Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Physica A 355 (2005) 224–231
PHYSICA A
www.elsevier.com/locate/physa

Herding and clustering: Ewens vs. Simon–Yule models

D. Costantini[a], S. Donadio[b], U. Garibaldi[c], P. Viarengo[d,*]

[a]*LABORatorio Riccardo Revelli, via Real Collegio 30, 10024 Moncalieri (Turin), Italy*
[b]*INFN, University of Genoa, Department of Physics, via Dodecaneso 33, 16146 Genoa, Italy*
[c]*IMEM-CNR, University of Genoa, c/o Department of Physics, via Dodecaneso 33, 16146 Genoa, Italy*
[d]*National Cancer Research Institute (IST), Largo Benzi 10, Genoa 16132, Italy*

Received 31 October 2004; received in revised form 11 January 2005
Available online 28 April 2005

Abstract

Clustering has often described by Ewens Sampling Formula (ESF). Focusing the attention on the evergreen problem of the size of firms, we discuss the compatibility of empirical data and ESF. In order to obtain a power law for all sizes in the present paper we shall explore the route inspired by Yule, Zipf and Simon. It differs from the Ewens model both for destruction and creation. In particular the probability of herding is independent on the size of the herd. Computer simulations seem to confirm that actually the mean number of clusters of size i (the equilibrium distribution) follows the corresponding Yule distribution. Finally we introduce a finite Markov chain, that resembles the marginal dynamics of a cluster, which drives the cluster to a censored Yule distribution.

PACS: 89.65.Gh; 89.75.Fb; 02.50.Ga

Keywords: Ehrenfest–Brillouin chain; Simon–Yule model; Ewens sampling formula; Zipf's law

*Corresponding author. National Cancer Research Institute (IST), Largo Benzi 10, Genoa 16132, Italy.
E-mail addresses: garibaldi@fisica.unige.it (U. Garibaldi), paolo.viarengo@istge.it (P. Viarengo).

0378-4371/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2005.02.087

1. Introduction

The clustering of agents in the market is a typical problem dealt with by the new approaches to macroeconomic modeling, that describe macroscopic variables in terms of the behavior of a large collection of microeconomic entities. Clustering [1] has often described by Ewens Sampling Formula (ESF) [2]. In Ref. [3], some of us have suggested an exact finitary characterization of the ESF, that admits a very nice interpretation in terms of rational vs. herd behavior. In this paper, we focus on the problem of the size of firms. Given that the compatibility of empirical data [4] and ESF is poor, we turn to an alternative dynamical model, traced to Simon [5]. We look for a microscopic explanation of the empirical power law based on a finite population of elementary units which interact as time goes by. We start from individual agents, that change job following a well-defined probabilistic rule. The resulting process is a finite homogeneous Markov chain, where time and state space can eventually achieve continuous limits. Starting with (continuous) stochastic process [6], even if results were satisfactory, would shadow the concrete (discrete) dynamics of agents that determines the final result. This is the case of most of the stochastic explanations quoted in Ref. [4]. Our microscopic explanations should be compared with the so-called agent-based computational models [7]. Our ambition is to bridge the gap between these models, where there is a lack of probabilistic insight, and stochastic processes, that appear "phenomenological" if they are non "agent-based".

2. The Ehrenfest–Brillouin dynamics and the Ewens limit

A dynamic system is composed of a population of n entities and $g>n$ cells (sites), whose state is described by the set of nonnegative occupation numbers $\{n_i : n_i \geqslant 0, \sum_{i=1}^{g} n_i = n\}$. Each site with $n_i>0$ hosts a cluster, the number of distinct cluster is $k \leqslant n$, so that $g-k>0$ sites are empty [3]. The state changes over discrete times as some units move from site to site. The probabilistic dynamics is modelled as an extraction of some units, which temporarily abandon the population, followed by a re-accommodation of the same units, usually into sites different from the original ones. Then size of the population shrinks as long as selection occurs, and it returns gradually to the original size n after all accommodations. The core of the Ehrenfest–Brillouin dynamics [8] in the Ewens limit [9] is as follows: (I) the selection of the moving agents is given by a random sampling; (II) the probability of re-accommodation (the creation probability) depends on the current size $v<n$ of the population and on $\theta>0$, the sole parameter of the dynamics. The parameter v is the total weight of joining existing clusters (herding), while θ is the total weight of founding a new cluster (pioneering). Hence at each accommodation a new cluster can be born with innovation probability $u=\theta/(\theta+v)$, while $1-u=v/(\theta+v)$ is the probability of joining the current herd. Introducing $\mathbf{z}=(z_1,\ldots,z_n)$, $\sum_{i=1}^{n} iz_i = n$, where z_i is the number of clusters of size i, the equilibrium distribution of $\mathbf{z}$ is just the Ewens Sampling Formula [3,9]. Passing to mean values, the mean number of cluster