J Econ Interac Coord (2007) 2:147–161
DOI 10.1007/s11403-007-0022-2

REGULAR ARTICLE

The two-parameter Ewens distribution: a finitary approach

U. Garibaldi · D. Costantini · P. Viarengo

Published online: 27 June 2007

Abstract The well-known Ewens Sampling Formula has been generalized by Pitman recently. We derive some essential feature of the model without introducing notions like frequency spectrum, structure distribution or size-biased permutation invariance, that are difficult to apply to concrete finite populations. A finite model of economic interacting agents whose equilibrium aggregation state is described by the two-parameter Ewens distribution is presented. The exact marginal description of a site is derived, wherefrom birth, life and death of clusters is easy to extract; and a code for computer simulations of the life of clusters is enclosed.

1 Introduction

The clustering of agents in the market is a typical problem dealt with by the new approaches to macroeconomic modelling, that describe macroscopic variables in terms of the behavior of a large collection of microeconomic entities. Quoting Aoki and Yoshikawa (2007), "..A cluster is a group of economic agents. It can be a sector, an industry, or any other group of economic agents with the same choice or same set of

U. Garibaldi (✉)
IMEM-CNR, c/o Department of Physics,
University of Genoa, via Dodecaneso 33, 16146 Genoa, Italy
e-mail: garibaldi@fisica.unige.it

D. Costantini
Health Physics' Laboratory, Department of Physics,
University of Genoa, via Dodecaneso 33, 16146 Genoa, Italy

P. Viarengo
Department of Aerospace Engineering,
University of Naples "Federico II", P. le Tecchio 80, 80125 Naples, Italy

attributes.... What Dynamics emerges in the processes of formation and dissolution of clusters comprising interacting agents?". Clustering (Aoki 1996, 2000, 2002) has often been described by Ewens Sampling Formula (ESF) (Ewens 1972). At variance to the usual complex derivations (Kingman 1978), we have suggested a finitary characterization of the ESF pointing to real economic processes (Garibaldi et al. 2003b, 2005), that admits a nice interpretation in terms of rational versus herding behavior. In this paper we extend our approach to the two-parameter Ewens distribution, an enrichment of the original treatise due to Pitman (1992), Pitman and Yor (1996) and Zabell (1992). Pitman's interest to the field is based on mathematical statistics, while Zabell aims at predictive inferences, in particular to a generalization of the Johnson-Carnap continuum for the case of universal generalization. A short history of the subject if found in Zabell (1996). Our field is neither sampling nor predicting, as we mean to apply this formula to formation and dissolution of clusters, due to some "physical" attitude that rules the behavior of economical agents. Given the partition of our finite population into some groups, a probabilistic mechanism of destruction-creation is assumed to move agents from cluster to cluster, or to form new ones, with the possibility of dissolution of some existing cluster if it is abandoned by all agents.

2 Two parameter Ewens distribution

The cluster description of n elements is based on the elementary fact that two units belong to the same cluster iff $X_i = X_j$, where the range of X is the set of allowed possibilities (categories, strategies). Given that categories are nothing but labels for distinguishing clusters, both sampling or predicting amount to give probability to a sequence of n random variables $Y_1, \dots, Y_n$, where $Y_1 = 1$ (the first observed category by definition) whatever X_1, $Y_2 = 1$ if $X_2 = X_1$, while $Y_2 = 2$ (the second observed category) if $X_2 \neq X_1$, and so on. Hence the range of Y is "the first, the second, ..., the $K_m - th$ appeared category in the sequence". The sentence "the forth agent joins the first cluster" means that the $X_4 = X_1$ while "the forth agent joins the second cluster" means that the $X_4 = X_2 \neq X_1$, or $X_4 = X_3 \neq X_2 = X_1$. These Y random variables are associated to temporal partitions (Zabell 1992), and are sufficient both for sampling and predicting. Given $m < n$ elements described by $Y_1, \dots, Y_m$, let $\mathbf{A}_m = (A_{1,m}, \dots, A_{K_m,m})$ the frequency vector of the K_m so far observed categories. Then $A_{j,m} = \#\{Y_i = j, i = 1, \dots, m\}$. Both sampling and predicting amount to characterize the following probability $P(Y_{m+1}|Y_1, \dots, Y_m)$, for $m = 0, 1, \dots, n-1$, being n the size of the finite population (or the sample). $P(Y_{m+1}|Y_1, \dots, Y_m)$ usually (under the exchangeability condition) simplifies to

$$P(Y_{m+1}|Y_1, \dots, Y_m) = P(Y_{m+1}|A_{1,m}, \dots, A_{K_m,m}) \tag{1}$$

The range of Y_{m+1} is $1, \dots, K_m, K_m + 1 \leq m + 1$, where $1, \dots, K_m$ denote one of the already appeared categories, and $K_m + 1$ a new one. Note that K_m is random, $1 \leq K_m \leq m$. If $Y_{m+1} = j \leq K_m$ the new observation augments the size of an already existing cluster, while $Y_{m+1} = K_m + 1$ gives rise to a new cluster. $\mathbf{A}_n$ is the frequency vector of the K_n categories observed in the sample, that is the number of elements classified into the first, ..., the K_n observed category. The resulting cluster size distribution on the $n-$sample is the partition vector $\mathbf{Z}_n=(Z_{1,n},\ldots,Z_{n,n})$, that is the frequency vector of $\mathbf{A}_n$, the end of the approach. Then $Z_{j,n}=\#\{A_{i,n}=j, i=1,\ldots,K_n\}$. F.i. $\mathbf{Z}_n=(n,0,\ldots,0)$ describes a population composed by all singletons, while $\mathbf{Z}_n=(0,0,\ldots,1)$ describes a population where all units stick together in the same cluster. Being $\mathbf{Z}_n$ a frequency (of frequencies) vector, it is invariant for the labels attached to the clusters.

The two parameter Ewens distribution is obtained by choosing the following probability function for (1):

$$P(Y_{m+1}=j|A_{1,m},\ldots,A_{K_m,m})=\begin{cases}\frac{A_{j,m}-\alpha}{m+\theta} & j\leq K_m\\ \frac{\theta+K_m\alpha}{m+\theta}, & j=K_m+1\end{cases} \tag{2}$$

with $0\leq\alpha<1, \theta+\alpha>0$, where $A_{j,m}=\#\{Y_i=j, i=1,\ldots,m\}$ for $j=1,\ldots,m$ are the frequencies of the distinct observed values in the order that they appear, and K_m is their number (Hansen and Pitman 2000). The statistical background is that of "species sampling", where the interest is focused on the abundancies of the different species (not in their names). Hence species are classified as "the first, the second,.." in the order that they appear in the sample. For predictive inferences this approach is useful when the possible categories of the observed units are non known in advance, or they are infinite, so that any event like $X_i=j$ has a vanishing probability.

2.1 The two-parameter Hoppe urn

It can be modeled by a two parameter Hoppe urn, described by Zabell (1992). "Imagine an urn containing both colored and black balls, from which balls are both drawn and then replaced. Each time a colored ball is drawn, it is put back into the urn, together with a new ball with the same color having a unit weight, each time a black ball is selected, it is put back into the urn, together with two new balls, one black and having weight α, and one of a new color, having weight $1-\alpha$. Initially the urn contains a single black ball (the mutator)". In words, if an already observed color is drawn, the updating mechanism is Polya (the weight of the color increases by one), while a new color receives an initial weight equal to $1-\alpha$. If $\alpha=0$ the model reduces to that of the Hoppe urn (Hoppe 1987), where the updating mechanism is always Polya for colors and nothing for the mutator. The case $\alpha>0$ is what we study in the following: it implies an increasing weight α for the mutator every time it is drawn (a new color appears).[1]

[1] It is apparent that the model can be extended to $\alpha<0$, in which case the weight of the mutator decreases every time it is drawn. If $\theta=k|\alpha|$, whence k colors have appeared no new color can be extracted. Hence the sampling is identical to a $k-$dimensional Polya sampling with uniform initial weights equal to $|\alpha|$. The only difference is that in the $k-$dimensional Polya sampling the names of the categories are known in advance, and the initial weights are not forced to be uniform; in the two-parameter Hoppe scheme with $\alpha<0, \theta=k|\alpha|$ the names of the categories are pure labels, that can be represented by their order of appearance (Hansen and Pitman 2000).

A repeated application of (2) gives the probability of any sequence $P_{\theta,\alpha}(Y_1, \ldots, Y_n)$, that is exchangeable, that is all sequences belonging to the same $\mathbf{A}_n$ are equiprobable, and the same for $\mathbf{Z}_n$. Thus introducing combinatorial factors we get $P_{\theta,\alpha}(\mathbf{A}_n)$ and finally

$$P_{\theta,\alpha}(\mathbf{Z}_n) = \frac{n!}{\prod_{i=1}^{K_n} A_{i,n}!} \frac{1}{\prod_{j=1}^{n} Z_{j,n}!} P_{\theta,\alpha}(Y_1, \ldots, Y_n)$$
$$= \frac{n!}{\prod_{j=1}^{n} j!^{Z_{j,n}} Z_{j,n}!} P_{\theta,\alpha}(Y_1, \ldots, Y_n), \qquad (3)$$

given that $\prod_{i=1}^{K_n} A_{i,n}! = \prod_{j=1}^{n} j!^{Z_{j,n}}$. It is an extension of Ewens Distribution, that for $\alpha = 0$ reduces to $P_{\theta,0}(\mathbf{Z}_n) = P_{\theta}^{Ewens}(\mathbf{Z}_n)$ (see Appendix). Note that $P_{\theta}^{Ewens}(\mathbf{Z}_n)$ is a multivariate distribution whose domain as a function of n is enormously large ($\approx \frac{1}{4\sqrt{3}n} Exp[\sqrt{\frac{2n}{3}}]$), difficult to handle, and quite unfit to be compared with the distributions (Pareto, Exponential, LogNormal, Power laws,...) usually met in the statistical analysis of such populations. The theoretical object which will be candidate to the comparison is proportional to the much simpler $E_{\theta,\alpha}(Z_i)$, the mean number of clusters for any size $i = 1, \ldots, n$.

2.2 Sampling from a random distribution

Let us suppose to have a system of n agents, and all that we know is: they are grouped in (some) K_n clusters, (some) Z_1 singletons, (some) Z_2 pairs,..., $\sum_{i=1}^{n} Z_{i,n} = K_n$. The sole sure thing is $\sum_{i=1}^{n} i Z_{i,n} = n$. We assume further that $P(\mathbf{Z}_n) = P_{\theta,\alpha}(\mathbf{Z}_n)$. Now $\mathbf{Z}_n$ is not completely unknown, it is still uncertain, but its uncertainty is ruled by $P_{\theta,\alpha}(\mathbf{Z}_n)$. While in statistics this uncertainty is based on Bayesian roots, in our case it has a realistic root, as we will suppose that the system changes in time, and $P(\mathbf{Z}_n)$ is proportional to the fraction of time in which the cluster statistical distribution is $\mathbf{Z}_n$. To sample from this random distribution means to take a snapshot of the moving population, and then to observe in sequence all agents. The sequence $Y_1, \ldots, Y_n$ describes the sampled units (without replacement). If the sampling hypothesis is that all sequences conditioned to the actual $\mathbf{Z}_n$ are equiprobable, the probability of any sequence is then $P_{\theta,\alpha}(Y_1, \ldots, Y_n)$, and the predictive sampling probability is then (2).[2] It means that the first observed agent belongs to the first observed cluster (by definition), the second belongs to the first with probability $\frac{1-\alpha}{1+\theta}$, or belongs to a new cluster (the second) with probability $\frac{\theta+\alpha}{1+\theta}$, and so on. After having observed m agents, the $(m+1)th$-one belongs to a not yet observed cluster with probability $\frac{\theta+K_m\alpha}{m+\theta}$, that is the expectation of novelty is an increasing function of the so far observed clusters, while the expectation of already observed clusters $\frac{m-K_m\alpha}{m+\theta}$ is decreasing.

[2] This duality between predictive inferences and sampling from a random distribution is well-known in the realm of Bayesian statistics, and it based on de Finetti's Representation Theorem on exchangeable sequences.

2.3 The distribution of the size-biased pick and the calculation of $E_{\theta,\alpha}(Z_i)$

We may ask: what is the probability that the first observed cluster has size i, that is $P_{\theta,\alpha}(A_{1,n}=i)$?

In order to calculate $P_{\theta,\alpha}(A_{1,n}=i)$ we have to marginalize (2) on the first observed cluster, that is

$$\begin{cases} P(Y_{m+1}=1|A_{1,m})=\frac{A_{1,m}-\alpha}{m+\theta} \\ P(Y_{m+1}>1|A_{1,m})=\frac{m-A_{1,m}+\theta+\alpha}{m+\theta} \end{cases} \tag{4}$$

The sequence $(Y_1=1,\ldots,Y_i=1,Y_{i+1}>1,\ldots,Y_n>1)$ has probability $\frac{\theta(1-\alpha)\ldots(1-\alpha+i-1)\cdot(\theta+\alpha)(\theta+\alpha+1)\ldots(\theta+\alpha+n-i-1)}{\theta^{[n]}}$, hence

$$P(Y_1=1,\ldots,Y_i=1,Y_{i+1}>1,\ldots,Y_n>1)=\frac{\theta(1-\alpha)^{[i-1]}(\theta+\alpha)^{[n-i]}}{\theta^{[n]}} \tag{5}$$

and the combinatorial factor is $\binom{n-1}{i-1}$, due to the fact that the first variable is bound to be $Y_1=1$, hence

$$P_{\theta,\alpha}(A_{1,n}=i)=\frac{\theta(1-\alpha)^{[i-1]}}{n(i-1)!}\frac{(\theta+\alpha)^{[n-i]}/(n-i)!}{\theta^{[n]}/n!} \tag{6}$$

where we put (6) in a form easy to compare with the usual (Garibaldi et al. 2003b, 2005)

$$P_{\theta,0}(A_{1,n}=i)=\frac{\theta}{n}\frac{\theta^{[n-i]}/(n-i)!}{\theta^{[n]}/n!}$$

An alternative form for (6) is $P_{\theta,\alpha}(A_{1,n}=i)=\binom{n-1}{i-1}\frac{(1-\alpha)^{[i-1]}(\theta+\alpha)^{[n-i]}}{(1+\theta)^{[n-1]}}=PD_{n-1;1-\alpha,\theta+\alpha}(i-1)$, that is the Polya Distribution with initial weights $1-\alpha$ and $\theta+\alpha$ for $i-1$ successes in $n-1$ trials.[3] Now (6) is essential to calculate the mean values of the statistical distribution $\mathbf{Z}_n$, so that a lot of interesting results can be obtained without writing explicitly $P_{\theta,\alpha}(\mathbf{Z}_n)$.In general $P(A_{1,n}=i)$ describes the probability that the first extracted cluster has size i, where it is understood that "the first extracted cluster" is the cluster to which belongs the first extracted unit, being all units equally probable to be extracted. It is a very general result that, conditioned to the cluster frequency $\mathbf{Z}_n$,there are $iZ_{i,n}$ units that point to i, hence $P(A_{1,n}=i|\mathbf{Z}_n)=\frac{iZ_{i,n}}{n}$, and by total probability theorem

$$P(A_{1,n}=i)=\sum_{\mathbf{Z}_n}\frac{iZ_{i,n}}{n}P(\mathbf{Z}_n)=\frac{iE(Z_{i,n})}{n} \tag{7}$$

[3] This is not surprising, as the two-parameter Hoppe urn, marginalized to the first extracted color, is just a usual bivariate Polya urn, with initial weights $1-\alpha$ and $\theta+\alpha$ and unitary prize. If in the remainig $n-1$ extractions you find $i-1$ times the first extracted color, its final size is i.

This deals with sampling, that is to say observing sequentially all the population, and revealing step-by-step its composition, given in this case by the cluster frequency $\mathbf{Z}_n$. This holds for any $P(\mathbf{Z}_n)$. But posing $P(\mathbf{Z}_n) = P_{\theta,\alpha}(\mathbf{Z}_n)$, it follows that a sequential observation of all units is ruled by (2), the size of the first observed cluster (the size-biased pick) is ruled by (6), and given (7) we have directly, for $i = 1, \dots, n$,

$$E_{\theta,\alpha}(Z_{i,n}) = \frac{n}{i} PD_{n-1;1-\alpha,\theta+\alpha}(i-1). \tag{8}$$

3 The accommodation process

Now we change scenario at all, and we consider a population of n agents that can choose among a very large number of possible strategies. We are not interested to the name of the strategies, rather we want to study how agents are partitioned into groups, being understood that two units belong to the same cluster iff $X_i = X_j$. The operational frame is that we make all agents choose sequentially, and we suppose that each agent's choice is conditioned by all previous declared ones. Also if the possible strategies are infinite, no more than n can be exploited by our population at the same time, so that it is convenient (specially for future developments) to label the possible choices inventing a system of $g > n$ cells or sites (if we think to firms as clusters of workers, for instance (Axtell 2001, 1999)). Of course these labels have no meaning except that of distinguishing different clusters, and they are posed equiprobable *ceteris paribus*. Then the first agent chooses the jth-site with uniform probability $\frac{1}{g}$. The second has two possibilities: if its choice is the same of the first ("herding"), he accommodates in the same jth-site; if it is different ("pioneering"), he accommodates in some ith-site, $i \neq j$, with probability $\frac{1}{g-1}$, uniform on the still empty sites. We suppose that the choice is probabilistic, and it is influenced by the results of all previous choices. This influence is twofold: the weight of "pioneering" is posed equal to $\theta + K_m\alpha, \theta > 0, 0 \leq \alpha < 1$, that of herding is posed equal to $m - K_m\alpha$, where K_m is the present number of clusters. Given herding, the agent chooses the jth-site proportionally to $N_{j,m} - \alpha$. Given pioneering, he chooses a random site among the $g - K_m$ empty sites. The resulting accommodation probability is

$$P(S_{m+1} = j | N_{1,m}, \dots, N_{g,m}) = \begin{cases} \frac{N_{j,m}-\alpha}{m+\theta} & N_{j,m} > 0 \\ \frac{1}{g-K_m}\frac{\theta+K_m\alpha}{m+\theta}, & N_{j,m} = 0 \end{cases} \tag{9}$$

$K_m = \#\{N_{j,m} > 0, j = 1, \dots, g\}$, and $\sum_{i=1}^{g} N_{i,m} = m$ is the actual size of the (declared) population. The parameter θ is the initial (unconditioned) weight for pioneering, while α is an additive weight due to the influence of each other pioneer in the population, so that $K_m\alpha$ is the contribution to pioneering due to the number of pioneers already present. Now (9) and (2) are quite similar, and we want to discuss their relationship.

The two-parameter Hoppe predictive probability (2) is a description of a sampling sequence in term of temporal labels (the time of appearance of the clusters within the present operation of sampling), and $\mathbf{A}_n$ is the frequency distribution in terms of these time-labels. Instead the accommodation probability (9) is a description of a physical process, and it uses site labels, $\mathbf{N}_n$, that is the frequency distribution in terms of these space-labels, that are usual permanent categories.

The essential difference between the two types of labels is given by the factor $\frac{1}{g-K_m}$ in case of novelty: for temporal labels the name of the new cluster is predetermined, while for sites it must be chosen. In this case one can pass from time labels to site ones fixing the K_n cluster labels independently on anything. If g labels are available, each of the $g(g-1)\dots(g-K_n+1) = \frac{g!}{(g-K_n)!}$ site descriptions are equally likely. It means that

$$P_{\theta,\alpha}(Y_1,\dots,Y_n) = \frac{g!}{(g-K_n)!}P_{\theta,\alpha}(S_1,\dots,S_n) \tag{10}$$

if the two sequences describe the same physical process. Observing that $\mathbf{A}_n$ and $\mathbf{N}_n$ describe the same clusters with different labels, and $\prod_{i=1}^{K_n} A_{i,n}! = \prod_{i=1}^{g} N_{i,n}! = \prod_{j=1}^{n} j!^{Z_{j,n}}$, from (3) we have

$$P_{\theta,\alpha}(\mathbf{Z}_n) = \frac{g!}{\prod_{j=0}^{n} Z_{j,n}!}\frac{n!}{\prod_{i=1}^{g} N_{i,n}!}P_{\theta,\alpha}(S_1,\dots,S_g) \tag{11}$$

where $Z_{0,n} = g - K_n$ is the number of voids sites.

As the clustering structure of the accommodation process is the same of (2), we can conclude that the size of the first created cluster (the size-biased pick) is ruled by (6), the created statistical cluster distribution has probability $P(\mathbf{Z}_n) = P_{\theta,\alpha}(\mathbf{Z}_n)$, and the mean number of clusters of size i is given by (8). In this frame all distribution have an objective meaning: if one constructs an ensemble of independent accommodation processes with the same initial conditions, the frequency in the ensemble of any considered event is roughly proportional to the corresponding probability. $\mathbf{Z}_n$ is "random", or better ruled by $P_{\theta,\alpha}(\mathbf{Z}_n)$, because is generated by a random mechanism (the accommodation process, the choices of agents). These probabilities have finite spreading also in the limit $n \to \infty$.

3.1 The marginal site distribution and a new calculation of $E_{\theta,\alpha}(Z_i)$

It is interesting to marginalize (9) on a fixed site, say the site 1 and calculate $P(S_1 = 1,\dots,S_i = 1, S_{i+1} \neq 1,\dots,S_n \neq 1)$ for $i \geq 1$.

By multiplication theorem we calculate first $P(S_1 = 1,\dots,S_i = 1) = \frac{1}{g}\frac{\theta(1-\alpha)\cdots(1-\alpha+i-1)}{\theta(\theta+1)\cdots(\theta+i-1)}$. The next term is $P(S_{i+1} \neq 1|S_1 = 1,\dots,S_i = 1) = \frac{\theta+\alpha}{\theta+i}$, the following term has probability $\frac{1-\alpha}{\theta+i+1}$ if $X_{i+2} = X_{i+1}$, or $\frac{\theta+2\alpha}{\theta+i+1}$ if $X_{i+2} \neq X_{i+1}$, whose sum is $\frac{\theta+\alpha+1}{\theta+i+1}$. Going on in the same way we get: $P(S_{i+1} \neq 1,\dots,S_n \neq 1|S_1 = 1,\dots,S_i = 1) = \frac{\theta+\alpha}{\theta+i}\frac{\theta+\alpha+1}{\theta+i+1}\cdots\frac{\theta+2+i-1}{\theta+n+1}$.

Finally $P(S_1 = 1,\dots,S_i = 1, S_{i+1} \neq 1,\dots,S_n \neq 1) = \frac{1}{g}\frac{\theta(1-\alpha)^{[i-1]}(\theta+\alpha)^{[n-i]}\dots}{\theta^{[n]}}$, that is $\frac{1}{g}$ times (5). It is apparent that any permutation of the observed values has the same probability (exchangeability of S_i), thus the combinatorial factor is $\binom{n}{i}$,due to the fact that the first variable is no more bound to be $Y_1 = 1$, hence from (8):

$$P_{\theta,\alpha}(N_{1,n} = i) = \frac{1}{g}\frac{n}{i} PD_{n-1;1-\alpha,\theta+\alpha}(i-1) \tag{12}$$

Now

$$E(Z_{i,n}) = \sum_{j=1}^{g} P(N_{j,n} = i) = gP(N_{1,n} = i) \tag{13}$$

being all sites symmetric. Then

$$E_{\theta,\alpha}(Z_{i,n}) = \frac{n}{i} PD_{n-1;1-\alpha,\theta+\alpha}(i-1)$$

The two quite general formulas (7) and (13) bridge the gap between the size probability of the first pick (respectively of a site) and $E(Z_{i,n})$, and then

$$P(N_{1,n} = i) = \frac{n}{gi} P(A_{1,n} = i) \tag{14}$$

is the main link about the two label descriptions. These relationships are useful both in case of sampling ($\mathbf{Z}_n$ is "fixed but unknown") and in the case of accommodating ($\mathbf{Z}_n$ is constructed step-by-step).

The site labels are able to indicate the persistence of clusters over different accommodation processes, as clusters are identified by the site that host them as long as they stay alive. They describe the accommodation of the agents into $g > n$ sites, whose meaning is that to distinguish "spatially" different clusters. The general accommodation term is $P(S_{m+1}|N_{1,m}, \ldots, N_{g,m})$, where $N_{1,m}, \ldots, N_{g,m}$ are the occupation numbers of the g sites, and $\sum_{i=1}^{g} N_{i,m} = m$ is the actual size of the population. Suppose that $P(S_{m+1}|N_{1,m}, \ldots, N_{g,m})$ represents an agent accommodating into the category X_{m+1}: if X_{m+1} is equal to some present category the agent accommodates in the corresponding site; if the category is "new" the agent accommodates randomly in some empty site, that always exists due to $g > n > m$. Note that the site variables (S_i) have the same range $1, \ldots, g$, that does not depend on i, while the the range of each time-ordered $Y_i = 1, \ldots, i$ depends on i. If we had to describe the same physical sequence starting from a void system, (2) is the partition probability function (Hansen and Pitman 2000), while (9) is the site accommodation probability function. The two sequences of frequency numbers $(A_{1,m}, \ldots, A_{k,m})$ and $(N_{i,m} > 0, i = 1, \ldots, g)$ have the same elements, so that the partition vector $\mathbf{Z}_m = (Z_1, \ldots, Z_m)$ is the same. Thus we are able to indicate the persistence of clusters over different accommodation processes, as clusters are identified by the site that host them as long as they stay alive.

4 Probabilistic dynamics

Clustering is a "physical" attitude that rules the behavior of economical agents. In the accommodation process we consider the abstract case of the sequential strategy choice of n agents. Actually the starting point of our analysis will result in some partition of our finite population into groups. This partition is not for ever, as a probabilistic mechanism of destruction-creation is assumed to move agents from cluster to cluster, or to found new ones, with the necessary consequence of the death of some old cluster. Suppose that at the (discrete) time t (say the $t-$day) our population is described by the site occupation vector $\mathbf{N}_n = (N_{1,n}, \ldots, N_{g,n})$.Each site can host a cluster, $\sum_{i=1}^{g} N_{i,n} = n$, the cluster size frequency $\mathbf{Z}_n$ is such that $\sum_{i=1}^{n} i Z_{i,n} = n$, and $\sum_{i=1}^{n} Z_{i,n} = K_n$ is the number of occupied sites, that is the number of represented strategies. The probabilistic dynamics is modelled to simulate the strategy changes of the n agents. The day-after some changing agents are extracted, and they temporarily abandon the system; when this "destruction" ends, the same units re-accommodate one by one into the site system ("creation"). Changes occur if their final site is different from the initial one. In the time interval from t to $t + 1$ the size of the population shrinks as long as selection occurs, and it returns gradually to the original size after all accommodations. Both destruction and creation are understood to be probabilistic, and the evolution of the system depends strongly on these probability functions. Quite generally we will speak of herding (resp: pioneering) whenever an accommodation occurs in an occupied (resp: empty) site. In the present paper we fix the accommodation (creation) probability in the form (9). Further a quite general property of this landscape is that a site represents a category (a strategy) as long as it is occupied, and we suppose that a strategy dies when the corresponding site becomes empty. When (with probability one) the same site is occupied again, it is understood that it represents a new strategy, where "new" is intended (with probability one) with respect of all past history of the system. Here in $P(S_{h+1} = j|N_{1,h}, \ldots, N_{g,h})$ the evidence represents the size of the K_h clusters which partition the h elements that are in front of the accommodating element. The form (9) implies that the accommodation depends only on the size (not on the age) of the present clusters. Hence in the case of a dynamical system temporal labels are meaningless as the identity of the clusters from day to day must be guaranteed by some non temporal label (Garibaldi et al. 2003b, 2005). The index $h+1$ of the accommodation is referred to the element that increases the size of the population from h to $h + 1$. If one wants to conserve at most the accommodation scheme or the previous chapter, one can represent all not selected agents as re-declaring the same strategies of the previous day. If destruction has shrinked the population from n to $n-m$, in general S_{h+1}, $h = n-m, n-m+1, \ldots, n-1$ is conditioned by the cluster distribution deriving by the initial state $N_{1,n-m}, \ldots, N_{g,n-m}$ (the not selected agents, that hold the previous strategy) plus the modifications due to the accommodation of the $h - (n - m)$ agents that have already re-declared.

5 The Ehrenfest–Pitman model

In order to complete the probabilistic engine, we state that the destruction probability results in a random sampling (without replacement) of m agents. This is a generalization of the famous Ehrenfest-urn-scheme (Ehrenfest and Ehrenfest 1907), with the great difference that the creation term is given by (9).

We limit ourselves to the "unary" case $m = 1$. In this simple case destruction transforms $\mathbf{N}_n = (N_{1,n}, \ldots, N_{g,n})$ in some $\mathbf{N}_{n-1} = (N_{1,n-1}, \ldots, N_{g,n-1})$, and the following creation transforms $\mathbf{N}_{n-1}$ in some $\mathbf{N}'_n$. The formal development is to construct the transition matrix $P(\mathbf{N}'_n(t+1)|\mathbf{N}_n(t))$, and thus to show that it is a homogeneous aperiodic Markov chain, whose equilibrium distribution $\pi(\mathbf{N}_n)$ is such that the distribution on the cluster frequency is just $P_{\theta,\alpha}(\mathbf{Z}_n)$, that is the two-parameter-Ewens distribution. The method is exactly the same as in Garibaldi et al. (2003b, 2005). But we follow an alternative method, that is more natural for economic applications.

5.1 The marginal chain of a site

Most of discussions about cluster growth mean to introduce some (logarithmic) random walk, whose role is to mimic the empirical data. In short you consider a cluster in contact with some environment, whose model is never investigated (Gibrat 1931). In our case we can concentrate the attention on a fixed site, and describe what happens in this site supposing that the whole population is ruled by some reasonable probabilistic mechanism. The marginal description of one fixed site is a reduced description of the system. In this case, the above-said dynamics is easy projected on one site exactly. Denoting by $N_{1,n} = i$ the site occupation number (the size), let us consider first the case $i > 0$, that implies the site to be occupied. The size increases by one if an external agent is extracted (with probability $\frac{n-i}{n}$) and it enters the site (with probability $\frac{i-\alpha}{\theta+n-1}$). It decreases by one if an internal agent is extracted (with probability $\frac{i}{n}$), and it joins the external clusters (with probability $\frac{n-i-(K_n-1)\alpha}{\theta+n-1}$) or it is a pioneer (with probability $\frac{\theta+K_n\alpha}{\theta+n-1}$). Both probabilities depend on the number K_n of present clusters, but the sum does not. Hence posing $w(i,j) := P(X_{s+1} = j|X_s = i)$, for unary changes the non vanishing entries of the transition matrix of the marginal chain are the following, for $i > 0$:

$$
\begin{aligned}
w(i,i+1) &= \frac{n-i}{n}\frac{i-\alpha}{\theta+n-1} \\
w(i,i-1) &= \frac{i}{n}\frac{n-i+\theta+\alpha}{\theta+n-1} \\
w(i,i) &= 1 - w(i,i+1) - w(i,i-1)
\end{aligned}
\tag{15}
$$

All external occupied sites are merged in a single one (the thermostat), whose weight is $n - i - (K_n - 1)\alpha$, plus $\theta + K_n\alpha$, that is the weight of the empty sites. Note that the cluster dynamics does not depend on the number of clusters K_n, and thus it is exactly representable. If the site is empty, $i = 0$, using the same method as Garibaldi et al. (2003b, 2005), we introduce the rebirthing term and its complement

$$
\begin{aligned}
w(0,1) &= \frac{1}{g-K_n}\frac{\theta+K_n\alpha}{\theta+n-1} \\
w(0,0) &= 1 - w(0,1)
\end{aligned}
$$

to avoid the state $N_{1,n} = 0$ to be absorbing. Of course $w(0, 1)$ depends on K_n, so that this term is an exogenous random variable. This is not essential for the equilibrium distribution of the cluster size. Starting from a cluster whose size is i we must distinguish the history of the cluster (that terminates when the size reaches $i = 0$) from that of the site, that sooner or later reactivates. Then we obtain an irreducible chain, whose stationary and equilibrium distribution satisfies the detailed balance condition. The equilibrium distribution of the chain (15) is

$$\begin{cases} P(i+1)\dfrac{i+1}{n}\dfrac{n-i-1+\theta+\alpha}{\theta+n-1} = P(i)\dfrac{n-i}{n}\dfrac{i-\alpha}{\theta+n-1}, & i = 1,\ldots,n-1 \\ P(1)w(1,0) = P(0)w(0,1) \end{cases}$$

It is easy to see that $\frac{P(i+1)}{P(i)} = \frac{n-i}{i+1}\frac{i-\alpha}{\theta+\alpha+n-i-1} = \frac{E_{\theta,\alpha}(Z_{i+1,n})}{E_{\theta,\alpha}(Z_{i,n})}$, where $E_{\theta,\alpha}(Z_{i+1,n})$ is just the mean number of clusters (8) in the two-parameter Ewens population, while $P_{\theta,\alpha}(N_{1,n} = i)$ is the marginal of the two-parameter Ewens population. As previously said the term $P(0) = P_{\theta,\alpha}(N_{i,n} = 0)$ depends on g and K_n. If we substitute K_n with $E_{\theta,\alpha}(K_n)$, the chain mimics at best what happens in the fixed site if its history is extracted from the exact joint dynamics of the whole population. But what is essential from the cluster point of view is that $w(0, 1)$ is positive. If for instance $g \gg n$, each fixed site will be empty for a long time. If one chooses a site at random, all sites being equiprobable, the site will be empty with probability $\frac{g-E_{\theta,\alpha}(K_n)}{g}$, and $\frac{E_{\theta,\alpha}(K_n)}{g}$ is the probability of finding a living cluster. To choose a cluster at random, all clusters being equiprobable, and to observe its size is then described by

$$P(N_{1,n} = i|i > 0) = \frac{E_{\theta,\alpha}(Z_{i,n})}{E_{\theta,\alpha}(K_n)}, \quad i = 1,\ldots,n \tag{16}$$

being understood that any observation regards living clusters (non empty sites). The attached code (Mathematica 5) shows how the marginal chain produces an empirical size distribution that tends to (16).

6 Appendix

6.1 Exchangeability

$P(Y_1,\ldots,Y_n) = f(A_{1,n},\ldots,A_{K_n,n})$ says the probability of a temporal partition $Y_1,\ldots,Y_n$ is a function of the sizes of the observed (or created) clusters. It means that all sequences that end in the same $A_{1,n},\ldots,A_{K_n,n}$ are equiprobable. But their number is not $\frac{n!}{\prod_{i=1}^{K_n} A_{i,n}!}$ as usual, because all possible sequence are equiprobable, but the variable are not equidistributed, as the possible sequences are bounded by the temporal meaning of the variables. All sequences start with $Y_1 = 1$, further $Y_i \geq i$, and $Y_i = j$ is possible only if for some $k < i$ it was $Y_k = j - 1$. In any case it is false that $P(Y_1,\ldots,Y_n) = P(Y_{\pi(1)},\ldots,Y_{\pi(n)})$, where by $\pi(i)$ we mean any permutation of $1,\ldots,n$, that is the usual definition of exchangeability. The resulting asymmetry among the first, the second, ..., the kth-cluster in order of appearance is essential in the size-biased sampling. Hence $P(A_{1,n},\ldots,A_{K_n,n})$ is not exchangeable with respect to the names of the clusters. This is the reason why the multiplicity factor of (3) is not apparent (see Hoppe 1987).

On the contrary $P(S_1,\ldots,S_n) = f(N_{1,n},\ldots,N_{g,n})$ makes all possible sequence equiprobable, and they are just $\frac{n!}{\prod_{i=1}^{g} N_{i,n}!}$

Further $P(N_{1,n},\ldots,N_{g,n})$ is exchangeable with respect to the names of the clusters. This simplifies the combinatorial factor that drives from the sequence probability to occupation number probability. Further all clusters are symmetrical with respect to their names, and this simplifies the combinatorial factor that drives from the occupation number probability to the partition vector probability. Hence the multiplicity factor of (11) is clear, as $\frac{n!}{\prod_{i=1}^{g} N_{i,n}!}$ counts the distinct patterns to $\mathbf{N}_n$, while $\frac{g!}{\prod_{j=0}^{n} Z_{j,n}!}$ counts the distinct permutations of the site labels that point to the same $\mathbf{Z}_n$.

6.2 Three different meanings of the frequency spectrum

Let $P(\mathbf{N}'_n, t|\mathbf{N}_n, 0)$ be the probability that the system prepared in the state $\mathbf{N}_n$ at time 0 is in the state $\mathbf{N}'_n$ at time t. If $\lim_{t\to\infty} P(\mathbf{N}'_n, t|\mathbf{N}_n, 0) = \pi(\mathbf{N}'_n)$, independent both on time and on the initial state, then $\pi(\cdot)$ is the equilibrium probability of the dynamical system. As $\mathbf{N}_n$ is embedded into arbitrary labels, we prefer to consider $\mathbf{Z}_n$, being all meanings transferred to it without loss. Then $\pi(\mathbf{Z}_n) = \frac{g!}{(g-\sum Z_{i,n})!Z_{1,n}!\cdots Z_{n,n}!}\pi(\mathbf{N}_n)$ has a lot of different meanings. The "ensemble" meaning is the following: if one prepares a large number of systems in the same initial state, then the number of systems in the state $\mathbf{Z}_n$ at time $t \gg 1$ is roughly proportional to $\pi(\mathbf{Z}_n)$. It is equivalent to suppose that a random selected system from the ensemble has $\pi(\mathbf{Z}_n)$ as sampling distribution. Sampling from the population at time t, that is observing units one-by-one in order to get information on $\mathbf{Z}_n$, is independent on t. If $\pi(\mathbf{Z}_n) = P_{\theta,\alpha}(\mathbf{Z}_n)$, it follows that this sampling distribution is (2).

But given that the chain of the system is ergodic, the invariant distribution is the limit of the visit time relative frequency. This to say that time averages (on a single realization) and ensemble averages are identical as long as equilibrium has been reached. Considering now the equilibrium mean number distribution $\left\{\frac{E_{\theta,\alpha}(Z_{i,n})}{E_{\theta,\alpha}(K_n)}, i = 1,\ldots,n\right\}$, that is normalized if $E_{\theta,\alpha}(K_n) = \sum_{i=1}^{n} E_{\theta,\alpha}(Z_{i,n})$. It is the mean fraction of clusters of size i in the ensemble at any time $t \gg 1$. But it has also a time average meaning: given that $f_{\theta,\alpha}(i) = \frac{E_{\theta,\alpha}(Z_{i,n})}{E_{\theta,\alpha}(K_n)}$ is equal to the equilibrium marginal distribution of any site when it is occupied, by ergodic theorem it is equal to the fraction of times a cluster has size i conditioned to be non empty. These consideration are essential in order to identify the probabilistic objects to be compared with empirical (respectively simulated) data. In the case of simulations, also for very small populations ($n = 100$) it is quite impossible to test $P_{\theta,\alpha}(\mathbf{Z}_n)$ comparing it with the visit time relative frequency, because the domain of $\mathbf{Z}_n$ has $1.99 \cdot 10^8$ elements (in the realistic landscape of Axtell (2001) $n \simeq 10^8$, the domain of $\mathbf{Z}_n$ has $\simeq 1.76 \cdot 10^{11131}$ elements...). What one can do is to calculate from the observed sequence $\{\mathbf{Z}_n(s), s = 0, 1, \ldots, t\}$ the time average of the number of clusters of size i, that is $\frac{\sum_{s=0}^{t-1} Z_{i.n}(s)}{t}$, and to compare it with $E_{\theta,\alpha}(Z_{i.n})$. Alternatively one can concentrate the attention on a fixed site (say 1), and calculate the time frequency of the size $i > 0$, that is $\frac{\sum_{s=0}^{t-1} 1[N_{1.n}(s)=i]}{t}$, the sum being extended on all times when the site is not empty, and to compare it with $f_{\theta,\alpha}(i)$.

In any case it is evident the main role of the function $E_{\theta,\alpha}(Z_{i.n})$. In statistics the function $E_{\theta,\alpha}(Z_{i.n})$ is considered only in the continuum limit $n \to \infty$, where it becomes a density, and it is called "frequency spectrum".

Starting from (6), $P_{\theta,\alpha}(A_{1,n} = i) = \frac{\theta(1-\alpha)^{[i-1]}}{n(i-1)!} \frac{(\theta+\alpha)^{[n-i]}/(n-i)!}{\theta^{[n]}/n!} \to \frac{1}{n} \frac{\Gamma(\theta+1)}{\Gamma(1-\alpha)\Gamma(\theta+\alpha)} \left(\frac{i}{n}\right)^{-\alpha} \left(1 - \frac{i}{n}\right)^{\theta+\alpha-1}$, hence $E_{\theta,\alpha}(Z_{i.n}) \approx \frac{1}{i} \frac{\Gamma(\theta+1)}{\Gamma(1-\alpha)\Gamma(\theta+\alpha)} \left(\frac{i}{n}\right)^{-\alpha} \left(1-\frac{i}{n}\right)^{\theta+\alpha-1} = K \frac{n^{\alpha}}{i^{\alpha+1}} \left(1-\frac{i}{n}\right)^{\theta+\alpha-1}$.

If we pass to the variable $x = \frac{i}{n}$, we get

$$f_{\theta,\alpha}(x) = \frac{1}{x} Beta(x; 1-\alpha, \theta+\alpha),$$

that is the density of clusters whose mean fractional size is x. Let's remark that this is a mean value, not an actual frequency. This density is strange as it diverges for $x \to 0$, but it is quite reasonable if one considers that for any finite n the mean number of clusters $E_{\theta,\alpha}(K_n) = \sum_{i=1}^{n} E_{\theta,\alpha}(Z_{i,n})$ is finite, but $E_{\theta,\alpha}(K_n) \to \infty$ for $n \to \infty$. In other words, for any $n \gg 1$ the domain of $f(x)$ is $\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\}$, so that this divergence is only formal. In the limit $n, \theta \gg 1$, given that $0 \leq \alpha < 1$, we can substitute $\left(1 - \frac{i}{n}\right)^{\theta+\alpha-1}$ with $Exp\left[-i\frac{\theta}{n}\right]$, so that the main probabilistic object of the analysis has the form

$$f_{\theta,\alpha}(i) \approx K \frac{Exp\left[-i\frac{\theta}{n}\right]}{i^{\alpha+1}}, \tag{17}$$

that is the (continuous) mean size approaches the limiting form $f_{\theta,\alpha}(y) = \frac{1}{y} Gamma(y; 1-\alpha, \theta/n)$.

We have discussed this form in Garibaldi et al. (2003b, 2005). We have an extension of the Ewens formula for small values of i, where we have all a family of power laws depending on $0 \leq \alpha < 1$. In any case we have an exponential cut-off.

6.3 Mean number of clusters

A typical example of the usefulness of the finite approach to the 2-par-Ewens distribution is the exact calculation of the equilibrium mean number of clusters as a function of α, θ, n. Let us consider the balance between death and birth. If the system is described by $\{Z_{i,n}\}$, considering unary changes, a cluster dies if and only if the extracted unit belongs to a singleton, hence the probability of a death is $\frac{Z_{1,n}}{n}$. A new cluster is born if the extracted unit is a pioneer, and given that the pioneering probability depends on the actual number of present cluster we must distinguish the case in which the accommodating unit comes from a singleton, where the actual number of clusters is $K_{n-1} = K_n - 1$, from the alternative one, where the actual number of clusters is

$K_{n-1} = K_n$. Hence

$$\begin{cases} P(death|\mathbf{Z}_n) = \frac{Z_{1,n}}{n} \\ P(birth|\mathbf{Z}_n) = \frac{Z_{1,n}}{n} \frac{\theta+(K_n-1)\alpha}{n-1+\theta} + \left(1 - \frac{Z_{1,n}}{n}\right) \frac{\theta+K_n\alpha}{n-1+\theta} \end{cases}$$

Now

$$\begin{aligned} &\frac{Z_{1,n}}{n}\frac{\theta+(K_n-1)\alpha}{n-1+\theta} + \left(1-\frac{Z_{1,n}}{n}\right)\frac{\theta+K_n\alpha}{n-1+\theta} \\ &= \frac{1}{n(n-1+\theta)}\big(Z_{1,n}(\theta+(K_n-1)\alpha-\theta+K_n\alpha)+n(\theta+K_n\alpha)\big) \\ &= \frac{1}{n(n-1+\theta)}\big(Z_{1,n}(\theta+(K_n-1)\alpha-\theta+K_n\alpha)+n(\theta+K_n\alpha)\big) \\ &= \frac{1}{n(n-1+\theta)}\big(-Z_{1,n}\alpha+n(\theta+K_n\alpha)\big) \end{aligned}$$

The unconditioned death (and birth) probabilities are:

$$\begin{cases} P(death) = \sum_{\mathbf{Z}_n} P(death|\mathbf{Z}_n)P(\mathbf{Z}_n) = \frac{E(Z_{1,n})}{n} \\ P(birth) = \frac{1}{n(n-1+\theta)}\big(-E(Z_{1,n})\alpha + n(\theta+E(K_n)\alpha)\big) \end{cases}$$

and the balance implies $E(Z_{1,n}) = -E(Z_{1,n})\frac{\alpha}{n-1+\theta} + \frac{\theta+E(K_n)\alpha}{n-1+\theta}$, that gives

$$E(K_n) = E(Z_{1,n})\frac{n-1+\theta+\alpha}{n\alpha} - \frac{\theta}{\alpha}. \tag{18}$$

Introducing in (18) the value $E(Z_{1,n}) = \theta n\frac{(\theta+\alpha)^{[n-1]}}{\theta^{[n]}} = n\frac{(\theta+\alpha)^{[n-1]}}{(\theta+1)^{[n-1]}}$ from (8) we find

$$E(K_n) = \frac{\theta}{\alpha}\left(\frac{(\theta+\alpha)^{[n]}}{\theta^{[n]}} - 1\right) \tag{19}$$

The fact that $E(K_n)$ from (19) is exactly equal to $\sum_{i=1}^n E(Z_{i,n})$ from (8) is easy to check numerically. Further it is easy to show that for $\alpha \to 0$, $\frac{(\theta+\alpha)^{[n]}}{\theta^{[n]}} = \frac{(\theta+\alpha)}{\theta}\frac{(\theta+\alpha+1)}{\theta+1}\ldots = \left(1+\frac{\alpha}{\theta}\right)\left(1+\frac{\alpha}{\theta+1}\right)\ldots\left(1+\frac{\alpha}{\theta+n-1}\right) = 1+\sum_{i=1}^n \frac{\alpha}{\theta+i-1}+O(\alpha^2)$. Hence $E(K_n) = \frac{\theta}{\alpha}\sum_{i=1}^n \frac{\alpha}{\theta+i-1} + O(\alpha^2) = \sum_{i=1}^n \frac{\theta}{\theta+i-1} + O(\alpha^2)$, where $\sum_{i=1}^n \frac{\theta}{\theta+i-1} = E_{0,\theta}(K_n)$ is the well-known mean number of clusters in the Ewens case.

In the opposite limit $\alpha \to 1$, we have $E(K_n) \to \theta\left(\frac{(\theta+1)^{[n]}}{\theta^{[n]}} - 1\right) = \theta\left(\frac{\theta+n}{\theta} - 1\right) = n$, as the population is composed of all singletons with probability one.

For large n, (19) can be approximated by $\frac{\theta}{\alpha}\left(\frac{\Gamma(\theta)}{\Gamma(\theta+\alpha)}n^\alpha - 1\right) \simeq \frac{\Gamma(\theta+1)}{\Gamma(\theta+\alpha)}\frac{n^\alpha}{\alpha}$.

References

Aoki M (1996) New approaches to macroeconomic modeling: evolutionary stochastic dynamics, multiple equilibria, and externalities as field effects. Cambridge University Press, New York

Aoki M (2000) Cluster size distributions of economic agents of many types in a market. J Math Anal Appl 249:32–52

Aoki M (2002) Modeling aggregate behavior and fluctuations in economics. Cambridge University Press, Cambridge, UK

Aoki M, Yoshikawa H (2007) Reconstructing macroeconomics: a perspective form statistical physics and combinatorial stochastic processes. Cambridge University Press, New York

Axtell RL (1999) The emergence of firms in a population of agents: local increasing returns . . . and power law size distribution, Working paper no 3, Brooking Institutions

Axtell RL (2001) Zipf distribution of U.S. firm sizes. Science 293:1818–1820

Costantini D, Garibaldi U (2004) The Ehrenfest model: from model to theory. Synthese 1398:107–142

Ehrenfest P, Ehrenfest T (1907) Ueber zwei bekannte Einwaende gegen Boltzmanns H-Theorem. Phys Zeitschrift 8(1):311–316

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3:87–112

Garibaldi U, Penco MA, Viarengo P (2003a) An exact physical approach for market participation models. In: Cowan R, Jonard N (eds) Heterogeneous agents, interactions and economic performance, Lecture Notes in Economics and Mathematical Systems, vol 521, Springer, Heidelberg

Garibaldi U, Costantini D, Viarengo P (2003b) A finite characterization of Ewens sampling formula. Adv Complex Syst 7(2):265–284

Garibaldi U, Costantini D, Viarengo P (2005) A finitary characterization of Ewens sampling formula. In: Lux T, Reitz S, Samanidou E (eds) Nonlinear dynamics and eterogeneous interacting agents. Lecture Notes in Economics and Mathematical Systems Series, pp 221–236, Springer, Berlin

Gibrat R (1931) Les inegalities economique:. . . la loi dell'effet proportionnel. Libraire du Recueil Sirey, Paris

Hansen B, Pitman J (2000) Prediction rules for exchangeable sequences related to species sampling. Statist Prob Lett 46:251–256

Hoppe FM (1987) The sampling theory of neutral alleles and an urn model in population genetics. J Math Biol 25(2):123–159

Karlin S, Taylor HM (1981) A second course in stochastic processes. Academic, New York

Kingman JFC (1978) The representation of partition structures. J Lond Math Soc 18:374–380

Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law, arXiv:cond-mat/0412004 v2

Pitman J (1992) The two-parameter generalization of Ewens' random partition structure, Tech Rep No 345, Department of Statistics, Berkeley University

Pitman J, Yor M (1996) The two-parameter Poisson-Dirichlet distribution derived by a stable subordinator. Ann Prob 28:855–900

Zabell S (1992) Predicting the unpredictable. Synthese 90:205–232

Zabell S (1996) The continuum of inductive methods revisited. In: Earman J, Norton J (eds) The cosmos of science. University Pittsburgh Press University Konstanz