Doclady Mathematics. V 69, N 3, 2004, pp.355-358

CONSTRUCTION OF ENTROPY CHARACTERISTICS BASED ON LYAPUNOV ENERGY FUNCTIONS

Yu. A. Pykh

Research Center for Interdisciplinary Environmental Cooperation, Russian Academy of Sciences, nab. Kutuzova 14, St. Petersburg, 191187, Russia
e-mail: inenco@mail.neva.ru

Presented by Academician A.B. Kurzhanskii October 25, 2003

Received November 13, 2003

"We always have to lay special emphasis on mathematical analogies because the concentration on them can promote the development of science"

A.N.Kolmogorov [1]

In this paper, we construct a family of Lyapunov energy functions for generalized replicator equations and show that virtually all existing entropy characteristics and measures of distance between probability distributions belong to this family of functions.

Generalized replicator equations determine the evolution of probability distributions $p(t) = (p_1(t), ..., p_n(t)) \in \sigma$, where $\sigma = \{p \in \mathbb{R}^n : p_i \geq 0, i = 1, ..., n, \mathbf{e}^T p = 1\}$ - is the standard simplex in Euclidean n-space $\mathbb{R}^n$, $\mathbf{e}$ - is the unit vector.

These equations are written as [2]

$$\dot{p}_i = h(p) f_i(p_i)\left(\sum_{j=1}^{n} w_{ij} f_j(p_j) - \theta^{-1}(p) \sum_{j,k=1}^{n} w_{jk} f_j(p_j) f_k(p_k)\right) i = 1, ..., n \qquad (1)$$

Here, f_i are nonlinear response functions satisfying the conditions $f_i(0) = 0$, $\partial f_i / \partial p_i > 0$ $p_i > 0$, and $\partial f_i / \partial p_i \geqslant 0$ for $p_i = 0$; $W = (w_{ij})$ is the matrix of interactions; the function $h : \sigma \to (0, \infty]$ is determined by the particular problem under consideration; $\theta(p) = \langle \mathbf{e}, f(p) \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product; and $f(p) = (f_1(p_1), ..., f_n(p_n))$. Obviously, since $\langle \dot{p}(t), \mathbf{e} \rangle \equiv 0$ and $f_i(0) = 0$, the simplex σ and each of its faces are invariant sets of system (1).

System (1) has a very wide range of applications, from mathematical genetics to neural networks [2, 3]. Recently, it was shown [4] that system (1) can be obtained from Boltzmann-like equations. Thus, there are grounds for believing that system (1) determines the evolution of probability distributions for a fairly wide variety of processes.

To state the main theorem, we need some preliminary results. First, it is convenient to pass to the matrix form of representation. In this form, system (1) becomes

$$\dot{p} = h(p)\mathcal{D}(f)\left(Wf - \mathbf{e}\theta^{-1}(p)\langle f, Wf\rangle\right) \tag{2}$$

where $\mathcal{D}(f) = diag(f_1, ..., f_n)$.

If the matrix W is nondegenerate, then system (2) has at most one isolated equilibrium point in $\mathrm{Int}\sigma$, which we call nontrivial.

Statement 1[2]. System (2) has a unique nontrivial equilibrium point $\hat{p} \in \mathrm{Int}\sigma$ if and only if the vector $W\mathbf{e}^{-1}$ is either strictly positive or strictly negative.

The coordinates of this equilibrium point are determined from the system

$$\hat{f} \big/ \left\langle \hat{f}, \mathbf{e} \right\rangle = W^{-1}\mathbf{e} \big/ \left\langle \mathbf{e}, W^{-1}\mathbf{e} \right\rangle \tag{3}$$

where $\hat{f} = \left(f_1(\hat{p}_1), ..., f_n(\hat{p}_n)\right)$. Let us introduce the notation $W^{-1}\mathbf{e} = b$; then, for $\hat{p}_i$, we obtain the following system of $(n+1)$ equations:

$\hat{p}_i = f_i^{-1}(\lambda b_i) \quad i = 1, ..., n, \quad \sum_{i=1}^{n} \hat{f}_i^{-1}(\lambda b_i) = 1$, where λ - is an auxiliary variable similar to the Lagrange multiplier. Obviously, this system has a unique solution if all the b_i are of the same sign.

Definition [5]. A continuous function on the phase space of a dynamical system is called a Lyapunov energy function if it has continuous time derivative positive on the set of nonwandering points constrained to the system.

Theorem 1 [2]. If the matrix W is symmetric, then the function

$$E(p) = \left\langle f(p), Wf(p) \right\rangle \theta^{-2}(p) \tag{4}$$

is a Lyapunov energy function for system (2).

Corollary [2]. If system (2) has a nontrivial equilibrium point $\hat{p} \in \mathrm{Int}\sigma$, then it is totally stable in $\mathrm{Int}\sigma$ if and only if the matrix W has $(n-1)$ negative characteristic numbers.

An estimate for the characteristic numbers of real matrices smaller than the maximum number is given in [6].

Now, we can state the main result.

Theorem 2. If $W = W^T$ and system (2) has a nontrivial equilibrium point $\hat{p} \in \mathrm{Int}\sigma$ which is totally stable in $\mathrm{Int}\sigma$, then the function

$$H_\varepsilon(p) = \sum_{i=1}^{n} \int_{\varepsilon_i}^{p_i} \frac{\hat{f}_i dx}{f_i(x)} \tag{5}$$

where $0 \le \varepsilon_i \le 1 \quad i = 1,...,n$ are some constants, is a Lyapunov energy function for system (2).

Proof. For the derivative of the function H constrained to system (2), we have

$$\dot{H}_\varepsilon = h(p)\left(\left\langle \hat{f}W, f\right\rangle - \left\langle \hat{f}, \mathbf{e}\right\rangle \theta^{-1}(p)\langle f, Wf\rangle\right).$$

Taking into account relations (3), we obtain

$$\dot{H}_\varepsilon = h(p)\hat{\theta}\theta\left(\left\langle \mathbf{e}, W^{-1}\mathbf{e}\right\rangle^{-1} - \langle f, Wf\rangle \theta^{-2}\right),$$

and taking into account (4), we obtain

$$\dot{H}_\varepsilon = h(p)\hat{\theta}\theta\left(E(\hat{p}) - E(p)\right) \ge 0. \tag{6}$$

This inequality is obvious, because the function E increases and attains its maximum at the point $\hat{p}$, which proves the theorem.

It is also obvious that, if $F : \mathbb{R}^1_+ \to \mathbb{R}^1_+$ - is some monotonically increasing smooth function and C is a number, then the function

$$H_F = F\left(H_\varepsilon(p)\right) + C \tag{7}$$

is also a Lyapunov energy function for system (2).
To go further, we need one more result. Note that function (4) can be rewritten as $E(x) = \langle x, Wx\rangle$, where the auxiliary variables $x = (x_1,...,x_n)$ are defined by the formulas

$$x_i(p) = f_i(p_i)/\theta(p) \quad i = 1,...,n. \tag{8}$$

Obviously, $x \in \sigma$. The natural question arises: When does transformation (8) have a one-to-one inverse? The answer to this question is rather unexpected.

Theorem 3 [2]. Transformation (8) has a one-to-one inverse if and only if the functions f_i are power functions of the same degree, i.e.,

$$f_i(p_i) = d_i p_i^\alpha$$

where $d_i > 0,\ i = 1,...,n,\ \ \alpha > 0.$

In this case, the inverse transformation is determined by the relations

$$p_i = (d^{-\frac{1}{\alpha}} x_i^{\frac{1}{\alpha}})\vartheta^{-1}(x), \quad \vartheta(x) = \sum d_i^{-\frac{1}{\alpha}} x_i^{\frac{1}{\alpha}}.$$

Note that the existence and stability of a nontrivial equilibrium point of the system are completely determined by the properties of the interaction matrix W and do not depend on the particular form of the functions $f_i(p_i), i = 1, 2, ..., n.$

Below, we consider examples which show how two known entropy characteristics can be assigned a system of form (2) such that the given characteristic are Lyapunov energy functions for this system. We consider only two examples, although it is easy to show that the approach applies to virtually all of the existing entropy characteristics [7].

Example 1. The Shannon entropy. Suppose that the matrix W satisfies the conditions of Theorem 2 and is stochastic. Take the following functions $f_i\left(p_i\right)$:

$$f_i\left(p_i\right)=\left(1-\ln p_i\right)^{-1}$$

Obviously, $f_i\left(0\right)=0$, $f_i\left(1\right)=1$, and $\partial f_i/\partial p_i>0$ on the interval [0, 1].

The condition $\mathrm{We}=\mathrm{e}$ implies that system (2) has a nontrivial position equilibrium, which is the uniform distribution $\hat{p}=\left(n^{-1},...,n^{-1}\right)$. Substituting the functions f_i thus defined into (5), taking into account that the function $H_0\left(p\right)$ is determined up to an arbitrary constant, and setting $\varepsilon_i=0$, $i=1,2,...,n$, we obtain the Shannon entropy

$$H_0\left(p\right)=-\sum_{i=1}^{n}p_i\ln p_i.$$

Example 2. The Tsallis entropy [8].

Suppose that the matrix W satisfies the conditions of Example 1. As the functions $f_i\left(p_i\right)$, we take

$$f_i\left(p_i\right)=p_i^{1-q},\quad 0<q<1 \tag{9}$$

Next, we apply (7). Setting $F\left(p\right)=qn\left(1-q\right)^{-1}H_0\left(p\right)$, and $C=-\left(1-q\right)^{-1}$, performing simple calculations, and taking for $\varepsilon_i=0,\ i=1,2,...,n$ we obtain the Tsallis entropy

$$S_q\left(p\right)=\left(\sum_{i=1}^{n}p_i^q-1\right)\Big/\left(1-q\right) \tag{10}$$

Note that this conclusion is valid only if $0<q<1$. The lower bound is determined by the requirement that the improper integrals in (5) must converge and the upper bound follows from the equalities $f_i\left(0\right)=0$.

Note that, if we remove the condition $f_i\left(0\right)=0$, i.e., pass to a system which is not replicator in the usual sense, then the constraint $0<q<1$ can be removed too. Indeed, consider system (2) with functions (9) but without the constraints $0<q<1$. Obviously, if $q<1$, then $\lim f_i\left(p_i\right)\to+\infty$ as $p_i\to 0$. Thus, in this case, we can consider system (2) only in $\mathrm{Int}\sigma$. The system of equations obtained in this way is called an *extended* replicator system. As opposed to the case of generalized replicator system (2), simplex σ and its faces are not invariant sets for this system. From the physical point of view, it is natural to consider only *extended* replicator systems for which $\mathrm{Int}\sigma$ is a positive invariant set. For the derivative of Tsallis entropy (10), Eqs. (2) with $h(p) = (1-q)/q$

$$\dot{S}_q(p) = \theta(p)\left(1 - n\left\langle p^{1-q}, Wp^{1-q}\right\rangle \theta^{-2}\right) \geq 0 \qquad (11)$$

where $p^{1-q} = \left(p_1^{1-q}, ..., p_n^{1-q}\right)$, $p \in \mathrm{Int}\sigma$; thus, the Tsallis entropy is a Lyapunov energy function for the extended replicator system at $q \neq 1$. Certainly, the same result can be obtained by considering function (7) at small $\varepsilon_i > 0$.

Recall that function (10) was introduced by Tsallis for the purpose of constructing the thermodynamics of nonextensive systems, and the parameter q is a nonextensivity measure of the system [9]. The creation of such a thermodynamics is necessary because large systems with so called long-range interactions cannot be described by the Boltzmann thermodynamics. The approach developed in this paper makes it possible to associate each Tsallis entropy with some set of dynamical systems which explicitly depend on the character of the interactions between the objects under examination, which, in the author's opinion, may be essential for solving particular problems. Another important point is that, as Theorem 3 shows, the distribution $p(t)$ has a one-to-one companion distribution $f_i(p_i)$ of form (8) precisely for power functions $x(t)$ (8).

In the following example, we consider one of the best known measures of distance between probability distributions.

Example 3. The Kullback--Leibler relative entropy [10]. Suppose that the matrix W satisfies the conditions of Theorem 2. We set $f_i(p_i) = p_i$ and $\varepsilon_i = \hat{p}_i$ for $i = 1, 2, ..., n$. Then (5) gives the Kullback--Leibler relative entropy

$$H_{\hat{p}}(p) = \sum_{i=1}^{n} \hat{p}_i \ln\left(p_i / \hat{p}_i\right)$$

For the first time, this function was used to analyze the stability of equilibrium points of replicator systems in [11]. Note that, since the functions f_i are linear in this case, we deal with the classical system of replicator equations. A detailed analysis of such equations was made in [12]; in particular, it was shown in [12] that, for such systems, we can remove the constraint $W = W^T$ and obtain results similar to those obtained above for matrices of the type $W = D_1 B D_2$, where D_1 and D_2 are diagonal matrices, $D_1 D_2 > 0$, and $B = B^T$, i.e., for the class of so-called diagonally symmetrizable matrices. A necessary and sufficient condition for a real matrix W to be diagonally symmetrizable was obtained in [13]. It is so far unclear whether the constraints $W = W^T$ can be removed in the case of nonlinear functions f_i .

Thus, the approach developed in this paper makes it possible to obtain not only various entropy characteristics but also so-called measures of distance between probability distributions [14].

Note that the function

$$H_F\left(p,\hat{p}\right)=F\left(\sum_{i=1}^{n}\int_{\hat{p}_i}^{p_i}\frac{\hat{f}_i dx}{f_i\left(x_i\right)}\right)+C$$

is a generalization of the f-divergence for distributions $\hat{p}$ and p suggested by Csiszar in [15]; if the dynamics of the process under examination is determined by system (2) and the conditions of Theorem 2 hold, then this function is a Lyapunov energy function.

It is seen from the examples given above that, if we consider entropy characteristics, then we should set $\varepsilon_i=0$ in the expression for $H_\varepsilon\left(p\right)$, and if we are interested in "distances" between distributions, then we should take $\varepsilon_i=\hat{p}_i$. We also emphasize that Eq. (6) establishes a relation between the derivative of the function $H_\varepsilon\left(p\right)$, which can be interpreted as a generalized entropy, and the function $E\left(p\right)$, which is often (in particular, in works on the theory of neural networks) considered as an analog of the energy.

In conclusion, we mention that all results stated above were obtained by formally analyzing systems of generalized replicator equations, which arise in very diverse fields of natural sciences and, therefore, can serve as a basis for finding analogies between these domains of natural sciences.

REFERENCES

1. A. N. Kolmogorov, Izv. Akad. Nauk SSSR, Ser. Mat. Vol.21, pp. 66-99 (1957).
2. Yu. A. Pykh, in Proceedings of International Conference "Physics and Control," St. Petersburg, Russia, 2003 (IFEE Publ., 2003), Vol. 1, pp. 271-276.
3. J. Hofbauer and K. Sigmund, Bull. Am. Math. Soc. Vol.40 N4, pp.479-519 (2003).
4. D. Helbing, Theory Decision Vol.40, N2, pp. 149-179 (1996).
5. K. Meyer, Am. J. Math. Vol.90, N 4, pp. 1031-1040 (1968).
6. Yu. A. Pykh, Dokl. Akad. Nauk SSSR, Vol. 211, N 6, pp. 1299-1301 (1973).
7. M. D. Esteban and D. Morales, Kybernetika, Vol.31, N4, pp. 337-346 (1995).
8. C. Tsallis, J. Stat. Phys. Vol.52, N3, pp. 479-487 (1988).
9. C. Tsallis, Chaos Solitons Fractals, Vol.13, N3, pp.371-391 (2002).
10. S. Kullback and R. Leiber, Ann. Math. Stat. Vol.22, N1, pp. 79-87 (1951).
11. Yu. A. Pykh, Dokl. Akad. Nauk SSSR Vol.203,N4, pp. 766-768 (1972).
12. Yu. A. Pykh, Equilibrium and Stability in Models of Population Dynamics. (Nauka, Moscow, 1983, 184p.) [in Russian].

13. Yu. A. Pykh, in Proceeding of the 5 IFAC Symposium "Nonlinear Control Systems," St. Petersburg, Russia (IFAC Publ., 2001), pp. 1655-1660.
14. I. J. Taneja, Adv. Imaging Electron. Phys. Vol.91, N1, pp. 37-135 (1995).
15. I. Csiszar, Publ. Math. Inst. Hung. Acad. Sci. Vol.8, N1, pp. 85-107 (1963).