



Research

Cite this article: Hernando A, Hernando R, Plastino A, Plastino AR. 2013 The workings of the maximum entropy principle in collective human behaviour. *J R Soc Interface* 10: 20120758.
<http://dx.doi.org/10.1098/rsif.2012.0758>

Received: 19 September 2012

Accepted: 24 October 2012

Subject Areas:

mathematical physics

Author for correspondence:

A. Hernando

e-mail: alberto.hernando@irsamc.ups-tlse.fr

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2012.0758> or via <http://rsif.royalsocietypublishing.org>.

The workings of the maximum entropy principle in collective human behaviour

A. Hernando¹, R. Hernando², A. Plastino^{3,4,6} and A. R. Plastino^{3,5}

¹Laboratoire Collisions, Agrégats, Réactivité, IRSAMC, Université Paul Sabatier, 118 Route de Narbonne 31062, Toulouse Cedex 09, France

²Social Thermodynamics Applied Research (SThAR), Ambrosio Vallejo 16, 28039 Madrid, Spain

³Instituto Carlos I de Física Teórica y Computacional and Departamento de Física Atómica, Molecular y Nuclear Universidad de Granada, Granada, Spain

⁴National University La Plata, Physics Institute (IFLP-CCT-CONICET) C.C. 727, 1900 La Plata, Argentina

⁵CREG-National University La Plata-CONICET C.C. 727, 1900 La Plata, Argentina

⁶IFISC (CSIC-UIB), Campus Universitat Illes Balears, 07122 Palma de Mallorca, Spain

We present an exhaustive study of the rank-distribution of city-population and population-dynamics of the 50 Spanish provinces (more than 8000 municipalities) in a time-window of 15 years (1996–2010). We exhibit compelling evidence regarding how well the MaxEnt principle describes the equilibrium distributions. We show that the microscopic dynamics that governs population growth is the deciding factor that originates the observed macroscopic distributions. The connection between microscopic dynamics and macroscopic distributions is unravelled via MaxEnt.

1. Introduction

1.1. Patterns in social systems

Orderliness, reflected in either scaling properties [1] or power laws [2–6], is encountered in different frameworks involving social groups. A salient example is Zipf's Law [7,8], a power law with exponent -2 for the density distribution function that is observed in describing urban agglomerations [9–11] and firm sizes all over the world [12]. This kind of 'order' has received immense attention in the literature. Pertinent regularities have been found in other scenarios as well, ranging from scientific collaboration networks [13] and total number of cites by physics journals [14] to Internet traffic [15,16] or Linux package links [17]. Another special regularity has also been found in the density distribution function of the number of votes in Brazilian elections [18]: a power law whose exponent is -1 (this kind of distribution is the origin of the so-called First-Digit Law, or Benford's Law, as shown in Weisstein [19]). This behaviour is also found in Hernando *et al.* [20] for the city size-distribution of the province of Huelva (Spain), and in the distribution of the number of votes in the 2008 Spanish General Elections. Undoubtedly, some aspects of human behaviour reflect a kind of 'universality'.

Common to all these disparate systems is the lack of a characteristic size, length or frequency for the observable of interest, which makes it scale-invariant. In our previous work [20], we have introduced an information-theoretic technique based upon the minimization of Fisher's information (MFI) measure [21] that allows for the formulation of a 'thermodynamics' for scale-invariant systems. The methodology establishes an analogy between such systems and physical gases that, in turn, shows that the two special power laws mentioned in the preceding paragraph lead to a set of relationships formally identical to those pertaining to the equilibrium states of a scale-invariant non-interacting system, the scale-free ideal gas (SFIG). The difference between the two distributions is thereby attributed to different boundary conditions on the SFIG.

Many social systems, however, are not easily included in the two previous descriptions. Inspired by opinion dynamics models [18,22,23], we described in a previous work [24] a numerical process that reproduces the shapes of the empirical city-population distributions. The model is based on a competitive

cluster growth process inside a scale-free ideal network (a scale-free network in which degree distribution is described as an SFIG). The finite size of the network introduces competition, and thus correlations in the cluster sizes. The larger the competitiveness, the larger the deviation from the SFIG distribution. The equivalence with the workings of physical gases was shown to be compelling indeed. However, some aspects of the concomitant problems defied full understanding, because an analytical prescription for the classification of the size-distribution of social groups was missing.

1.2. Entropy measure of information

In order to do look for such analytical understanding, we will appeal here to a variational approach centred on information theory. On the basis of ideas by Hernando *et al.* [25] we deal with the well-established MaxEnt technique [26–29] by including, within the associated Lagrangian, information regarding equations of motion, i.e. ‘dynamical’ information of a kind that goes beyond the customary one, based upon expectation values.

We have demonstrated that proportional growth and hyper-exponential growth (also known as q -growth where q is a real value that parametrizes the growth dynamics) can both be described in these terms, accurately predicting the equilibrium size-distribution for such systems [25]. This is tantamount to giving a dynamical interpretation to the ‘information cost’ introduced in Baek *et al.* [30]. Our approach relates the microscopic dynamics of the system elements with the observed macroscopic distribution. We remark that the introduction of dynamical information results in a net decrease of the entropy. Take for example geometric Brownian motion. In the case of a pure diffusive process, the final state covers all space with equiprobability, while in the geometric Brownian process the final state does not *uniformly* cover all space. Of course, a lower value of the entropy ensues then. See in this vein Corominas-Murtra & Solé [8], where Zipf’s Law has been shown to be the asymptotic solution for a wide range of entropy losses. The important concept of entropy loss is lucidly described in Harremoës & Topsoe [31,32]. See also Gabaix [11] and Zanette & Manrubia [33] for interesting related work.

1.3. Aim of this work

Our main present purpose is to apply what we learned in Hernando *et al.* [25] to empirical social systems, thus demonstrating the applicability of MaxEnt to collective human behaviour, and the potential use of an explicit *social thermodynamics*. The applicability of this theoretical procedure will allow us to reduce the large number of microscopic degrees of freedom to a few macroscopical parameters. In this work we test two different tableaus for the migration dynamics underlying the empirical city population:

- (i) Proportional growth with a constraint in the total population. The city populations evolve as geometrical Brownian walkers but total population remains constant.
- (ii) q -growth with total-population constraint plus proportional drift. This more sophisticated evolution uses q -metric Brownian walkers [25] to reproduce migration patterns and adds some drift of proportional nature to simulate the growth of the total population.

We fit each microscopic dynamical model to the empirical data on city-populations, and compare the MaxEnt prediction

for macroscopic size-distribution with the actual one. *If compatibility is encountered between the microscopic and macroscopic measure, our theoretical description should be correct.*

This work is organized as follows. In §2 we describe the empirical dataset used here. In §3, we revisit the theoretical MaxEnt approach to be followed for tableau (i) and (ii). In §4, we present the results, first for the proportional growth approximation and then for the q -growth with drift. We analyse the results for a numerical simulation with random walkers as the control case, and apply our methodology to the empirical data for the Spanish provinces. The ensuing results clearly show the usefulness of the theoretical method. In §5, we discuss some noteworthy features of the results and their possible uses in other contexts. A summary is given in §6 and, finally, some technical aspects are the subject of appendix A.

2. Empirical dataset

The raw data are obtained from the Spanish state institute INE [34] and cover the years in the 1996–2010 (with the exception of 1997) period. It encompasses 8116 municipalities (the smallest Spanish administrative unit) distributed within 50 provinces (the building blocks of the autonomous communities). The autonomous cities of Ceuta and Melilla are not included. We use provinces and municipalities as the closest representatives of the ideal of a closed system’s fundamental elements. However, some arbitrariness remains in the data, because (i) there are many municipalities that actually possess more than one population-centre, (ii) there are provinces that contain more than a single socioeconomic cluster, and (iii) there are economic regions that extend beyond the border of a province. Those facts introduce systematic errors into the data, but we retain enough accuracy for the purposes of this work.

The data cover a total Spanish population of 39 106 917 inhabitants in 1996, reaching 47 254 510 in 2010. For this last year, the largest municipality (Madrid) covers 3 273 049 people and the smallest one just five persons (Illán de Vacas, Toledo). The total population of each province ranges between 6 458 684 inhabitants (Madrid) and 95 258 (Soria), and the number of municipalities, between 371 (Burgos) and 34 (Las Palmas). Figure 1 shows the spatial distribution of the Spanish municipalities.

The number of municipalities in each province is not large enough to build up an accurate macroscopic size-density distribution. Accordingly, we have systematically worked with rank-distributions (RDs) instead. RDs are usually built from the vector $\mathbf{x} = \{x_i\}_{i=1}^n$, where x_i is the population of the i th municipality and n is the total number of municipalities in that province. We assign rank numbers ranging from 1 to n , from the largest ($\max(\mathbf{x}) \rightarrow r = 1$) to the lowest in population ($\min(\mathbf{x}) \rightarrow r = n$). However, in order to compare with theoretical, continuous RDs, we have found it to be more accurate to assign ‘middle-point’ values from $r = 0.5$ to $r = n - 0.5$, instead. We show some examples of RD for different provinces in figures 2 and 4 (the RDs for all the 50 provinces can be found in the electronic supplementary material). For our microscopic dynamical analysis, we generate, for each province and for each time-period, the pair (u_i, \dot{u}_i) , with the logarithm of the population u_i correlated with the relative population increment \dot{u}_i . They are computed as

$$u_i = \frac{\log[x_i(t_2)x_i(t_1)]}{2} \quad \text{and} \quad \dot{u}_i = \frac{\log[x_i(t_2)/x_i(t_1)]}{t_2 - t_1}, \quad (2.1)$$

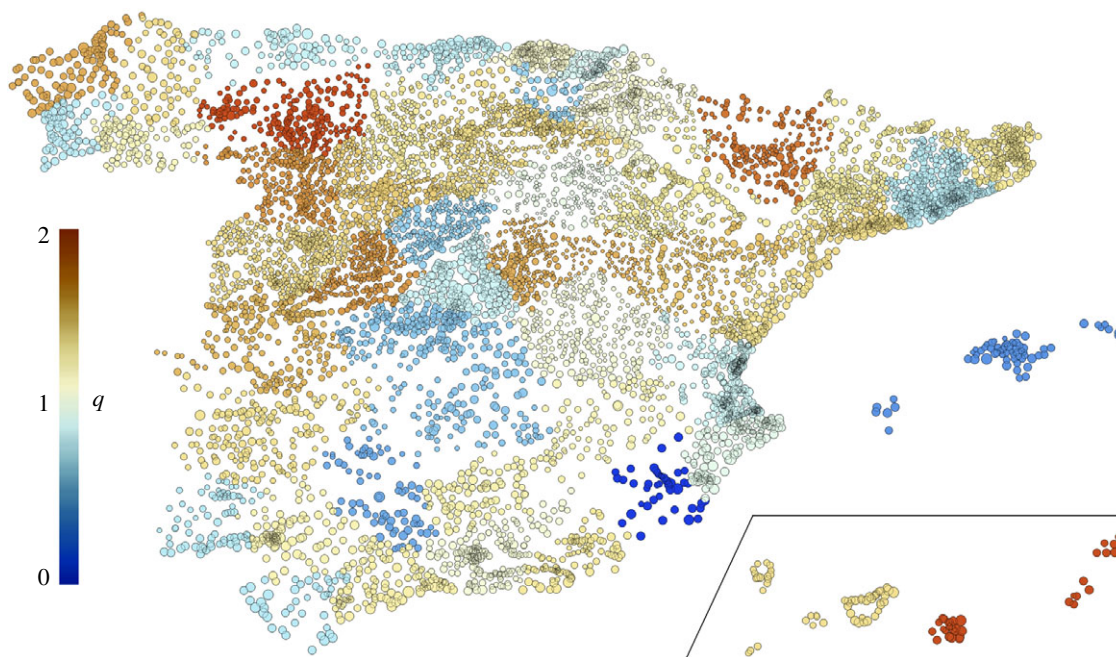


Figure 1. Spatial map of the municipalities of Spain. The size is proportional to the logarithm of the population, and colour indicates the q value obtained from the rank-distributions of that province when applicable, or from the dynamics instead (see text). (Online version in colour.)

where t_1 and t_2 are consecutive years for which data are available. We show in the insets of figures 2 and 4 the corresponding dynamics for each province. To unravel the relationship between these dynamic pictures and city-population RDs is the main purpose of this work.

3. Theoretical approach

3.1. Basics

Following the methodology described in Hernando *et al.* [25], let us define our relevant quantities. They are

- (i) N , the total population of a province,
- (ii) n , the total number of municipalities into which the population is apportioned,
- (iii) $x_i(t)$, the population of the i th municipality at time t ,
- (iv) $\dot{x}_i(t)$, the time-derivative of $x_i(t)$,
- (v) *a priori* knowledge of the dynamics at hand, i.e. $\dot{x}_i(t) = g[x_i(t)]$, where $g(x)$ describes the functional dependence of the dynamics in the population x ,
- (vi) x_0 , the minimum possible population (at least $x_0 = 1$), and
- (vii) $p_X(x)$, the number of cities with exactly an x -population.

Considering the continuous limit of the distribution $p_X(x)$, the conservation of both n and N guarantees

$$\int_{x_0}^{\infty} dx p_X(x) = n \quad \text{and} \quad \int_{x_0}^{\infty} dx p_X(x)x = N. \quad (3.1)$$

The variables $x_i(t)$ and $\dot{x}_i(t)$ describe the explicit microscopic state of the system, and they are those involved in the dynamical equation. On the other hand, quantities as N , n , x_0 or the distribution $p_X(x)$ can be obtained from macroscopic observations. The theoretical approach that follows is our attempt to describe macroscopically the state of the system univocally using the latter macro-observables. To

this aim, we introduce in the MaxEnt principle the dynamical information of the equation of motion. We do that in the fashion of Hernando *et al.* [25], i.e. turning on MaxEnt and conjecturing, à la Jaynes, that in dynamical equilibrium the appropriate distribution $p_X(x)dx$ is the one that maximizes Shannon's entropy S subject to the above constraints. One thereby concocts the variational problem

$$\delta[S - \mu n - \lambda N] = 0, \quad (3.2)$$

with μ and λ the corresponding Lagrange multipliers. The functional form of the dynamics $g(x)$ affects the way Shannon's entropy is measured [25]. The equilibrium city size-distributions $p_X(x)$ are determined univocally by $g(x)$, x_0 , N and n , as will be seen next for two different guesses for $g(x)$: proportional growth ($g(x) \propto x$) and q -growth ($g(x) \propto x^q$).

3.2. MaxEnt for proportional growth dynamics

Hernando *et al.* [24] showed that city population expansion can be modelled by means of cluster growth in complex networks. Such a process (diffusion) in networks generally starts (i) using a node as a seed, (ii) its first neighbours being added to the cluster in the first iteration, (iii) the neighbours of those neighbours afterwards and so on. Mathematically, if Δt is the interval of time for each iteration and $x(t)$ is the population of the cluster at the time t , we can write

$$x_i(t + \Delta t) = x_i(t) + \Delta t \sum_{j=1}^{x_i(t)} k_{ij}(t), \quad (3.3)$$

where $k_{ij}(t)$ is the number of neighbours of the j th node accepted to the cluster per unit time at time t . Because it is our intention to pass to the continuum, we appeal to the central limit theorem so as to generalize this equation to

$$x_i(t + \Delta t) \simeq x_i(t) + \Delta t (k_i(t)x_i(t) \pm \sigma_k \sqrt{x_i(t)}), \quad (3.4)$$

where $k_i(t)$ is a mean value of accepted neighbours per unit time at time t and σ_k its standard deviation. For large enough values of

the population the last term—related to *finite size effects*—can be neglected; so we write for the continuous limit

$$\dot{x}_i(t) = k_i(t)x_i(t), \quad (3.5)$$

where the dot represents the time derivative. This equation describes the so-called proportional growth, whose generalization to a random process leads to the so-called geometrical Brownian motion. A system governed by these dynamics exhibits *scale-invariance* since a scale transformation in the fashion $x'(t) = cx(t)$, where c is an arbitrary constant, leaves invariant the dynamical equation. Equation (3.5) is linearized by introduction of the new variable $u_i = \log(x_i/x_0)$, which leads to $\dot{u}_i(t) = k_i(t)$. Considering that k is a Wiener coefficient (following a normal distribution with covariance $\langle k_i(t)k_j(t') \rangle = \langle k_i(t)^2 \rangle \delta(t-t')\delta_{ij}$), the Shannon measure is expressed in terms of u as

$$S = \int_0^\infty du p(u) \log[p(u)]. \quad (3.6)$$

where $p(u)du = p_X(x)dx$. With our two conservation rules (equation (3.1)) written in u -terms, the MaxEnt variational problem becomes

$$\delta \left[S - \mu \int_0^\infty du p(u) - \Lambda \int_0^\infty du p(u) e^u \right] = 0, \quad (3.7)$$

where we have used the definition $\Lambda = x_0\lambda$. The solution is the equilibrium density $p(u) du = nZ^{-1} \exp(-\Lambda e^u) du$, that one now recasts in terms of the observable x getting

$$p_X(x) dx = \frac{n e^{-\Lambda x/x_0}}{Z} dx. \quad (3.8)$$

From the conservation rules, we can easily obtain for the pertinent constraints the values

$$\text{and } \left. \begin{aligned} Z &= \Gamma(0, \Lambda) \\ \frac{e^{-\Lambda}}{\Lambda \Gamma(0, \Lambda)} &= \frac{N}{nx_0} \end{aligned} \right\} \quad (3.9)$$

with $\Gamma(a, z)$ the so-called incomplete Gamma function. In this way, the $2n$ microscopic degrees of freedom (each x_i and \dot{x}_i pair) are reduced to four parameters (n, N, x_0 and Λ) related between them via this last equation. The shape of the macroscopic distribution is then determined by a single value N/nx_0 .

The concomitant cumulative function $P_X(x)$ reads

$$P_X(x) = n \left[1 - \frac{\Gamma(0, \Lambda x/x_0)}{\Gamma(0, \Lambda)} \right], \quad (3.10)$$

and the associated RD—obtained from the inversion of the cumulative one—becomes

$$x = \frac{x_0}{\Lambda} \Gamma^{-1} \left[0, \frac{\Gamma(0, \Lambda)r}{n} \right], \quad (3.11)$$

where r is the (continuous) rank from 0 to n , and $\Gamma^{-1}(z)$ denotes the inverse function of Γ : $\Gamma(\Gamma^{-1}(z)) = z$. We derive from here what we call the *Gamma Scaling Law*

$$\text{and } \left. \begin{aligned} x' &= \frac{x\Lambda}{x_0} \\ r' &= \frac{r\Gamma(0, \Lambda)}{n} \end{aligned} \right\} \quad (3.12)$$

obtaining a ‘scaled’ RD

$$x' = \Gamma^{-1}(0, r'), \quad (3.13)$$

which no longer depends on N, n or x_0 .

3.3. Beyond proportional growth generalizing equation (3.5)

The description based on equation (3.5) has limitations that we discuss in §4.1. A different possibility consist in appealing to the more general expression

$$\dot{x}_i(t) = k_i(t)[x_i(t)]^q, \quad (3.14)$$

where q is a dimensionless parameter. Using the definitions of the q -logarithm and q -exponential of Tsallis’ statistics [35–37]

$$\left. \begin{aligned} \log_q(x) &= \frac{x^{1-q} - 1}{1-q}; \quad (x > 0) \\ \text{and } e_q(v) &= [1 + (1-q)v]_+^{1/(1-q)}, \end{aligned} \right\} \quad (3.15)$$

linearized by means of the new variable $v_i = \log_q(x_i/x_0)$. The subindex $+$ takes care of the so-called Tsallis cut off [38]: e_q vanishes if the bracket becomes negative. We obtain the dynamical relationship $\dot{v}(t) = k(t)$, and so the Shannon entropy is measured as

$$S = \int_0^{v_m} dv p(v) \log[p(v)], \quad (3.16)$$

where $v_m = 1/(q-1)$ if $q > 1$ or infinite otherwise. The concomitant MaxEnt problem is now of the form

$$\delta \left[S - \mu \int_0^{v_m} dv p(v) - \Lambda \int_0^{v_m} dv p(v) e_q(v) \right] = 0, \quad (3.17)$$

whose solution is $p(v) dv = nZ^{-1} \exp[-\Lambda e_q(v)] dv$. Changing back to the observable x , we find

$$p_X(x) dx = \frac{n e^{-\Lambda x/x_0}}{Z} dx. \quad (3.18)$$

The constraints derive from the conservation rules in the usual manner as

$$\text{and } \left. \begin{aligned} Z &= \left(\frac{\Lambda}{x_0} \right)^{q-1} \Gamma(1-q, \Lambda) \\ \frac{\Gamma(2-q, \Lambda)}{\Lambda \Gamma(1-q, \Lambda)} &= \frac{N}{nx_0}. \end{aligned} \right\} \quad (3.19)$$

The current cumulative function turns out to be

$$P_X(x) = n \left[1 - \frac{\Gamma(1-q, \Lambda x/x_0)}{\Gamma(1-q, \Lambda)} \right], \quad (3.20)$$

and for the associated RD one has

$$x = \frac{x_0}{\Lambda} \Gamma^{-1} \left[1-q, \frac{\Gamma(1-q, \Lambda)r}{n} \right]. \quad (3.21)$$

We now consider a q -equilibrium system subject to proportional drift that may account for the ‘natural’ population-growth or fluctuations of proportional nature. One is here including a kind of noise that affects the underlying dynamical equation via

$$\dot{x} = k_1 x(t) + k_q x(t)^q, \quad (3.22)$$

or, in terms of $u = \log(x)$,

$$\dot{u} = k_1 + k_q e^{(q-1)u}, \quad (3.23)$$

where k_1 characterizes proportional drift and k_q the hyper-exponential growth. Considering both k_1 and k_q as independent Wiener coefficients, the system departs equilibrium via proportional diffusion. As seen in Hernando *et al.* [25], the kernel function for that kind of diffusion is a lognormal

distribution. We face the convolution

$$p_X(x|\Lambda, x_0, q, \sigma) dx = \frac{n dx}{(\Lambda/x_0)^{q-1} \Gamma(1-q, \Lambda)} \times \int_{x_0}^{\infty} dx' \frac{e^{-\Lambda x'/x_0} e^{-\log^2(x'/x)/2\sigma^2}}{x'^q x \sqrt{2\pi\sigma}}, \quad (3.24)$$

where σ is the drift-scale parameter. The mean value reads

$$\frac{\Gamma(2-q, \Lambda)}{\Lambda \Gamma(1-q, \Lambda)} e^{\sigma^2/2} = \frac{N}{nx_0}. \quad (3.25)$$

In this model, the $2n$ microscopic degrees of freedom are then reduced to six parameters (n , N , x_0 , Λ , q and σ) related between them via this last equation. The shape of the macroscopic distribution is then determined by the value of N/nx_0 , q and σ .

Note that we have derived our distributions using MaxEnt (with the standard Shannon entropy, but with modified variables). Many other authors use MaxEnt in conjunction with different entropic forms for which one of the four Khinchin's axioms is modified (the axiom of extensivity) [38,39], but keeping the original variables, not changing them as we do.

3.4. Numerical simulation with random walkers

We have performed numerical simulations with random walkers to test the theoretical results of our connection between the microscopic dynamics and the macroscopic state via MaxEnt. This is achieved following the algorithm described in Hernando *et al.* [25] for the dynamics. We first find the equilibrium state attained by n q -metric walkers following equation (3.14) with a given value of q , a initial total population N , and a minimum allowed population x_0 (or the equivalent value of u_0 if working with the linearized variable u). Without this lower bound, the dynamics may reach a peculiar stable state in which the population tends to zero owing to the eventual disappearance of the settlements. The effect of this kind of constraint on the final size-distribution is also studied in Blank & Solomon [2], where the above-referred peculiarity is avoided by recourse to introducing an exponential growth that takes place if one forces the average of k to be greater than zero. In our case, growth is introduced as a proportional drift, via equation (3.23), with a positive-valued mean for k . Accordingly, the system evolves in the following fashion:

- (i) walkers 'move' following the q -dynamics, and populations are corrected as described in Hernando *et al.* [25] to reproduce the total population,
- (ii) we make the walkers to evolve by adding a proportional noise,
- (iii) the ensuing drift generates population-growth and the value of N obtained afterwards is updated accordingly, and
- (iv) we keep going on during a specified number m of iterations.

The walkers' final macroscopic size-distribution and their dynamics are analysed in the same way as we did for empirical data in §2. This final distribution is used as a control case. We will show in §4 that our approach is able to macroscopically describe the effects of (i) the lower bound and (ii) the

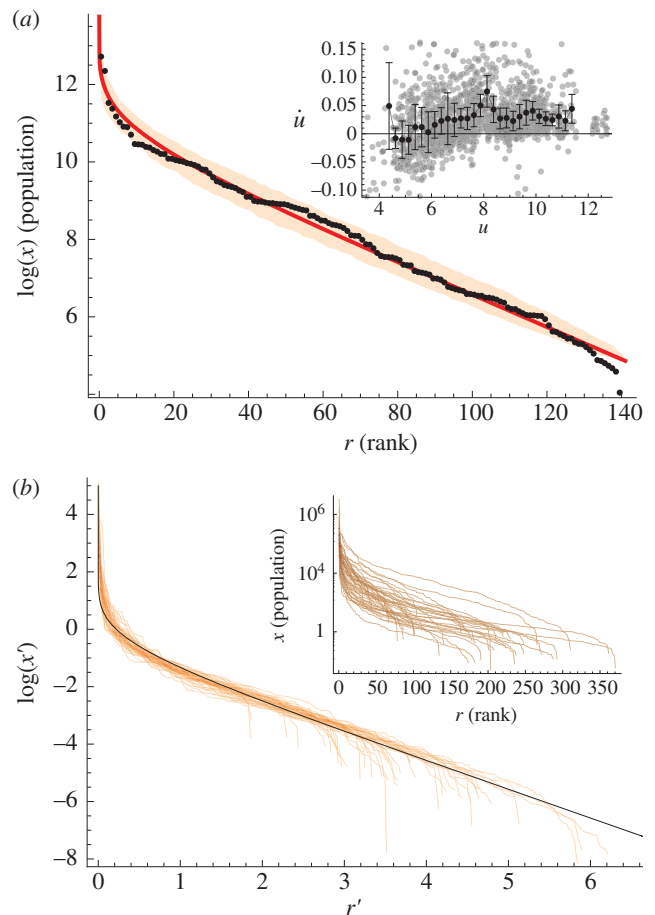


Figure 2. (a) Macroscopic rank-distribution of the province of Alicante (year 2010), compared with the corresponding MaxEnt prediction for proportional growth. The shadowed area represents the 90% CI for $n = 141$. Inset: relative change versus natural logarithm of the population for this province for the last 15 years, where each grey dot represents one municipality for each annual period. Black dots denote the mean value in each interval $\Delta u = 0.25$ and the error bars the standard deviation at that interval. (b) Gamma-scaled rank-distribution of all 50 provinces. Inset: raw data (2010). (Online version in colour.)

positive mean growth (microscopic dynamics). Also, the preceding discussion is revisited in §5.2.

4. Results

4.1. Proportional growth analysis

We use as paradigmatic example of proportional growth the province of Alicante. We also use it to illustrate our empirical analysis. The inset in figure 2a depicts u versus \dot{u} obtained using equation (2.1) for all the municipalities in the last 15 years. The correlation coefficient between the two microscopic variables is found to be 0.16, small enough to consider, at zeroth-order approximation, that the growth is of proportional nature, and thus independently of the size and obeying equation (3.5).

As seen in §3.4, the macroscopic equilibrium RD predicted for proportional growth by MaxEnt follows equation (3.8), where Λ is univocally determined via equation (3.9), with the given total population $N = 1\,926\,285$ (in 2010), the number of municipalities $n = 141$ and an estimation of the minimum population x_0 . N and n are well determined in all cases. The simplest way of estimating x_0 is via extrapolation in the rank-plot of a linear fit to the logarithm of the lowest populations. A more sophisticated method (that we

use here) uses a nonlinear fitting of the raw data to the MaxEnt rank-distribution via this single parameter, using equation (3.11) together with the definition of Λ equation (3.9). We have obtained $\log(x_0) = 4.83$, with a correlation coefficient of 0.9997. The comparison of the raw data with the analytical rank-distribution is displayed in figure 2*a*. The shadowed area represents the 90% CI owing to the finite value of n (numerically estimated as described in appendix A). The empirical RD falls inside the CI, and the observed microscopic dynamics are compatible with the dynamical assumption of equation (3.5). We consider then that this case constitutes strong evidence for the applicability of the MaxEnt principle for scale-invariant social systems.

We have applied the same approximation of proportional growth to the rest of the Spanish provinces, also including the values of the total populations N in the fitting procedure. Using the corresponding values of n together with the estimated values of x_0 and N , we get the value for Λ , and apply then the Gamma Scaling Law described in §3.4. The scaled RDs are displayed in figure 2 (raw data in the inset), and the numerical values of the fitted results can be found in electronic supplementary material, table S1. A quite nice adjustment ensues in general but not in all instances. This ‘failure’ is linked to the strong correlations between \dot{u} and u for some provinces, that reaches significant values (of up to 0.57 for the province of Lugo). Although such correlations may compromise the validity of the approximation of proportional growth, the scaling features exhibited by this plot are remarkable indeed.

It is worth mentioning that the fitted value for the total population N is found to be systematically lower than the actual value for it. Such a scenario usually changes when the capital city of the province is not considered in the fitting process, and thus in the estimation of N . This indicates that the population of the capital city is systematically larger than what one would expect from the MaxEnt prediction (of the 90% confidence level). This is not surprising because, as mentioned earlier, provinces are not ideal, isolated systems: the actual administrative municipality for capital cities is usually the sum of the historical ones plus some near neighbours, and their economy is expected to be highly correlated with that of other capital cities. Although only 50 cities is a small sample, we have studied the rank-distribution of these capitals and their dynamics to shed some light on this observation. We have found the pleasing result that these capitals form a *scale-free system of their own*. We compare in figure 3 the RD with the MaxEnt prediction using the actual value $N = 15528025$ with the only fitting parameter $\log(x_0) = 10.80(5)$ (with a correlation of 0.99989). When studying the relative increment \dot{u} versus the log-size u , we find very low size-dependencies (correlation of 0.15, lower than the one prevailing for Alicante). This observation confirms the appropriateness of a proportional growth dynamics. Again, the capital city of Madrid exhibits a larger-than-expected population (just in the limit of the 90% confidence level), indicating the possible existence of a higher, international system of correlations.

4.2. q -Growth with proportional drift analysis

The main limitation of the proportional growth approximation is that the model is not able to describe some dynamical patterns such as, for example, the generalized migration from small villages to big cities. To reproduce

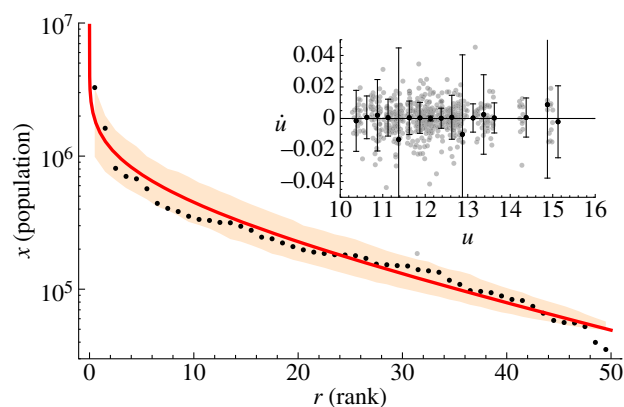


Figure 3. Rank-distribution for the capital cities (2010), compared with the corresponding MaxEnt prediction for proportional growth. The shadowed area represents the 90% CI for $n = 50$. Inset: relative change versus natural logarithm of the population for the last 15 years, where each grey dot represents one city for each annual period. Black dots denote the mean value in each interval $\Delta u = 0.25$ and the error bars the standard deviation for that interval. (Online version in colour.)

such behaviour, some dependence of the relative growth \dot{u} with the population via u is needed. We start from a Taylor expansion in u up to second-order $\dot{u} \simeq a + bu + cu^2$. Considering it as the second-order expansion of an exponential function, we write

$$\dot{u} \simeq a + bu + cu^2 \simeq k_1 + k_q e^{(q-1)u}, \quad (4.1)$$

where $a = k_1 + k_q$, $b = k_q(q-1)$ and $c = k_q(q-1)^2/2$. Thus, we face a microscopic ‘ q -dynamics’ with proportional noise (equation (3.23)) whose macroscopic distribution will follow equation (3.24). The actual q -value can then be directly determined either from the macroscopic RD or from the microscopic dynamical data. If both estimations match, the theoretical procedure should be correct. For the first estimation, we obtain from the RDs the parameters x_0 , Λ , q and σ of equation (3.24) using the fitting method described in appendix A. For the second estimation, we fit the ‘local’ mean value of \dot{u} in equation (3.23) to the expression (see appendix A)

$$\langle \dot{u}(u) \rangle = \langle k_1 \rangle + \langle k_q \rangle e^{(q-1)u}. \quad (4.2)$$

4.2.1. Results on random walkers simulation

To test both fitting procedures, we have performed a numerical simulation of drifted q -metric walkers as described in §3.4. We have arbitrarily used $n = 200$ walkers with an initial total population of $N = 50\,000$, minimum allowed population $x_0 = 1$ and a dynamical value $q = 1.2$. We use in the simulation a time interval of $\Delta t = 0.03$, and generate normally distributed random numbers for k_1 and k_q with standard deviations $\sigma_1 = 0.6$ and $\sigma_q = 0.1$, respectively. We also set a mean value $\langle k_1 \rangle = 0.05$. We stop after $m = 2650$ iterations. At the end of the simulation both the total population N and the minimum size x_0 have grown by a factor of $\exp[m\Delta t\langle k_1 \rangle] = \exp(3.98)$. Proportional noise contributes to the final width with $\sigma = \Delta t\sqrt{n}\sigma_1 = 0.93$. The final Λ value predicted by MaxEnt via equation (3.25) using these values is $\log(\Lambda) = -7.8$. We show the final RD in figure 4, with the dynamical data obtained using equations (2.1) in the inset.

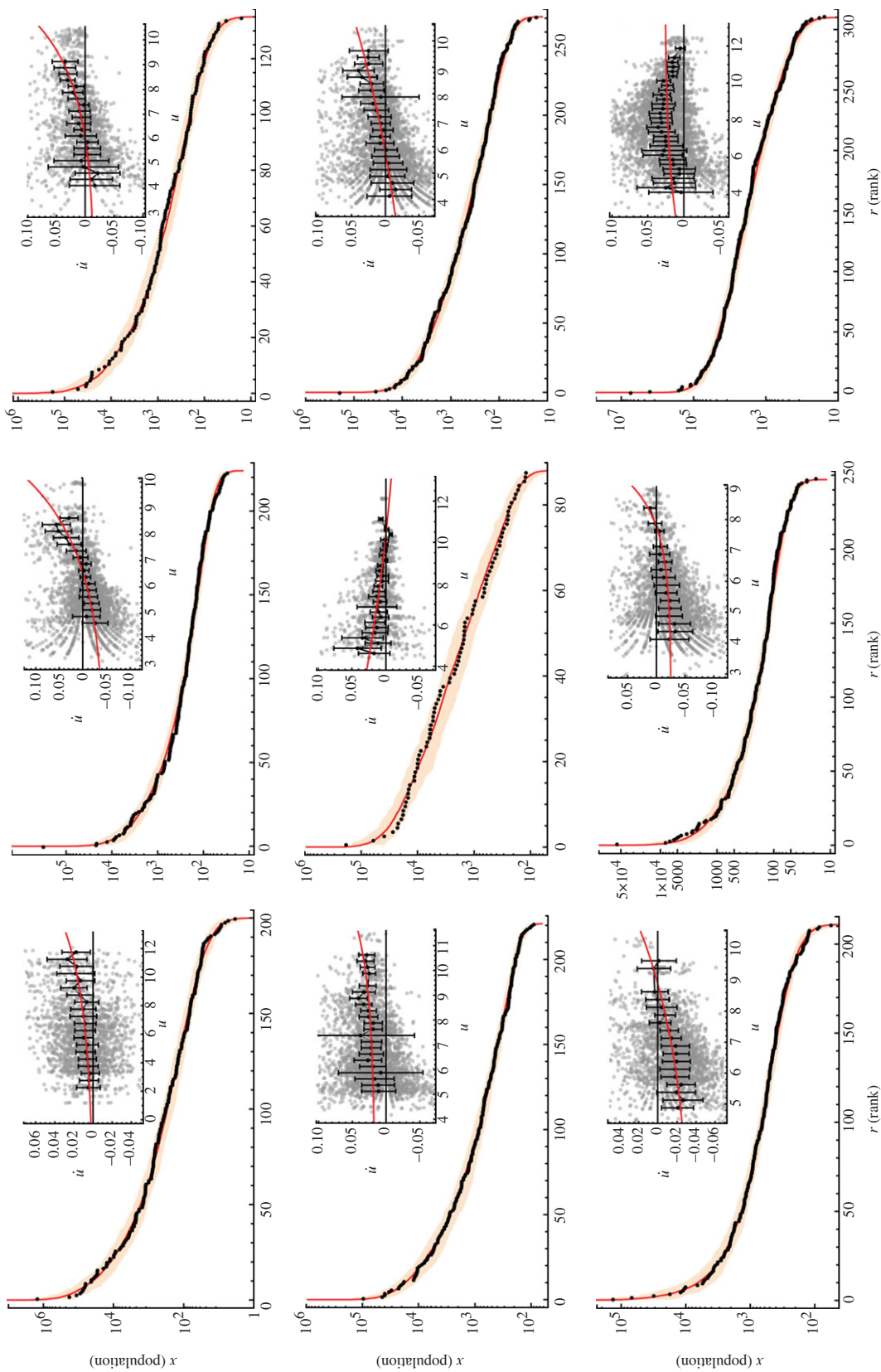


Figure 4. Following the reading order, macroscopic rank-distribution of a numerical simulation of a numerical simulation of q -metric random walkers (see text), and empirical rank-distribution for the provinces of Valladolid, Castellón, Girona, Guipúzcoa, Navarra, León, Ávila and Barcelona (2010), compared with the pertinent MaxEnt prediction. The shaded area represents the 90% CI for each n value. Insets: microscopic dynamics with the relative growth versus logarithm of the population. Black dots denote the mean value in each interval $\Delta u = 0.25$ and the error bars the standard deviation for that interval. The red line shows the fit to equation (4.2). (Online version in colour.)

Submitting the RD data to the fitting procedure obtains the estimated values $\log(\Lambda) = -8.8$, $q = 1.32$, $\log(x_0) = 3.56$ and $\sigma = 1.15$. We show in the first panel of figure 4 the fit as a red line to the data. Comparing these values with the actual ones, we find a general discrepancy of 10 per cent (in particular, 20% for σ). This result gives us the order of the error of our fitting methodology. From the fit to the dynamics, we have obtained $q = 1.16 \pm 0.10$ (red line in the inset of first panel of figure 4), which is closer to the actual value of 1.2 with a discrepancy of 3 per cent. The factor between the two independent q -fits is 0.89, which gives an idea of the accuracy of the method, approximately 10 per cent.

4.2.2. Results on empirical city-population

As a general trend, we can fit the macroscopic RDs without too many technical complications. In many cases, the capital city has to be excluded, as found in §4.2.1. Also, in some other instances, a few of the largest cities have to be excluded. We have also found 'outliers' for very low population centres (details can be found in appendix A). We find for all the Spanish provinces (mean value \pm s.d.): $\log(x_0) \sim 5.5 \pm 1.3$; $\log(\Lambda) \sim -4.3 \pm 1.6$; $q \sim 1.20 \pm 0.45$; and $\sigma \sim 0.43 \pm 0.24$. We show in figure 4 some interesting examples of RD (plots for all provinces are found in the electronic supplementary material, together with the table with all the numerical values). Most of the provinces follow the analytical curve quite well, with very few exceptions. The most dramatic of those exceptions correspond to Salamanca, Orense and Zamora, where a few cities account for the main part of the province's total population and the small villages follow a lognormal distribution. It is expected that owing to their small size, the finite effect of the last term in equation (3.4) becomes important. Thus, the distribution is found to be noisy in these cases.

Proportional drift noise, scaled by σ , has shown to be negligible in some cases but quite important in others. In general, better fits are obtained for those provinces for which the distributions' evolution during the last 15 years has been smooth and slow. High rates of change correlate with high σ . This rather surprising result tells us that if the system is able to reach dynamic equilibrium, it converges to the MaxEnt prediction. To illustrate this, we plot in figure 5 the evolution in time of q for some provinces. We have also found that the correlation between the pair σ and $\log(\Lambda)$ and the pair q and $\log(\Lambda)$ is rather important ($R = -0.40$ and -0.37 , respectively). Indeed, equation (3.25) relates these parameters in an equation-of-state fashion.

On the other hand, the microscopic dynamical fit presents some preliminary difficulties. The main one is high noise for small city sizes (owing to the last term of equation (3.4), which cannot be neglected). Anyhow, because the mean value of this noise is zero, we can safely use equation (4.2) to fit the 'local' mean value of \dot{u} via q . However, in few cases, we have found technical troubles when the value of $\langle k_q \rangle$ is small in relation to the drift. Indeed, a low value of $\langle k_q \rangle$ generates a nearly linear dependence of \dot{u} with u , neglecting the second-order term of equation (4.1). This generates an ambiguous result for q , falling down to a generic $q = 1$ value in the fitting process. We have found this in nine of the 50 provinces. Unfortunately, an accurate estimation of q for these cases cannot be achieved from the dynamics.

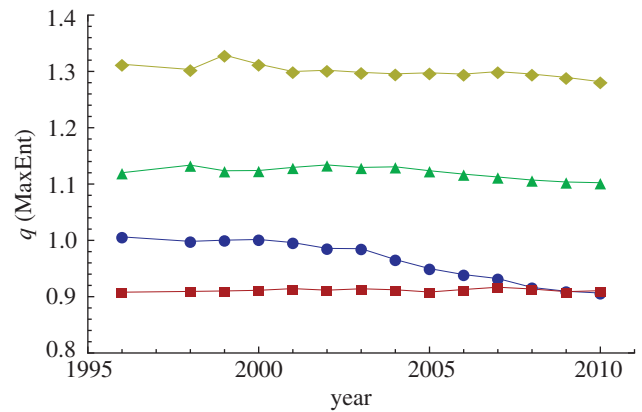


Figure 5. Evolution of q via MaxEnt distributions for Alicante (circles), Guipúzcoa (squares), Girona (diamonds) and Navarra (triangles). Alicante is the province with larger σ (0.15) of these four examples, correlated with its larger time-variation. (Online version in colour.)

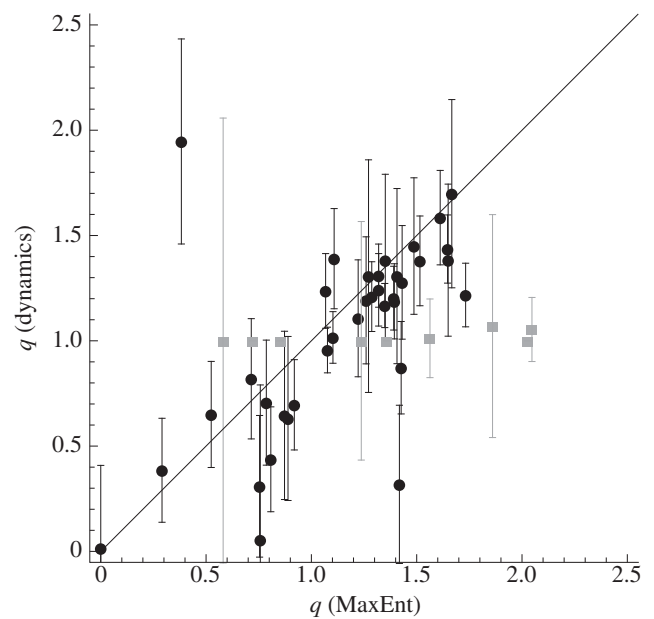


Figure 6. Comparison of independent measures of q , via the comparison MaxEnt distribution versus dynamics. (Grey colour represents those cases where the dynamical fit fails, see text.)

We depict in figure 6 the comparison between both independent sources for q , macroscopic versus microscopic, which represent *the main result of this work* (we display in grey the cases where the dynamical fit fails, near the dynamical value $q = 1$). We find a mean proportionality of 0.85 with a correlation of 0.97, and compared with the proportionality found in the control case (the random-walker simulation), we can assume that the discrepancy is the expected one using the fitting procedure.

In such instances, *the microscopic dynamical process determines the macroscopic rank-distribution of the city-population. The equilibrium distribution is that found via maximization of the Shannon entropy in the terms discussed above.*

5. Further remarks

5.1. Origin of the q -dynamics

In §4.2.2, we have successfully parametrized the pertinent dynamics *without discussing possible underlying mechanisms.*

We encounter two general trends: (A) migration from small villages to big cities, and (B) saturation for large cities. The cases of Valladolid, Castellón, Navarra, León and Ávila constitute good examples of the first scenario, the most frequent for Spain's provinces. We usually find a value $q > 1$ for them. Several of these examples illustrate the second possibility for cities with a population larger than 10 000, with a relative growth lower than expected. Such a situation is due to the finiteness of the resources needed to make a city grow, which impedes arbitrarily large growth rates. We also find examples, as shown for Guipúzcoa, for which the system follows the B-trend with values $q < 1$. These cases clearly indicate a generalized migration from cities to small towns, owing to local socioeconomic conditions. In other remarkable cases, both situations A and B occur in *the same region*, as in Madrid and Barcelona, where a migration to medium-sized towns occurs, a sign of non-monotonic behaviour for \dot{u} with respect to u , which in such instances compromises the dynamical fit (depicted here for Barcelona). On the other hand, the case of Girona deserves some comment. Even if it presents a tiny migration tendency, the system is very near total equilibrium, with an excellent fit for the RD.

The use of power laws to model both migration and saturation is here of a heuristic nature. We have no evidence of any 'microscopic' mechanism generating them, and the value of q is merely obtained from empirical observation. Identifying the mechanism that generates the dynamics constitutes a formidable challenge. It is worth mentioning that growth with saturation has been traditionally modelled using the so-called logistic function, or its generalization $X(t)$ with parameters α , K and v that follow a differential equation of the type (http://en.wikipedia.org/wiki/Generalised_logistic_function)

$$\dot{X}(t) = \alpha \left\{ 1 - \left[\frac{X(t)}{K} \right]^v \right\} X(t), \quad (5.1)$$

which formally is identical to equation (3.22) with $v = q - 1$, $\alpha = k_1$ and $\alpha/K = k_q$. Accordingly, migration and growth with saturation relates with hyper-exponential growth. However, as far as we know, a pure theoretical determination of the q -value from the underlying mechanism is still unknown.

5.2. The importance of the constraints in the microscopic dynamics

The constraints imposed on both the dynamics and the MaxEnt treatment play a fundamental role in the success of our approach. Indeed, fixing a minimum allowed population x_0 , different from zero, prevents the eventual 'death' of the city, and thus maintains constant the total number of cities in the system. Without this condition, the dynamics could lead to a stable state in which the population tends to zero, a problem that could be overcome by introducing a continuous growth via a positive value of the mean of the growth. However, this situation is very hard to maintain for a long period, so that it cannot be considered as a normal situation on which to base a general theory. On the other hand, we need a mean growth different from zero to describe migration patterns. These could lead to an unstable situation of continuous growth. The condition of constant total population for internal migration solves this issue in a natural fashion, and macroscopic equilibrium configurations can then be arrived at.

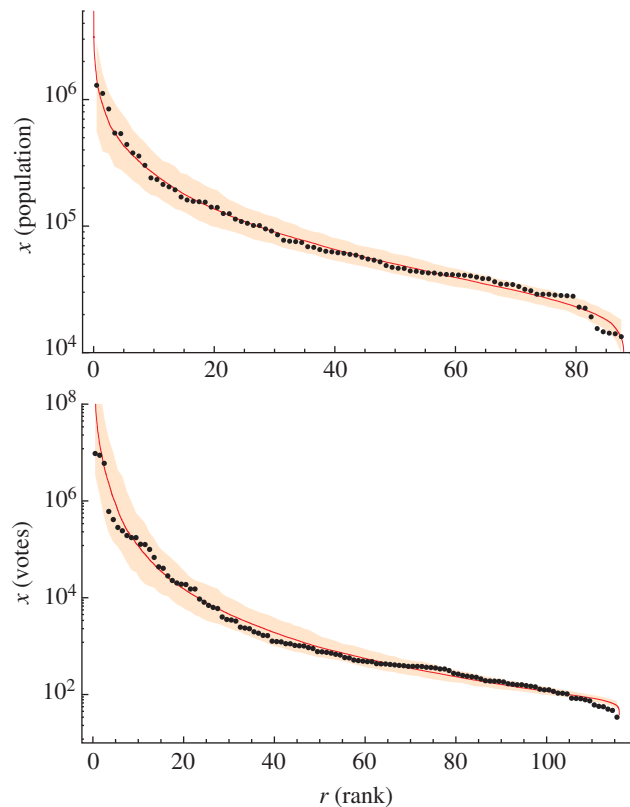


Figure 7. Same as figure 2 for the state of Ohio (USA) in 2010 (*a*), and the general electoral results in United Kingdom in 2005 (*b*). (Online version in colour.)

The above-mentioned constraints, together with a smooth growth for the total population, allow one to simultaneously describe the RDs for both large populations and small ones. Thus, our distribution equation (3.24) describes the tails in a better fashion than simple power laws. Indeed, the tail for small sizes was usually not considered if one wants to check the adequacy of a power law or other functional form. The use of administrative data for calculating the size of a settlement is subject to some limitations: (i) the arbitrary administrative area may be too small in the case of large cities, or (ii) too large in the case of small settlements, including more than one town so that their actual population cannot be properly assessed (as commented in §2). The latter problem is particularly evident when one considers the difference between the number of municipalities in France (36 781) and Spain (8116), although the population of France is only 38 per cent higher than that of Spain. Equation (3.24) overcomes this limitation. *Per contra*, as we have seen in §4.2.2, it fails in some cases (and for a few populations in the distribution tail, due to 'outliers', see appendix A) indicating that a case-by-case analysis is needed to describe some areas in order to identify whether the hypothesis of the model match the empirical situation.

5.3. Beyond Spanish city populations

We have also found evidence of MaxEnt equilibrium both in other countries and in other kinds of social systems, such as electoral results, in current and previous years. As an example, we plot in figure 7*a* the rank distribution with theoretical fit for the state of Ohio [40] ($\log(\Lambda) = -3.7$, $q = 1.62$, $\log(x_0) = 10.1$ and $\sigma = 0.36$). We also show in the bottom panel the electoral results for the 2005 UK general election [41] ($\log(\Lambda) = -13.5$,

$q = 13.5$, $\log(x_0) = 4.15$ and $\sigma = 0.05$). Deeper understanding of each territory/election correlated with the kind of study presented here may help to better appreciate the effects of regional policies. Work along such lines is in progress.

We emphasize that the formalism applied here is of a quite general nature and can be applied to other kind of systems as well. If a dynamical equation

- (1) can be parametrized, and
- (2) a change of variables able to linearize it can be concocted,

one would be able to (i) recover the present scenario and (ii) comfortably use the present procedure.

6. Summary

We have shown that the observed macroscopic city-population distributions of the Spanish provinces are, in general, MaxEnt equilibrium distributions, according to the extant microscopic growth dynamics. The MaxEnt variational process is constrained by appropriate constraints (minimum allowed size different from zero and conservation of the total population for internal migration flows). Our procedure in this communication can be summed up as follows.

- (i) We first considered a proportional growth approach to the problem, finding that the empirical distributions can be nicely scaled following the Gamma Scaling Law, derived from the equilibrium distribution of a scale-free system.
- (ii) Afterwards, we dealt with q -growth processes, with some proportional drift. We accounted in this way for both migration and 'natural' growth of the population, obtaining better fits to the data than in case (i).

We have checked that the value of the dynamical exponent q , estimated from the RDs, is equivalent to the one estimated directly from the dynamics. We read this result as a confirmation of the validity of our approach, establishing thus a connection between microscopic and macroscopic worlds via MaxEnt.

We greatly appreciate the insightful comments of the two reviewers and their help in the improvement of the manuscript. This work was partially supported by ANR DYNHELIUM (ANR-08-BLAN-0146-01) Toulouse, and the Projects FQM-2445 and FQM-207 of the Junta de Andalucía, PIP1177 of CONICET (Argentina), FIS2008-00781/FIS (MICINN) and FEDER (EU; Spain, EU). A.P. acknowledges support from the Senior Grant CEI Bio-Tic GENIL-SPR.

Appendix A

A.1. Estimation of the distribution's 90% CI

The confidence levels of the RDs for a given number n are estimated as follows.

- (i) A list of n random numbers is generated, following the desired distribution, by inverse transform sampling.
- (ii) The list is sorted from largest to lowest in order to obtain the rank-distribution. The list is saved for further use.
- (iii) A large number of lists is generated following (i) and (ii), obtaining a distribution of values for each rank.

- (iv) The 0.95th and 0.05th quantiles are obtained from each rank, determining the lower and upper limits of the 90% CI.

A.2. Fit for q -exponential growth

A first method to estimate the MaxEnt q -distributions' parameters Λ , x_0 and q for a given empirical data is a direct fit of the rank-distribution to equation (3.21). We have used the MATHEMATICA software (Wolfram Research Inc. 1988–2011) to this end, by means of the Nonlinear Model Fit function using different guesses for the initial values. We also compare the pertinent results with a solution obtained from the reproduction of the first three moments of the distribution, finding, statistically and numerically, better stability using the logarithmic moments $\langle (\log(x/x_0))^m \rangle$ instead of $\langle (x/x_0)^m \rangle$. We have tabulated them as functions of q and Λ in the form $M_m(q, \Lambda) = \langle [\log(x/x_0)]^m \rangle$, which are defined via

$$\left. \begin{aligned} \left\langle \log\left(\frac{x}{x_0}\right) \right\rangle &= \frac{1}{\Gamma(1-q, \Lambda)} G_{23}^{30} \left(\Lambda \left| \begin{matrix} 11 \\ 00(1-q) \end{matrix} \right. \right) \\ \left\langle \left[\log\left(\frac{x}{x_0}\right) \right]^2 \right\rangle &= \frac{2}{\Gamma(1-q, \Lambda)} G_{34}^{40} \left(\Lambda \left| \begin{matrix} 111 \\ 000(1-q) \end{matrix} \right. \right) \\ \text{and } \left\langle \left[\log\left(\frac{x}{x_0}\right) \right]^m \right\rangle &= \frac{n!}{\Gamma(1-q, \Lambda)} G_{(m+1)(m+2)}^{(m+2)0} \left(\Lambda \left| \begin{matrix} 1 \dots 1 \\ 0 \dots (1-q) \end{matrix} \right. \right), \end{aligned} \right\} \quad (\text{A1})$$

where G stands for the so-called Meijer G -functions. The associated system of equations is

$$\begin{aligned} M_m(q, \Lambda) &= E \left[\log^m \left(\frac{x}{x_0} \right) \right] \\ &= \sum_{m'=0}^m (-1)^{m'} \binom{m}{m'} E[\log^{m-m'}(x)] \log^{m'}(x_0), \end{aligned} \quad (\text{A2})$$

which we deterministically solve using $m = 1, 2, 3$ via Λ , x_0 and q , with the empirical expected values $E[\log^m(x)] = \sum_{i=1}^m \log^m(x_i)/n$.

We have also introduced into our equations the proportional drift parametrized with σ . Its inclusion in the above equations can be easily materialized using the additive property of cumulants for convolutions (Wikipedia, en.wikipedia.org/wiki/Cumulant). Taking into account that for the centred normal distribution only the second cumulant does not vanish, we face the following system of equations:

$$\sum_{i=0}^{\lfloor m \rfloor} T_i^m \sigma^{2i} M_{m-2i}(q, \Lambda) = \sum_{m'=0}^m (-1)^{m'} \binom{m}{m'} E[\log^{m-m'}(x)] \log^{m'}(x_0), \quad (\text{A3})$$

where T_i^m is the i th element of the m th row of the triangle of Bessel numbers [42] $T_i^m = m!/(i!2^i(m-2i)!)$.

The ensuing system of equations is solved using the FindRoot functionality of MATHEMATICA with different starting points so as to ensure the best possible result. We accept a set of parameters if the three different results (fit, and equations' system with/without drift) are reasonably similar. When that does not happen, we look for outliers, excluding some points from the tails (largest or smallest) until reaching a satisfactory convergency of the three results. The most frequently outlier case obtains for the capital city. The second case refers to a few small villages with undersized population. All outliers are indicated in the electronic

supplementary material, table S1, together with the results of the three estimations. On the other hand, as commented in the text, we have found three notorious cases—Salamanca, Orense, and Zamora—where no satisfactory outcome has been achieved with this procedure.

A.3. Fitting the dynamical data

As commented in the text, we have fitted the mean value $\langle \dot{u}(u) \rangle$ to equation (4.2) via $\langle k_1 \rangle$, $\langle k_q \rangle$ and q . We have used again the Nonlinear Model Fit function, with different

guesses for the initial values. The dataset for $\langle \dot{u}(u) \rangle$ (and also the standard deviation) is systematically computed in bins of size $\Delta u = 0.25$ for all the provinces. Very few points can be found in some of the bins, which introduces high numerical error. Accordingly, we include the bin in the dataset if, and only if, a minimal number of points exists. We have assumed, for all the provinces, that this number is the 15 per cent of the bin with the larger number of points. The bins used in each fit are shown in the inset of the rank-distribution plots.

References

- Batty M. 2008 The size, scale, and shape of cities. *Science* **319**, 769–771. (doi:10.1126/science.1151419)
- Blank A, Solomon S. 2000 Power laws in cities population, financial markets and internet sites *Physica A* **287**, 279–288. (doi:10.1016/S0378-4371(00)00464-7)
- Gabaix X, Ioannides YM. 2004 The evolution of city size distributions. In *Handbook of regional and urban economics*, vol. 4 (eds J. V. Henderson, J. F. Thisse), pp. 2341–2378, 1st edn. Amsterdam, The Netherlands: Elsevier.
- Reed WJ. 2002 On the rank-size distribution for human settlements. *J. Reg. Sci.* **42**, 1–17. (doi:10.1111/1467-9787.00247)
- Newman MEJ. 2005 Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351.
- Pareto V. 1896 *Cours d'Economie Politique*. Geneva, Switzerland: Droz.
- Zipf GK. 1949 *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.
- Corominas-Murtra B, Solé RV. 2010 Universality of Zipf's law. *Phys. Rev. E* **82**, 11102. (doi:10.1103/PhysRevE.82.011102)
- Malacarne LC, Mendes RS, Lenzi EK. 2001 q-Exponential distribution in urban agglomeration. *Phys. Rev. E* **65**, 017 106–017 109. (doi:10.1103/PhysRevE.65.017106)
- Marsili M, Zhang YC. 1998 Interacting individuals leading to Zipf's law. *Phys. Rev. Lett.* **80**, 2741–2745. (doi:10.1103/PhysRevLett.80.2741)
- Gabaix X. 1999 Zipf's law for cities: an explanation. *Q. J. Econ.* **114**, 739–767. (doi:10.1162/003355399556133)
- Axtell RL. 2001 Zipf distribution of US firm sizes. *Science* **293**, 1818–1820. (doi:10.1126/science.1062081)
- Newman MEJ. 2001 Scientific collaboration networks. *Phys. Rev. E* **64**, 016 131–016 139. (doi:10.1103/PhysRevE.64.016131)
- Hernando A, Puigdomènech D, Villuendas D, Vesperinas C, Plastino A. 2009 Zipf's law from a Fisher variational-principle. *Phys. Lett. A* **374**, 18–23. (doi:10.1016/j.physleta.2009.10.027)
- Albert R, Barabási AL. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2074**, 2047.
- Newman MEJ, Barabási AL, Watts DJ. 2006 *The structure and dynamics of complex networks*. Princeton, NJ: Princeton University Press.
- Maillard T, Sornette D, Spaeth S, von Krogh G. 2008 Empirical tests of Zipf's law mechanism in open source Linux distribution. *Phys. Rev. Lett.* **101**, 218701. (doi:10.1103/PhysRevLett.101.218701)
- Costa-Filho RN, Almeida MP, Andrade JS, Moreira JE. 1999 Scaling behavior in a proportional voting process. *Phys. Rev. E* **60**, 1067–1068. (doi:10.1103/PhysRevE.60.1067)
- Weisstein EW. 2012 *Benford's law*. From MathWorld See <http://mathworld.wolfram.com/BenfordsLaw.html>.
- Hernando A, Vesperinas C, Plastino A. 2010 Fisher information and the thermodynamics of scale-invariant systems. *Phys. A* **389**, 490–498. (doi:10.1016/j.physa.2009.09.054)
- Frieden R, Plastino A, Plastino AR, Soffer BH. 1999 Fisher-based thermodynamics: its Legendre transform and concavity properties. *Phys. Rev. E* **60**, 48–53. (doi:10.1103/PhysRevE.60.48)
- Castellano C, Fortunato S, Loreto V. 2009 Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646. (doi:10.1103/RevModPhys.81.591)
- Fortunato S, Castellano C. 2007 Scaling and universality in proportional elections. *Phys. Rev. Lett.* **99**, p138 701–138 705. (doi:10.1103/PhysRevLett.99.138701)
- Hernando A, Villuendas D, Vesperinas C, Abad M, Plastino A. 2010 Unravelling the size distribution of social groups with information theory in complex networks. *Eur. Phys. J. B* **76**, 87–97. (doi:10.1140/epjb/e2010-00216-1)
- Hernando A, Plastino A, Plastino AR. 2012 MaxEnt and dynamical information *Eur. Phys. J. B* **85**, 147. (doi:10.1140/epjb/e2012-30009-3)
- Jaynes ET. 1957 Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630. (doi:10.1103/PhysRev.106.620)
- Jaynes ET. 1957 Information theory and statistical mechanics. II. *Phys. Rev.* **108**, 171–190. (doi:10.1103/PhysRev.108.171)
- Jaynes ET. 1968 Prior probabilities. *IEEE Trans. Syst. Sci. Cyb.* **4**, 227–241. (doi:10.1109/TSSC.1968.300117)
- Katz A. 1967 *Principles of statistical mechanics: the information theory approach*. San Francisco, CA: W. H. Freeman.
- Baek SK, Bernhardsson S, Minnhagen P. 2011 Zipf's law unzipped. *New J. Phys.* **13**, 043004. (doi:10.1088/1367-2630/13/4/043004)
- Harremoes P, Topsoe F. 2001 Maximum entropy fundamentals. *Entropy* **3**, 191–226. (doi:10.3390/e3030191)
- Harremoes P, Topsoe F. 2002 Zipf's law, hyperbolic distributions and entropy. In *IEEE International Symposium on Information Theory: Proceedings*, Lausanne, Switzerland.
- Zanette DH, Manrubia SC. 1997 Role of intermittency in urban development: a model of large-scale city formation. *Phys. Rev. Lett.* **79**, 523–526. (doi:10.1103/PhysRevLett.79.523)
- National Statistics Institute of Spain. 2012 Government of Spain. See www.ine.es.
- Tsallis C. 1988 Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52**, 479–487. (doi:10.1007/BF01016429)
- Tsallis C. 1998 Generalized entropy-based criterion for consistent testing. *Phys. Rev. E* **58**, 1442–1445. (doi:10.1103/PhysRevE.58.1442)
- Kowalski AM, Martin MT, Plastino A, Zunino L. 2009 Tsallis' deformation parameter q-quantifies the classical-quantum transition. *Physica A* **388**, 1985–1994. (doi:10.1016/j.physa.2009.01.036)
- Plastino A, Plastino AR. 1999 Tsallis entropy and Jaynes' information theory formalism. *Braz. J. Phys.* **29**, 50–60. (doi:10.1590/S0103-97331999000100005)
- Hanel R, Thurner S, Gell-Mann M. 2011 Generalized entropies and the transformation group of superstatistics. *Proc. Natl Acad. Sci. USA* **108**, 6390. (doi:10.1073/pnas.1103539108)
- Census Bureau. 2012 Government of USA. See www.census.gov
- Electoral Commission. 2012 Government of UK. See www.electoralcommission.org.uk.
- Deutsch E. 2005 Sequence A100861. In *The online encyclopedia of integer sequences*. See <http://oeis.org>.