2) da_1 da_2 = X(0) M^1(t, -q_1(t)) \\ Y(t) &= Y(0) \int_A \exp((q_2(t) - a_3 t)) P^2(0; a_3) da_3 = Y(0) \exp(q_2(t)) M^2(-t) \end{aligned} \tag{7.12}$$

$$\begin{aligned} P^1(t; a_1, a_2) &= x(t; a_1, a_2) / X(t) = \exp(a_1 t - a_2 q_1(t)) / M^1(t, -q_1(t)) P^1(0; a_1, a_2) \\ P^2(t; a_3) &= y(t; a_3) / Y(t) = \exp(-a_3 t) / M^2(-t) P^2(0; a_3) \end{aligned} \tag{7.13}$$

$$E_t^1[a_2] = \left(\int_A a_2 \exp(a_1 t - a_2 q_1(t)) P^1(0; a_1, a_2) da_1 da_2 \right) / M^1(t, -q_1(t)), \text{ hence}$$

$$E_t^1[a_2] X(t) = X(0) \frac{\partial}{\partial \lambda_2} M^1(t, -q_1(t)) \tag{7.14}$$

Finally, we obtain a closed system of non-autonomous equations:

$$\begin{aligned} dq_1 / dt &= Y(0) \exp(q_2(t)) M^2(-t) \\ dq_2 / dt &= X(0) \frac{\partial}{\partial \lambda_2} M^1(t, -q_1) \end{aligned} \tag{7.15}$$

Now that we have a solution to the Cauchy problem for this system with zero initial values, we can get explicit formulas for total populations' sizes (7.12) and current distribution of the parameters (7.13), which completely solve the problem. In particular, the current mean values of the parameters:

$$\begin{aligned} E_t^1[a_1] &= \frac{\partial}{\partial \lambda_1} \ln(M^1(t, -q_1(t))) \\ E_t^1[a_2] &= \frac{\partial}{\partial \lambda_2} \ln(M^1(t, -q_1(t))) \\ E_t^2[a_3] &= \frac{\partial}{\partial \lambda_3} \ln(M^2(-t)) \end{aligned} \tag{7.16}$$

One can check that the obtained formulas coincide with the formulas, which follow from Theorem 5, MA. The current information discriminations for the inhomogeneous Volterra' model:

$$\begin{aligned}
 I[P^1_t : P^1_0] &= E_0[(a_1t - a_2q_1(t)) \exp(a_1t - a_2q_1(t)) / M^1_0(t, -q_1(t))] - \ln M^1_0(t, -q_1(t)) \\
 I[P^2_t : P^2_0] &= E^2_0[(a_1t - a_2q_1(t)) \exp(a_1t - a_2q_1(t)) / M^1_0(t, -q_1(t))] - \ln M^1_0(t, -q_1(t))
 \end{aligned}
 \tag{7.17}$$

The distributions (7.13) provide the minimum of information discriminations equaled to (7.17) over all distributions compatible with constraints (7.16).

Integrating the equations of system (7.8) over the parameters we obtain the system:

$$\begin{aligned}
 dX / dt &= X(E^1_t[a_1] - E^1_t[a_2]Y) \\
 dY / dt &= Y(E_t^1[a_2]X - E_t^2[a_3])
 \end{aligned}
 \tag{7.18}$$

These equations for total sizes of inhomogeneous populations have the same form as the initial Volterra’ system (7.5); the difference is that now the parameter values are not constants but vary over time according to formulas (7.16). The phase-parametric portrait of “homogeneous” Volterra’ model is well known (see, e.g., [14]). The dynamics of system (7.18) is determined by the parametric point $(E^1_t[a_1], E^1_t[a_2], E^2_t[a_3])$, which moves across the parametric portrait of model (7.5). This phenomenon, which may be referred to as “traveling across the parametric portrait of a homogeneous model” is a common feature of corresponding inhomogeneous models. It was well observed on the example of discrete-time models [13]. For Volterra-type model of two inhomogeneous populations with logistic reproduction rates and ratio-dependent predator functional response the phenomenon was studied in detail in [15].

7.3. Competition of Two Inhomogeneous Populations

The dynamics of two populations competing for a common resource can be described by the following logistic-like model (see, e.g., [14], ch.4):

$$\begin{aligned}
 dx / dt &= ax(1 - (x + \alpha y) / A) \\
 dy / dt &= by(1 - (y + \beta x) / B)
 \end{aligned}
 \tag{7.19}$$

where A, B are the capacities of the ecological niches for both populations, and α, β are the coefficients of interspecies competitions. A more general Allee-like model has a form:

$$\begin{aligned}
 dx / dt &= ax((x - L)(A - x) - \alpha y) \\
 dy / dt &= by((y - M)(B - y) - \beta x)
 \end{aligned}
 \tag{7.20}$$

where L, M are the lower threshold sizes for both populations.

Consider the inhomogeneous versions of these models, supposing that the reproduction rates a, b are distributed, and the competition is defined by the total sizes $X(t), Y(t)$ of the population.

Then instead of the logistic model we obtain the following model:

$$\begin{aligned}
 dx(t, a) / dt &= ax(t, a)(1 - (X(t) + \alpha Y(t)) / A) \\
 dy(t, b) / dt &= by(t, b)(1 - (Y(t) + \beta X(t)) / B)
 \end{aligned}
 \tag{7.21}$$

and correspondingly:

$$\begin{aligned} dx(t, a) / dt &= ax(t, a)((X(t) - L)(A - X(t)) - \alpha Y(t)) \\ dy(t, b) / dt &= by(t, b)((Y(t) - M)(B - Y(t)) - \beta X(t)) \end{aligned} \tag{7.22}$$

instead of the Allee-type model. Both systems have a form:

$$\begin{aligned} dx(t, a) / dt &= ax(t, a)(u(X(t)) - \alpha Y(t)) \\ dy(t, b) / dt &= by(t, b)(v(Y(t)) - \beta X(t)) \end{aligned} \tag{7.23}$$

where u, v are appropriate functions.

Let $P^1(0; a)$ and $P^2(0; b)$ be the initial distributions of the Malthusian rates a, b and $M^1(\lambda), M^2(\lambda)$ be corresponding mgf-s. In order to study model (7.23), we apply Theorem 5, MA and consider the 4-dimension escort system:

$$\begin{aligned} dq_1^1 / dt &= u(X(0)M^1(q_1^1 - \alpha q_2^1)) \\ dq_2^1 / dt &= Y(0)M^2(q_1^2 - \beta q_2^2) \\ dq_1^2 / dt &= v(Y(0)M^2(q_1^2 - \beta q_2^2)) \\ dq_2^2 / dt &= X(0)M^1(q_1^1 - \alpha q_2^1) \\ q_1^1(0) &= q_2^1(0) = q_1^2(0) = q_2^2(0) = 0 \end{aligned} \tag{7.24}$$

Suppose that Cauchy problem (7.24) has a unique global solution at $t \in [0, T], 0 < T \leq \infty$.

Define $K_t^1(a) = \exp(a(q_1^1(t) - \alpha q_2^1(t))), K_t^2(b) = \exp(b(q_1^2(t) - \beta q_2^2(t))$.

Then the solution to model (7.23) is:

$$\begin{aligned} x(t, a) &= x(0, a)K_t^1(a) = x(0, a)\exp(a(q_1^1(t) - \alpha q_2^1(t))) \\ y(t, b) &= y(0, b)K_t^2(b) = y(0, b)\exp(b(q_1^2(t) - \beta q_2^2(t))) \\ X(t) &= X(0)M^1(q_1^1(t) - \alpha q_2^1(t)) \\ Y(t) &= Y(0)M^2(q_1^2(t) - \beta q_2^2(t)) \end{aligned} \tag{7.25}$$

The mean values of the Malthusian rates at t moment:

$$\begin{aligned} E_t^1[a] &= \frac{\partial}{\partial \lambda} \ln(M^1(q_1^1(t) - \alpha q_2^1(t))) \\ E_t^2[b] &= \frac{\partial}{\partial \lambda} \ln(M^2(q_1^2(t) - \beta q_2^2(t))) \end{aligned} \tag{7.26}$$

The current pdf-s:

$$\begin{aligned} P^1(t; a) &= P^1(0; a)\exp(a(q_1^1(t) - \alpha q_2^1(t)) / M^1(q_1^1(t) - \alpha q_2^1(t)) \\ P^2(t; b) &= P^2(0; b)\exp(b(q_1^2(t) - \beta q_2^2(t)) / M^2(q_1^2(t) - \beta q_2^2(t)) \end{aligned} \tag{7.27}$$

The information discrimination:

$$\begin{aligned}
 I[P_t^1 : P_0^1] &= E_0^1[aQ^1(t) \exp(aQ^1(t)) / E_0^1[\exp(aQ^1(t))] - \ln E_0^1[\exp(aQ^1(t))] \\
 I[P_t^2 : P_0^2] &= E_0^2[bQ^2(t) \exp(bQ^2(t)) / E_0^2[\exp(bQ^2(t))] - \ln E_0^2[\exp(bQ^2(t))]
 \end{aligned}
 \tag{7.28}$$

where we denoted:

$$Q^1(t) = q_1^1(t) - \alpha q_2^1(t), Q^2(t) = q_1^2(t) - \beta q_2^2(t)
 \tag{7.29}$$

The distributions P_t^j (7.27) for $j = 1,2$ at each moment t provide minimum of $I[P_t^j : P_0^j]$ given by (7.28) over all probability distributions compatible with the constraints (7.26).

A simpler logistic model (7.21) can be reduced to a two-dimension escort system:

$$\begin{aligned}
 dq^1 / dt &= X(0)M^1(t - (q^1 + \alpha q^2) / A) \\
 dq^2 / dt &= Y(0)M^2(t - (\beta q^1 + q^2) / B) \\
 q^1(0) &= q^2(0) = 0
 \end{aligned}
 \tag{7.30}$$

The solution to model (7.21) is given by the following formulas:

$$\begin{aligned}
 x(t, a) &= x(0, a) \exp(a(t - (q^1(t) + \alpha q^2(t) / A))) \\
 y(t, b) &= y(0, b) \exp(b(t - (\beta q^1(t) + q^2(t) / B))) \\
 X(t) &= X(0)M^1(t - (q^1(t) + \alpha q^2(t) / A) \\
 Y(t) &= Y(0)M^2(t - (\beta q^1(t) + q^2(t) / B)
 \end{aligned}
 \tag{7.31}$$

The current pdf-s:

$$\begin{aligned}
 P^1(t; a) &= P^1(0; a) \exp(a(t - (q^1(t) + \alpha q^2(t) / A)) / M^1(t - (q^1(t) + \alpha q^2(t) / A) \\
 P^2(t; b) &= P^2(0; b) \exp(b(t - (\beta q^1(t) + q^2(t) / B)) / M^2(t - (\beta q^1(t) + q^2(t) / B)
 \end{aligned}
 \tag{7.32}$$

The mean values of the Malthusian rates at moment t :

$$\begin{aligned}
 E^1_t[a] &= \frac{\partial}{\partial \lambda} \ln(M^1(t - (q^1(t) + \alpha q^2(t) / A)) \\
 E^2_t[b] &= \frac{\partial}{\partial \lambda} \ln(M^2(t - (\beta q^1(t) + q^2(t) / B))
 \end{aligned}
 \tag{7.33}$$

The information discrimination:

$$\begin{aligned}
 I[P_t^1 : P_0^1] &= E_0^1[aQ^1(t) \exp(aQ^1(t)) / E_0^1[\exp(aQ^1(t))] - \ln E_0^1[\exp(aQ^1(t))] \\
 I[P_t^2 : P_0^2] &= E_0^2[bQ^2(t) \exp(bQ^2(t)) / E_0^2[\exp(bQ^2(t))] - \ln E_0^2[\exp(bQ^2(t))]
 \end{aligned}
 \tag{7.34}$$

where we denoted:

$$Q^1(t) = t - (q^1(t) + \alpha q^2(t) / A), Q^2(t) = t - (\beta q^1(t) + q^2(t) / B)
 \tag{7.35}$$

Distributions (7.32) provide the minimum (7.34) of discrimination information for model (7.21) at each time moment among all distributions subject the mean values (7.33).

8. Discussion and Conclusions

In this paper we develop a method of solving selection systems that describe replica dynamics based on the reduction theorem, and show that the solutions obey the Principle of minimum discrimination information, MinxEnt.

The selection system is a dynamical model of an inhomogeneous biological community or a population of individuals, each of which is characterized by a set of qualitative traits; the values of these traits determine the reproduction rate of the individual. The model takes into account the processes of replication and selection, but not mutation and immigration (at least, not explicitly). The main problem of interest is the dynamics of joint distribution of these traits in each population depending on the initial distribution, as well as on correlations between the traits and interconnections between the populations. The dynamics of a distribution of a selection system is governed by the replicator equation.

The MinxEnt (in other equivalent form, the Principle of maximum information entropy, or cross-entropy, MaxEnt), was successfully applied to various statistical, physical and biological problems as a method for inference of unknown distribution, subject to some given constraints. MinxEnt offers an efficient algorithm for construction of the minimum discrimination information probability distributions; the “MinxEnt distributions” are well known both in physics as the Boltzmann-Gibbs distributions, as well as in mathematics as the exponential family of distributions. The method can incorporate interactions between different traits and variables, *i.e.*, in the form of moments of their joined distribution and requires only testable information in the form of mean values of the traits.

There exists an “observer-dependent” view of the cross-entropy concepts (defended by Jaynes [5, 6] and, subsequently, by many other authors). Briefly, the authors claimed that entropy is a property of our description of a system rather than a property of the system itself. Then, if MaxEnt is fundamentally an algorithm of Bayesian statistical inference from partial information it is not clear why one can expect it to work as a description of nature. This problem has been discussed in many of papers over the last 50 years. An acceptable answer was formulated by Dewar [16]: “...MaxEnt predicts that behavior, which is selected *reproducibly* by nature under the imposed constraints”. In the recent paper [9] the author has suggested that MaxEnt “is not a physical principle but, rather, an inference algorithm that passively translates physical assumptions into macroscopic predictions”.

Nevertheless, the problem cannot be considered to be completely solved. Actually, an approach to obtain a satisfactory solution was pointed out by A. Einstein [17], who argued that the statistics of a system should follow from its dynamics and, in principle, could not be postulated a priori.

In this paper we show that, for a wide class of models for replica dynamics (the selection systems), the dynamical version of the Principle of minimal information production is neither inference algorithm no external principle but a mathematical assertion that can be derived from the system dynamics instead of being postulated.

We provide an algorithmic approach to find the solution to the replicator equation through reduction of the initial model (which may have infinite dimensionality) to a corresponding escort system of ordinary differential equations, which can be of significantly smaller dimensionality than the original system; the actual dimensionality is equal to the number of traits that define the reproduction rate. The solution to any replicator equation from the considered class is a time-dependent Boltzmann

distribution, whose parameters solve the escort system. With the solution to the replicator equation we can compute the current mean values of the traits at any instant. Then, treating these mean values as constraints, we can show that the “MinxEnt distribution” coincides with the solution of the replicator equation, which was obtained independently of the MinxEnt algorithm. Hence, the principle of the minimum of discrimination information can be considered as the variation principle that governs the selection system dynamics. Overall, both the reduction theorem and the Dynamical MinxEnt principle stem from the fact that the solution to the replicator equation belongs to the exponential family.

On the other hand, it is easy to show that any Boltzmann distribution with time-dependent parameters solves the corresponding replicator equation [8] and hence provides the minimum of discrimination information for the distribution of the associated selection system. It means that the replica dynamics is the “natural habitat” for the MinxEnt principle; within the framework of selection systems, we cannot choose whether or not to ascribe the property of minimization of the information discrimination to the selection system. It is an intrinsic property of any solution to the replicator equations that is fulfilled due to the system dynamics at any point of the system trajectory. More generally, the MinxEnt principle is an *internal property of the process of natural selection* at every moment of the system evolution.

We showed that the discrimination information is the Legendre transform of logarithm of the partition function; the auxiliary variables and the constraints are conjugated variables under this transformation. This assertion clarifies the meaning of the auxiliary variables and the role of the escort system. It also implies that the dynamics of selection system can be equally well described either in terms of the auxiliary variables or in terms of the constraints, which are the current mean values of the traits. Hence, the solution to the replicator equation is completely determined by the current mean values of the traits subject the important condition that the initial distribution is known.

Our approach is illustrated for inhomogeneous versions of some conceptual models of mathematical biology, namely, the logistic and Ricker’ models of populations and the Volterra’ models of communities. Formally, the inhomogeneous versions can be written in the same form as the initial models, with their current mean values substituted for the fixed parameters, as it was done in s.7 for the Volterra’ system. As a result, one obtains a complex non-linear integro-differential system, which hardly can be studied directly. The developed approach allows us to reduce this complex model to a system of two non-autonomous differential equations, which is specific for every initial distribution of the parameters. The same method works for other models and allows us to find the distribution of the systems at any time moment.

Looking beyond the formal solutions, we can now reveal a general optimization principle that governs the replica dynamics in many problems in mathematical biology, evolutionary game theory [3], the Eigen quasispecies theory [18], *etc.* As a result of the selection process, these systems evolve in such a way that the discrimination information is minimized at each time moment, given the mean values of the accounting traits. The values of the constraints and corresponding minimal value of discrimination information can be determined for each specific system by the developed method; we can conclude that for this type of models the MinxEnt principle follows from the system dynamics.

Acknowledgements

I appreciate four anonymous referees and the Guest Editor of the volume Alexander Gorban for their valuable comments and criticisms from which the paper has greatly benefited.

Mathematical Appendix

1. Information Discrimination and the Legendre Transform

Let us recall some definitions and statements (see e.g., [19], section 12 for rigorous definitions and theorems). Let $W(\mathbf{\alpha})$ be a convex function of r -dimensional vector $\mathbf{\alpha}$. The Legendre transform of the function $W(\mathbf{\alpha})$ is the function:

$$L(\mathbf{\beta}) = \sup \{ \mathbf{\alpha} : (\mathbf{\alpha}, \mathbf{\beta}) - W(\mathbf{\alpha}) \} \tag{A.1}$$

where $(\mathbf{\alpha}, \mathbf{\beta}) = \sum_{i=1}^r \alpha_i \beta_i$. The function $L(\mathbf{\beta})$ is again a convex function, and its Legendre transform is $W(\mathbf{\alpha})$. The functions $W(\mathbf{\alpha})$ and $L(\mathbf{\beta})$ are conjugate.

If the function $W(\mathbf{\alpha})$ is smooth, then the transform (A.1) coincides with the classical Legendre transform defined as follows. Let $\mathbf{\alpha} = \mathbf{\alpha}(\mathbf{\beta})$ be the solution of the equation:

$$\nabla W(\mathbf{\alpha}) = \mathbf{\beta}, \text{ i.e., } \partial W(\mathbf{\alpha}) / \partial \alpha_i = \beta_i. \text{ Then } L(\mathbf{\beta}) = (\mathbf{\alpha}(\mathbf{\beta}), \mathbf{\beta}) - W(\mathbf{\alpha}(\mathbf{\beta}))$$

The variables $\mathbf{\alpha}, \mathbf{\beta}$ are conjugate. Let

$$W(\boldsymbol{\lambda}) = \ln E_0 [\exp(\sum_{i=1}^n \lambda_i \varphi_i(\mathbf{a}))]$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$. It is a convex smooth function on $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$. Then:

$$\beta_j \equiv \partial W(\boldsymbol{\lambda}) / \partial \lambda_j = E_0 [\varphi_j \exp(\sum_{i=1}^n \lambda_i \varphi_i)] / E_0 [\exp(\sum_{i=1}^n \lambda_i \varphi_i)]$$

Now, let $\lambda_i = q_i(t)$. Then:

$$W(\mathbf{q}(t)) = \ln E_0 [\exp(\sum_{i=1}^n q_i(t) \varphi_i)] = \ln E_0 [K_t]$$

$$\beta_j = E_0 [\varphi_j K_t] / E_0 [K_t] = E_t [\varphi_j] = A_j(t) \text{ for all } j \leq n$$

Hence, taking $\beta_j = A_j(t)$, we obtain:

$$\sum_{i=1}^n \lambda_i(\boldsymbol{\beta}) \beta_i = \sum_{i=1}^n q_i(t) E_t [\varphi_i] = E_t [B(\mathbf{q}(t))]$$

$$W(\boldsymbol{\lambda}(\boldsymbol{\beta})) = E_0 [K_t] \text{ and finally:}$$

$$L(\boldsymbol{\beta}) = (\boldsymbol{\lambda}(\boldsymbol{\beta}), \boldsymbol{\beta}) - W(\boldsymbol{\lambda}(\boldsymbol{\beta})) = E_t[B(\mathbf{q}(t))] - \ln E_0[K_t] = I[P_t : P_0]$$

So, $I[P_t : P_0]$ at given constraints $E_t[\varphi_j] = A_j(t)$ solves the optimization problem:

$$I[P_t : P_0] = \sup \left\{ \boldsymbol{\alpha} : \sum_{i=1}^n \alpha_i A_i(t) - W(\boldsymbol{\alpha}) \right\} \tag{A.2}$$

and hence $q_i(t) = \alpha_i^*$ where α_i^* provide supremum in (A.2).

We have thus proven Theorem 4.

2. Inhomogeneous Models of Communities: the Main Theorems

Below we formulate the main theorems about model of inhomogeneous community (7.1)-(7.4). We omit the proofs because these theorems are similar to the corresponding theorems for inhomogeneous populations ([7, 8]) up to more complex technical details.

The model (7.1)-(7.4) can be reduced to an *escort system* of ordinary non-autonomous equations of dimensionality $r \times n$:

$$dq_i^j / dt = u_i^j(t, \mathbf{G}_i^*(t)), \quad q_i^j(0) = 0, \quad j = 1, \dots, r, \quad i = 1, \dots, n. \tag{A.3}$$

Here:

$$\mathbf{q}^j(t) = (q_1^j(t), \dots, q_n^j(t))$$

$$G_i^{*j}(t) = N^j(0)\Phi^j(g_i; \mathbf{q}^j(t))$$

in particular:

$$N^{*j}(t) = N^j(0)\Phi^j(1; \mathbf{q}^j(t))$$

and $\Phi^j(r; \boldsymbol{\lambda})$ is the generating functional (2.5) for the initial distribution $P^j(0, \mathbf{a})$.

Denote $K_t^j(\mathbf{a}) = \exp(\sum_{i=1}^n q_i^j(t)\varphi_i(\mathbf{a}))$.

Theorem 5. *Suppose that Cauchy problem (A.3) has a unique global solution at $t \in [0, T)$, $0 < T \leq \infty$. Then the functions:*

$$l^j(t, \mathbf{a}) = l^j(0, \mathbf{a})K_t^j(\mathbf{a})$$

$$G_i^j(t) = N^j(0)\Phi^j(g_i; \mathbf{q}^j(t))$$

$$N^j(t) = N^j(0)\Phi^j(1; \mathbf{q}^j(t))$$

satisfy system (7.1)- (7.3) at $t \in [0, T)$. The pdf:

$$P^j(t, \mathbf{a}) = P^j(0, \mathbf{a})K_t^j(\mathbf{a}) / E_0[K_t^j] = P^j(0, \mathbf{a})K_t^j(\mathbf{a}) / \Phi^j(1; \mathbf{q}^j(t)) \quad (\text{A.4})$$

solves the replicator equations (7.4).

It follows from Theorem 5, that:

$$dN^j / dt = N^j \left(\sum_{i=1}^n u_i^j(t, \mathbf{G}_i) E_t^j[\varphi_i] \right)$$

where $E_t^j[\varphi_i] = \Phi^j(\varphi_i; \mathbf{q}^j(t)) / \Phi^j(1; \mathbf{q}^j(t))$.

Theorem 6.

i) Let $P_t^j = P^j(t, \mathbf{a})$ be the solution (A.4) to replicator equation (7.4). Then at each moment t the distribution P_t^j provides minimum of $I[P_t^j : P_0^j]$ over all probability distributions compatible with the constraints $A_i^j(t) = E_t^j[\varphi_i]$, $i = 1, \dots, n$;

ii) The constraint values can be computed at each time moment by the formula:

$$A_i^j(t) = E_0^j[\varphi_i K_t] / E_0[K_t] = \Phi^j(\varphi_i; \mathbf{q}^j(t)) / \Phi^j(1; \mathbf{q}^j(t))$$

iii) The discrimination information $I[P_t^j : P_0^j]$ can be computed with the help of the following formulas:

$$I[P_t^j : P_0^j] = E_t[B^j(t, \mathbf{a})] - \ln(N^j(t) / N^j(0))$$

$$I[P_t^j : P_0^j] = E_0^j[B^j K_t^j] / E_0^j[K_t^j] - \ln E_0^j[K_t^j]$$

where $B^j(t, \mathbf{a}) = \sum_{i=1}^n q_i^j(t) \varphi_i(\mathbf{a})$.

References

- Schuster, P.; Sigmund, K. Replicator dynamics. *J. Theor. Biol.* **1983**, *100*, 533-538.
- Akin, E. Exponential Families and game dynamics. *Can. J. Math.* **1982**, *XXXIV*, 374-405.
- Hofbauer, J.; Sigmund, K. Evolutionary game dynamics. *Bull. Am. Math. Soc.* **2003**, *40*, 479-519.
- Kullback, S. *Information Theory and Statistics*; John Wiley: New York, NY, USA, 1959.
- Jaynes, E.T. Information theory and statistical mechanics I. *Phys. Rev.* **1957**, *106*, 620-630.
- Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
- Karev, G.P. On mathematical theory of selection: continuous time population dynamics. *J. Math. Biol.* **2010**, *60*, 107-129.
- Karev, G.P. Replicator equations and the principle of minimal production of information. *Bull. Math. Biol.* **2010**, DOI 10.1007/s11538-009-9484-9.

9. Dewar, R.C. Maximum entropy production as an inference algorithm that translates physical assumptions into macroscopic predictions: don't shoot the messenger. *Entropy* **2009**, *11*, 931-944.
10. Chentsov, N.N. *Statistical Decision Rules and Optimal Inference*; AMS: Providence, RI, USA, 1982.
11. Gorban, A.N. Selection theorem for systems with inheritance. *Math. Model. Nat. Phenom.* **2007**, *2*, 1-45.
12. Ackleh, A.S.; Marshall, D.F.; Heatherly, H.E.; Fitzpatrick, B.G. Survival of the fittest in a generalized logistic model. *Math. Model. Meth. Appl. Sci.* **1999**, *9*, 1379-1391.
13. Karev, G.P. Inhomogeneous maps and mathematical theory of selection. *JDEA* **2008**, *14*, 31-58.
14. Bazykin, A.D. *Nonlinear Dynamics of Interacting Populations*; World Scientific: Singapore, 1998.
15. Karev, G.; Novozhilov, A.; Koonin, E. Mathematical modeling of tumor therapy with oncolytic viruses: Effects of parametric heterogeneity on cell dynamics. *Biol. Dir.* **2006**, *3*, 1-30.
16. Dewar, R.C. Maximum entropy production and the fluctuation theorem. *J. Phys. A: Math. Gen.* **2005**, *38*, L371-L381.
17. Einstein, A. In *The Collected Papers of Albert Einstein, Volume 3: The Swiss Years: Writings, 1909-1911*, Klein, M.J., Kox, A.J., Renn, J., Schulmann, R., Eds.; Princeton University Press: Princeton, NJ, USA, 1993; pp. 286-312.
18. Eigen, M.; McCaskill, J.; Schuster, P. The molecular quasi-species. *Adv. Chem. Phys.* **1989**, *75*, 149-263.
19. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.

© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).