Applications of Exchangeable Random Partitions to Economic Modeling

Masanao Aoki
Department of Economics
University of California, Los Angeles
Fax Number 310-825-9528, e-mail aoki@econ.ucla.edu
Feb., 2003

Abstract

This paper\textsuperscript{1} sketches methods to model a large number of interacting agents that are based on combinatorial analysis of random decomposable systems, and points to some applications of the methods. This paper also discusses a possibility that hitherto unknown types of agents may appear at some future time. Results of the one- and two-parameter inductive methods of Ewens, Pitman and Zabell (1992, 1997) are mentioned in this connection.

More specifically, we use the notion of exchangeable random partitions of a finite set to produce expressions for the probabilities of entries by new or known types, conditional on the observed data. Then application of Ewens distribution for the sizes of clusters is introduced to examine market behavior, especially when a few types of agents are dominant.

We also suggest that the approaches outlined in this paper are also relevant to agent-based simulations because holding times can be used to randomly select agents that "act" first.

\textsuperscript{1}Supported in part by the UCLA Academic Senate grant
1 Introduction

Economists often face problems of modeling collective behavior of a large number of interacting agents, possibly of several different types. This paper discusses methods that are useful in such context. We indicate how the methods may explain diverse phenomena such as equilibrium size distributions of clusters, that is, of subgroups formed by agents, market shares by different types of goods, and emergence of macroeconomic regularities as the number of agents increases towards infinity, among other things.

We explicitly assume that there are several types of agents in our models, the number of which may not be known in advance, and that agents of new types may enter the models at any time. We cannot assume in advance that we know all of them because new rules or new goods may be invented in the future. This is the so-called problem of unanticipated knowledge in the sense of Zabell (1992). In biology this problem is known as the sampling of species problem. In probability and statistics it is called laws of sucession, that is, how to specify the conditional probability that the next sample is a new type, given available sets of observation up to now. See Zabell (1982). In addition, agents may change their minds at any time about the decisions or behavioral rules they use. In other words, agents may change their types any time.

We view large economic structures as decomposable random combinatorial structures. Of the many possible structures due to many possible configurations which a large number of components may assume, we wish to deal with "typical" structures of large sizes. By typical we mean structures which have high probabilities of being realized, or observed from the set of all possible structures, when chosen at random in some sense from the set of such candidate structures.

This is the main reason for borrowing or adapting tools and concepts from combinatorial stochastic processes, and population genetics literature. Some methods are borrowed from the field of decomposable random combinatorial analysis. They are useful in modeling economic structures composed of a large number of possibly heterogeneous agents, components or

\[\text{\footnotesize 2 Zabell describes the problem faced by statisticians in classifying samples of insects, say, collected in unexplored regions, since they may contain new species. The naive Bayesian approach is not applicable. See, however, Antoniak (1969) on non-parametric Bayesian approach. He obtained the same distribution as the Ewens sampling formula, Ewens (1972).} \]

\[\text{\footnotesize 3 There is no lock-step behavior by agents.} \]
basic units. This chapter indicates some of their potential applications in 
macroeconomics and finance.

We interpret the word "type" broadly. Types may refer to choices or 
decisions made by agents, some characteristics related to risk taking, behav-
ioral patterns or rules that are used to partition the set of agents, and so 
on. In other words, the word refers to any characteristics which are used to 
distinguish one subgroup of agents from others. We assume that the number 
of types are at most countable.

Many of the methods described in this chapter are not in the tool kit of 
traditionally trained economists or econometricians who mostly rely on the 
notion of representative agents\(^4\), but we have found them to be useful for un-
derstanding macroeconomic or financial phenomena from our new perspec-
tives. Stochastic combinatorial tools are used to show how agents form clus-
ters, and jump Markov processes are used to model how the clusters evolve 
over time through interaction among agents of several types. To describe 
dynamic phenomena, master equations (backward Chapman-Kolmogorov 
equations) are used to describe probability distributions of states of models 
under discussion. We use a new notion of states, called partition vector, 
which is more appropriate in dealing with exchangeable or delabelled agents 
and category, i.e., type indices.

Some of the questions we examine are: How do we describe the process 
by which agents form clusters, that is, subgroups in modeling a collection 
of interacting agents? What are the stationary distributions of sizes of 
fractions of agents of different types? What are the market shares of the 
largest cluster, two largest clusters, and so on? Distributions of cluster 
 sizes matter, because a few of the larger clusters, if and when formed, may 
approximately determine the market excess demands for whatever goods in 
the markets, and the nature and magnitudes of fluctuations.

The methods mentioned in this paper have diverse origins. To discuss 
clusters and entries by agents of new types, such as new goods, new busi-
ness models, new (sub)optimization procedures, and so on, we borrow from 
the literature of population genetics such as Ewens (1972, 1990), Watterson 
(1976), and Watterson and Guess (1977), and from statistics and stochas-
tic processes such as Kingman (1978a, b), Arratia and Tavaré (1992), and 

The Ewens sampling formula is an example of one-parameter inductive 
model. It is specified by a single parameter \(\theta\), which controls the rate of

\(^4\)See Kirman (1992) for his criticism of this notion, for example
entries of new types, and correlations among agents of different types, hence the emergence of clusters of various sizes. We also describe its two-parameter extension by Pitman (1992) and Zabell (1997), which is specified by two parameters, $\alpha$ and $\theta$ to be introduced later.

We use partition vectors as state vectors, and exchangeable partitions induced by agents of different types to examine our models. We describe distributions of cluster sizes governed by the residual allocation process, the so-called GEM distribution in some cases, and the one- and two-parameter version of the Poisson-Dirichlet distribution for the order statistics of the shares by types. We also point out that behavior of the moments of the shares may be non-self-averaging phenomena.\footnote{See Derrida (1997) or Sonette (2000, Sec.16.3) on non-self-averaging shares or weights generated by residual allocation or stick-breaking processes.}

As examples of economic and financial applications of the methods mentioned here, we discuss models of business cycles, the distributions of market shares, behavior of rates of returns, and volatilities of returns: In Aoki (2002b, c) power laws of returns of financial assets, and volatility switching are discussed. We briefly compare the approach of the traditional economists in allocating capital stock between two sectors as formulated by Dixit (1989), and our alternative modeling procedure which drops the assumption of instantaneous equalization of marginal products of factors of production. Our analysis and simulation are in terms of continuous-time Markov chains with a large but finite number of interacting agents, as in Aoki (2002a, Chapt.8) for example. A new approach to the Diamond search model from our perspective is in Aoki and Shirai (2001). A new approach to modeling growth and fluctuations (business cycles) is in Aoki and Yoshikawa (2002). Some policy analysis in our framework is illustrated in Aoki, Yoshikawa, and Shimizu (2002).

2 New Concepts

We collect here several concepts mentioned in the introduction. They are not in the traditional economic literature but are deemed to have useful economic applications.
Partition vector

When it is known that there are $n$ agents of $K$ distinct types, state vector is commonly defined as $n = (n_1, n_2, \ldots, n_K)$ where $n_i$ is the number of agents of type $i$, $i = 1, 2, \ldots, n$.

This choice seems natural, and appears to be satisfactory. There is, however, another choice of state which is better when identities of agents of various types are not the issue. In some cases, only the numbers of agents of different types may matter. Labels we assign to agents may be merely for convenience of reference. Permuting these labels often leaves nothing of substance changed. For example, agents may be labelled in the order we sample or examine them, or in the order they enter the market, but there may be no essential meaning or substance to the labels. They are used merely for convenience of referring to them. In such cases permuting labels assigned to agents cause no substantial changes in our conclusions about the models. Then, agents are called delabelled, or exchangeable in the technical sense defined in probability literature.\(^6\) For a collection of exchangeable agents their joint probability is invariant to permutations of indices assigned to agents in order to refer to them.

When agents are exchangeable agents, labels of types, that is, of categories may also be exchangeable. Agents are partitioned by types into distinct clusters. Labels of the clusters may again be for convenience. Category indices may again be for mere convenience of reference with no substance. If categories are delabelled, then the probability is also invariant with respect to permutations of category indices. This is the notion of exchangeable partitions.

State of a population is described by the (unordered) set of type-frequencies i.e., fractions or proportions of different types without stating which frequency belongs to which type. In the context of economic modeling, this way of description does not require model builders to know in advance how many or what types of agents are in the population. It is merely necessary to recognize that there are $K_n$ distinct types in his sample of size $n$, and that there are $a_j$ types with $j$ agents or goods in the sample.

The vector $\mathbf{a}$ with these components is called partition vector by Zabell (1992), and we adopt this name. Note that $\sum_j a_j = K_n$, and $\sum j a_j = n$. The first relation counts the number of occupied boxes, and the second total number of agents.\(^7\) Partition vector is the right notion to discuss models

\(^6\)See Feller (1968).

\(^7\)In the occupancy problem unlabelled or indistinguishable balls are allocated among
with delabelled agents and delabelled categories. The same concept is known under different names in Kingman (1980), and Sachkov (1996).

Exchangeable random sequences and partitions

We formalize next what is stated at the end of the previous section: Let $X_1, \ldots, X_n, \ldots$ be an infinite sequence of random variables taking on any of a finite number of values, say $1, 2, \ldots, k$. The subscripts on $X$ are time index, or the order in which samples are taken or agents enter the system.

The sequence is said to be exchangeable if for every $n$, the cylinder set probabilities

$$\Pr(X_1 = j_1, \ldots, X_n = j_n) = \Pr(j_1, j_2, \ldots, j_n)$$

are invariant under all possible permutations of the time index. Two sequences have the same probability if one is a rearrangement of the other, or the probability is the function of the frequency vector, $n = (n_1, n_2, \ldots, n_k)$. The observed frequency counts, $n_j = n_j(X_1, X_2, \ldots, X_n)$ are sufficient statistics for the sequence in the sense that probabilities conditional on the frequency counts depend only on the frequency vector

$$\Pr(X_1, X_2, \ldots, X_n|n) = \frac{n_1!n_2! \cdots n_k!}{n!}.$$

A partition of a finite set $F$ into $K$ blocks is an unordered collection of non-empty, disjoint sets $\{A_1, \ldots, A_K\}$ whose union is $F$. We follow Zabell (1992) and Pitman (2002).

Economic agents of the same type do not have individuality. Lables we give to agents or to goods are merely for convenience of referring to them. We simply have the first kind of agent, then at some later time a different kind of agents, and so on. The relevant information is the partition of $[n] = \{1, 2, \ldots, n\}$ by a sample of $n$ agents. To be definite, we may use a convention that the blocks of partitions are listed in the order of appearance, that is, by the least elements of the blocks. For example, such partitions can be constructed as follows: The first type occurs at some sets of points

$$A_1 := \{1, t^2_1, \ldots; 1 < t^2_1 < t^3_1 < \cdots\}$$

where $t^1_1$ is necessarily 1,

$$A_2 := \{t^1_2, t^2_2, \ldots; t^2_1 < t^2_2 < \cdots,\}$$

unlabelled or indistinguishable boxes.
and so on. Let $X_i, i = 1, 2, \ldots, n$ be random variables with values on $[k]$. These $X$s are grouped into subsets of $k$ or less and induce a partition of $[n]$.

Any partition of $[n]$ defines a composition of $n$ by using the size of the sets $A_i$, $n_i = |A_i|$, $n = n_1 + n_2 + \cdots + n_K$.

A partition $\Pi_n$ of $[n]$ is exchangeable if its distribution is invariant with respect to permutations, i.e.,

$$\Pr(\Pi_n = \{A_1, A_2, \ldots, A_k\}) = p(|A_1|, \ldots, |A_k|),$$

where $p(\cdots)$ is some symmetric function of the components.

Consider a random composition of $n$ such that given $K_n = k$, $n = n_1 + n_2 + \cdots + n_k$, and the random variables $n_i, i = 1, \ldots, k$ are exchangeable. The corresponding random partition $\Pi_n$ of $[n]$ is constructed by first picking at random uniformly one of $n!/n_1! \cdots n_k!$ of ordered partitions $(A_1, A_2, \ldots, A_k)$ of $[n]$ with $n_i = |A_i|$, for each $1 \leq i \leq k$. Then, let $\Pi_n$ be the unordered partition $\{A_1, A_2, \ldots, A_k\}$. This is an exchangeable partition of $[n]$ with exchangeable partition probability function $p(n_1, n_2, \ldots, n_k)$, where $p(\cdots)$ is some symmetric function of the components.

Exchangeable random partition is then

$$\Pr(n_1, n_2, \ldots, n_k) = \frac{n!}{n_1!n_2!\cdots n_k!} \frac{1}{k!} p(n_1, n_2, \ldots, n_k).$$

The probability of exchangeable partition is a function of partition vector $a$. Two partitions with the same vector $a$ are equiprobable when the partitions are exchangeable. Sequences associated with exchangeable random partitions are exchangeable sequences.

**Limit behavior of large fractions: Poisson-Dirichlet distributions**

Suppose that a large number of agents interact in a market where each agent uses one of $K$ available trading rules, say, where $K$ is large. Then the set of agents is partitioned at most into $K$ clusters. The number of clusters depends crucially on the correlations among agents which affects the probability that two randomly chosen agents in the market are of the same type, e.g., using the same trading rule. Now, arrange the sizes of clusters in non-increasing order as $n_{(1)} \geq n_{(2)} \geq \cdots$. When agents are highly correlated, a small number of large clusters tend to form. When correlations among agents are small, many smaller clusters are likely to emerge, as we later
show. In high correlation cases, we know from Watterson and Guess (1977) that the sum of the sizes of the first two largest clusters, \( n_{(1)} + n_{(2)} \), alone in some cases account for the majority, 70 per cent say, of the total number of participants. This fact is useful in characterizing aggregate behavior when it happens. See Aoki (2002b) for an example in asset markets which has this feature.

Order statistics of the fractions have a well-defined limit distribution, called the Poisson-Dirichlet distribution, Kingman (1978), as the number of agents go to infinity. Pitman (1996) generalized it to a two-parameter version, to be discussed in Section 4.

**Residual allocation model, and size-biased sampling**

Next we introduce the notion of size-biased permutation or sampling. Given a sequence \( \{P_n\} \), generate another sequence \( \{\tilde{P}_n, n = 1, 2, \ldots\} \) as follows.

\[
\Pr(\tilde{P}_1 = P_n | P_1, P_2, \ldots) = P_n,
\]

and

\[
\Pr(\tilde{P}_{j+1} = P_n | \tilde{P}_1, \ldots, \tilde{P}_j, P_1, P_2, \ldots) = \frac{P_n}{1 - \tilde{P}_1 - \ldots - \tilde{P}_j},
\]

if \( P_n \) is not one of \( \tilde{P}_k, k = 1, \ldots j \). This scheme is known as size-biased sampling, and is described in Kingman (1992) or Carlton (1999), among several others.

Now, if the sequences \( \{\tilde{P}_n\} \) has the same distribution as the original one, \( \{P_n\} \), then, the original sequence is said to be invariant under size-biased permutation or sampling.

We next state some facts from the statistics literature. Let a parameter \( \alpha \) be in \([0, 1]\), and \( \theta + \alpha \) be positive. Let \( V_i, i = 1, 2, \ldots \) be a sequence of independent random variables, distributed as \( Be(1-\alpha, \theta + i\alpha) \), where \( Be(a, b) \) is the Beta distribution. Define \( Q_1 = V_1 \), and \( Q_k = V_k \prod_{j=1}^{k-1}(1 - V_j), k = 2, 3, \ldots \). This process of generating a sequence of random variables \( Q_k \)'s is called residual allocation process, or random alm problem in Halmos (1944). It is also known as the stick-breaking process. With \( \alpha = 0 \), the distribution is known as the GEM distribution named for Griffiths, Engen and McCloskey by Ewens. See also Johnson, Kotz, and Balakrishnan (1997). The former is denoted by \( GEM(\alpha, \theta) \), and the latter by \( GEM(0, \theta) \). Given \( \{P_n\} \) which has a GEM \((\alpha, \theta)\) distribution, the non-increasing ranked sequence \( P_{(1)} \geq P_{(2)} \geq \cdots \) is said to have a Poisson-Dirichlet\((\alpha, \theta)\) distribution.
With $\alpha = 0$, the random variables $V$s are $\text{Beta}(1, \theta)$, we have Poisson-Dirichlet distribution $PD(0, \theta)$ as the distribution for the non-increasing order statistics.

We have the following

**Theorem (McCloskey 1965)** When $\{P_n\}$ is the Poisson-Dirichlet distribution, PD$(0, \theta)$, then, the size-biased permutaion has GEM$(0, \theta)$ distribution.

**Theorem (Pitman 1996)** Let $\{P_n\}$ be such that $P_n > 0$ a.s. for all $n$, and they sum to one. Let $\{P_n\}$ be a residual allocation process generated with independent $V_i$ as above. Then $\{P_n\}$ is invariant with respect to size-biased permutation if and only if $\{P_n\}$ has a GEM$(\alpha, \theta)$ distribution.

**Non-self-averaging behavior**

A model is said to be non-self-averaging when a random variable $X_N$, associated with the model which is an extensive quantity, where $N$ is a parameter to indicate the "size" of a model, and $X_N$ is a random variable with sample dependent properties. When the expression $\langle X_N^2 \rangle - \langle X_N \rangle^2$, where the bracket indicates average, which is taken with respect to sample realization, does not go to zero as $N$ tends to infinity, we say $X_N$ has non-self-averaging property. See Sornette (2000, Sec. 16.3). Derrida (1997) describes the same phenomena for some sequence of weights, or fractions $w_i$, $i = 1, \ldots, N$, with sum 1. Let $Y_k$ be the $k$th moment, $Y_k = \sum_i w_i^k$. If $\langle Y_k^2 \rangle - \langle Y_k \rangle^2$ does not vanish as the number $N$ goes to infinity, $Y_k$ does not have self-averaging property.

**3 Dynamics of clustering processes**

Recall that agents and goods are classified into clusters or subsets by associating types with strategies or choices of agents.

As agents interact, new clusters may form or some existing clusters may break up into smaller ones. The transitions of these processes are captured by specifying how the partition vector $a$ is transformed over a small time intervals. For example, transition rate specification

$$w(a, a + e_i) = \lambda(n),$$

where $e_i$ is a vector with the only non-zero component is at the $i$th component, refers to an event that a single agent enters the market without joining
any existing cluster, while

\[ w(a, a + e_{j+1} - e_j) = ja_j\lambda(n) \]

specifies the event that one agent joins a cluster of size \( j \), thereby increasing the number of clusters of size \( j + 1 \) by one, and reducing that of size \( j \) by one. The right-hand side specifies that the rate of this event is proportional to some constant \( \lambda(n) \), possibly depending on the model size \( n \), and \( ja_j \) which counts the total number of agents in the clusters of size \( j \). There are many other possibilities, of course.

With this much explained, our model building procedures may now be summarized as follows: We start with a collection of a large, but a finite number of microeconomic agents in some economic or financial context. We first select state space and specify a set of transition rates on it to model agent interactions stochastically. Agents may be households, firms, or sectors of economy depending on the context of models. Transition rates we specify are usually state-dependent, that is functions of states, to model endogenously generated aggregate effects, called field effects in Aoki (1996). They represent effects on individual agents of aggregate behavior of all the agents in the model, such as total outputs, crowding, fashion, group pressures, and so on. In addition, individual agents must evaluate consequences of their individual choices subject to uncertainty or imperfect information that affect evaluations of discounted present values of alternative choices.

We describe by master equation the dynamics of the joint probability of the components of a state vector for the model which incorporates specified transition rates. Stationary or nonstationary solutions of the master equations are then examined to deduce model aggregate dynamic behavior. In models which focus on the decomposable random combinatorial aspects, distributions of a few of the largest order statistics of the cluster size distributions are examined to draw economic consequenc

In open models with growth, we also need to evaluate conditional probabilities of the next entrant to the model in assessing model behavior. Given a sequence of exchangeable random variables, \( X_j, j = 1, 2, \ldots \), taking values on \([k]\) de Finetti theorem says that

\[
Pr(X_1 = j_1, X_2 = j_2, \ldots, X_n = j_n) = \int p_1^{n_1}p_2^{n_2}\cdots p_k^{n_k}d\mu(p_1, p_2, \ldots, p_k),
\]

over the simplex \( \Delta_k \) of \( p_j \)'s which sum to one.
Once the prior $d\mu$ is implicitly or explicitly specified, it is immediate that
\[
\Pr(X_{n+1} = j_i | X_1, X_2, \ldots, X_n) = \Pr(X_{n+1} = j_i | \mathbf{n}).
\]
Such a conditional probability is known as a rule of succession.

Johnson’s sufficiency postulate stipulates that the conditional probability that the next agent which enters is of type $i$, given the current state vector, is a function of the existing number of agents of type $i$ and that of the total number of agents in the model. The rule of succession simplifies to
\[
\Pr(X_{n+1} = i | \mathbf{n}) = f(n_i, n).
\]
If $X_1, X_2, \ldots$ is an exchangeable sequence satisfying the sufficiency postulate, and $k \geq 3$, then assuming that the relevant conditional probabilities exist
\[
\Pr(X_{n+1} = i | \mathbf{n}) = \frac{n_i + \alpha}{n + k\alpha}.
\]
See Zabell (1982).

How different are the estimates of probability of new types with the Ewens distribution and multinomial distribution? Ewens (1996) has some numerical examples which show that they are quite different. With the multinomial approach, $a_1$ (the number of singleton) is critical. With the Ewens formula, only the total number of types, $\sum a_j$, is relevant.

In modeling industrial sectors, $n_i$ may refer to the number of agents of type $i$, or to the size of the "production line", that is, a measure of capacity utilization by firm producing typ $i$ good. Zabell (1982) proved that under the assumption of exchangeable partitions the functional form of $f$ is specified by
\[
 f(n_i, n) = \frac{n_i}{n + \theta}, \quad (1)
\]
with some positive scalar parameter $\theta$. This may be interpreted as the limit of $\alpha$ going to zero, while $K$ going to infinity in such a way that the product $K\alpha$ converges to $\theta$. Then, the entry rate of a new type is given by $\theta/(n + \theta)$. More generally, with $K$ types, it is of the form
\[
 w(n, n + e_k) = \frac{\alpha + n_k}{K\alpha + n}, \quad (2)
\]
which reduces to (1) in the limit of $\alpha$ going to zero, and $K$ to infinity while their product approaches $\theta$, and
\[
 w(n, n - e_j) = \frac{n_j}{n}.
\]
See Costantini (2000), Costantini and Garibaldi (1979), or Zabell (1982) for circumstances under which these transition rates arise. See Aoki and Yoshikawa (2001) and Aoki (2002, Sec.8.6) for an application of this type of transition rates in models of economy or sectors of economy.

4 Ewens distribution

In this section we describe the equilibrium distribution for sizes of clusters, originally developed in the population genetics literature. This distribution describes that of the cluster sizes of exchangeable random partitions. Later, we discuss another equilibrium distribution based on the Gibbs distribution that are used to construct some state-dependent transition rates. The former has been applied to model asset market and return behavior in Aoki (2000 b, 2002 b, c). The latter has been applied to describe business cycles in Aoki (1998), and to model technical diffusion process in industry in Aoki (1996, Sec.5.10), but equally applicable in many other situations where agents face binary choices.

Here we follow Aoki (1996, 1998, 2000 a,b, 2002 a, b, c) and sketch the basic ingredients for our modeling procedure without too much detail. The reader is asked to consult the cited references for detail.

As convenience of exposition dictates, we use vector \(\mathbf{n}\) some of the times and and \(\mathbf{a}\) at other times. Define a state vector \(X_t\) which takes on the value \(\mathbf{n} = (n_1, n_2, \ldots, n_K)\), called frequency or occupancy vector, where \(n_i\) is the number of agents of type \(i\), \(i = 1, 2, \ldots, K\), \(n = n_1 + n_2 + \cdots + n_K\).

In our model we need to specify entry rates, exit rates and rates of type changes. Over a small time interval \(\Delta\), rates are multiplied by the length of interval to approximate the conditional probabilities up to \(o(\Delta)\). For example, entry rates by an agent of type \(j\) may be specified by

\[w(\mathbf{n}, \mathbf{n} + \mathbf{e}_j) = \phi_j(n_j, \mathbf{n}),\]

where \(\mathbf{e}_j\) is a vector with the only nonzero element of one at component \(j\).

Exit rate of an agent of type \(k\) is specified by

\[w(\mathbf{n}, \mathbf{n} - \mathbf{e}_k) = \psi_k(n_k, \mathbf{n})\]

and transition rate of type \(i\) agent changing into type \(j\) agent by

\[w(\mathbf{n}, \mathbf{n} - \mathbf{e}_i + \mathbf{e}_j) = \lambda_{i,j} \nu(n_i, n_j, \mathbf{n}).\]
We now specialize the transition rates to derive the Ewens distribution:

\[ w(n, n + e_k) = c_k(n_k + h_k), \]

for \( n_k \geq 0 \),

\[ w(n, n - e_j) = d_j n_j, \]

\( n_j \geq 1 \), and

\[ w(n, n - e_j + e_k) = \lambda_{jk} d_j n_j c_k(n_k + h_k), \]

with \( \lambda_{jk} = \lambda_{kj} \), and where \( j, k = 1, 2, \ldots K \). We assume that \( d_j \geq c_j > 0 \), and \( h_j > 0 \), and \( \lambda_{jk} = \lambda_{kj} \) for all \( j, k \) pairs.

The first transition rate specifies entry rate of type \( k \) agents, and the second that of the exit or departure rate by type \( j \) agents and the last specifies the transition intensity of changing types by agents from type \( j \) to type \( k \). In the entry transition rate specification \( c_k n_k \), with positive \( c_k \) may stands for attractiveness of larger group, such as network externality which makes it easier for others to join the cluster or group. The term \( c_k h_k \) stands for the innovation effects which is independent of the group size. These transition rates for type changes are in Kelly (1979). We need interactions or correlations among agents. It turns out that parameter \( \theta \), to be introduced in connection with (4) below, plays this role. See Aoki (2000a, 2002b). The jump Markov process thus specified has the steady state or stationary distribution

\[ \pi(n) = \prod_{j=1}^{K} \pi_j(n_j), \]

where

\[ \pi_j(n_j) = (1 - g_j)^{-h_j} \binom{-h_j}{n_j} (-g_j)^{n_j} \]

where \( g_j = c_j / d_j \).

These expressions are derived straightforwardly by applying the detailed balance conditions to the transition rates. See Kelly (1979, Chapt.1), or Aoki (2002, p. 148) for example.

To provide simpler explanation, suppose that \( g_j = g \) for all \( j \). Then, noting that \( \prod_j (1 - g)^{-h_j} = (1 - g)^{-\sum_j h_j} \), the joint probability distribution is expressible as

\[ \pi(n) = \left( -\sum_{n} h_k \right)^{-1} \prod_{j=1}^{K} \binom{-h_j}{n_j}. \]

(3)
By a suitable limiting process this distribution goes to the Ewens distribution. To see this suppose that $K$ becomes very large and $h$ very small, while the product $Kh$ approaches a positive constant $\theta$. We note that the negative binomial expression
\[
\binom{-h}{j}^{a_j}
\]
approaches $(h/j)^{a_j}(-1)^{ja_j}$ as $h$ becomes smaller. Suppose $K_n = k \leq K$. Then, there are
\[
\frac{K!}{a_1!a_2!\cdots a_n!(K - k)!}
\]
many ways of realizing a vector. Hence
\[
\pi(a) = \binom{-\theta}{n}(-1)^n\frac{K!}{a_1!a_2!\cdots a_n!(K - k)!} \prod_j \frac{h}{j}^{a_j}. \quad (4)
\]
Noting that $K!/(K - k)! \times a^k$ approaches $\theta^k$ in the limit of $K$ becoming infinite and $h$ approaching 0 while keeping $Kh$ at $\theta$, we arrive, in the limit, at the probability distribution, known as the Ewens distribution, or Ewens sampling formula very well known in the genetics literature, Ewens (1972), and Kingman (1987).
\[
\pi_n(a) = \frac{n!}{\theta^{[n]}} \prod_{j=1}^{n} \frac{(-\theta)^{a_j}}{a_j!} \frac{1}{a_j!},
\]
where $\theta^{[n]} := \theta(\theta + 1)\cdots(\theta + n - 1)$. This distribution has been investigated in several ways. See Arratia and Tavaré (1992), or Hoppe (1987). Kingman (1980) states that this distribution arise in many applications. There are other ways of deriving this distribution. We next examine some of its properties.

**The number of clusters and value of $\theta$**

Ewens sampling formula has a single parameter $\theta$. Its value influences the number of clusters formed by the agents. Smaller values of $\theta$ tends to produce a few large clusters, while larger values produce a large number of smaller clusters.
To obtain some quick feel for the influences of the value of $\theta$, take $n = 2$ and $a_2 = 1$. All other $a$s are zero. Then

$$\pi_2(a_1 = 0, a_2 = 1) = \frac{1}{1 + \theta}.$$ 

This shows that two randomly chosen agents are of the same type with high probability when $\theta$ is small, and with small probability when $\theta$ is large. In fact, $\theta$ controls correlation between agents' types or classification. Furthermore, the next two extreme situations may convey the relation between the value of $\theta$ and the number of clusters. We note that the probability of $n$ agents forming a single cluster of size $n$ is given by

$$\pi_n(a_j = 0, 1 \leq j \leq (n - 1), a_n = 1) = \frac{(n - 1)! \theta}{\theta^{[n]}}$$

while the probability that $n$ agents form $n$ clusters of size one each is given by

$$\pi_n(a_1 = n, a_j = 0, j \neq 1) = \frac{\theta^{n - 1}}{(\theta + 1)(\theta + 2) \cdots (\theta + n - 1)}.$$ 

With $\theta$ much smaller than one, the former probability is approximately equal to 1, while the latter is approximately equal to zero. When $\theta$ is much larger than $n$ the opposite is approximately true.

We can show that

$$P_n(K_n = k) = \frac{1}{\theta^{[n]}} c(n, k) \theta^k,$$

where $c(n, k)$ is known as the unsigned Stirling numbers of the first kind, and is defined by the generating function

$$\theta^{[n]} = \sum_{k=1}^{n} c(n, k) \theta^k.$$


**Delabelled Configuration Process**

Assume that agents have more than two and actually many choices. Denote the number of population by $n$ as before and the number of choices or subgroups by $K$. We regard both $n$ and $K$ as large. We let $K$ goes to
infinity while the product $K\epsilon$ is such that it goes to a finite positive limit $\theta$ by letting $\epsilon$ go to zero appropriately. The parameter $\epsilon$ will be introduced shortly.

We examine the phenomenon of $N$ agents somehow separating into $K$ categories or types by the choices they make or by the optimization problems they solve, i.e., the algorithms they use.

One natural extension of our earlier exposition is to regard the pattern

$$n = n_1 + n_2 + \cdots + n_K$$

as being multinomial distribution

$$Pr(n_1 = a_1, n_2 = a_2, \ldots, n_K = a_K | \sum_i n_i = n) = \frac{n!}{a_1! \cdots a_K!} p_1^{a_1} p_2^{a_2} \cdots p_K^{a_K}$$

where $p_i$ is the fraction of agents of type $i$, $\sum_{i=1}^{K} p_i = 1$, if we know these $p$s. Generally, we do not know these $p$s. Following Watterson (1976) and Chen (1978) we assume that $p$s are drawn from a symmetric Dirichlet distribution, and average the multi-nominal distribution with respect to the $p$s. Then under some technical conditions we arrive at the Ewens sampling formula, due to Ewens (1972), a famous distribution in the literature of population genetics. See also Aoki (2002b).

We assume that

$$\phi(p_1, p_2, \ldots, p_K) = \frac{\Gamma(K\epsilon)}{\Gamma(\epsilon)^K} \prod_{i=1}^{K} p_i^{\epsilon-1}$$

is the joint density function of $p$s, where we substitute $1 - p_1 - p_2 - \cdots - p_{K-1}$ for $p_K$.

Suppose that we take a random sample of size $r$ from the population of agents. Let $k$ be the number of agent types present in this sample. Rearrange $n$s in non-increasing order, $n_{(1)} \geq n_{(2)} \geq \cdots \geq n_{(k)}$. After averaging over $p$s, we obtain the (joint) probability of

$$Pr(n_{(1)}, n_{(2)}, \ldots, n_{(k)}; k) = \frac{r!}{n_{(1)}! \cdots n_{(k)}!} \frac{\Gamma(K\epsilon)}{\Gamma(\epsilon)^K} \frac{\prod_{i=1}^{K} \Gamma(\epsilon + n_{(i)})}{\Gamma(K\epsilon + r)} \times M,$$

where $r = n_{(1)} + \cdots + n_{(k)}$, and $M$ is the number of ways the positive values of the order statistics $n_{(1)} \geq n_{(2)} \geq \cdots \geq n_{(k)}$ can be distributed among $K$ types:

$$M = \frac{K!}{a_1! a_2! \cdots a_r! (K-k)!},$$

16
where $a_j$ is the number of $n$s equal to $j$, $j = 1, 2, \ldots r$. Note that $\sum_{j=1}^r a_j = k$, and $\sum_{j=1}^r ja_j = r$.

Now we let $K$ goes to infinity and $\epsilon$ to zero such that $K\epsilon$ approaches some constant $\theta \geq 0$. Noting that $K!/(K-k)!(\Gamma(\epsilon)\epsilon)^k$ approaches $\theta^k$ in the limit, we obtain

$$\Pr(n_{(1)}, n_{(2)}, \ldots, n_{(k)}; k) = \frac{r!\theta^k\Gamma(\theta)}{n_{(1)}n_{(2)}\ldots n_{(k)}a_1!a_2!\ldots a_r!\Gamma(\theta + r)}.$$ 

This is another way of writing the Ewens sampling formula since the product of $n_{(i)}$ equals $1^{a_1}2^{a_2} \ldots r^{a_r}$.

This is the sampling distribution for the ‘delabelled configuration process’ in which agents and types are treated as exchangeable, i.e., when we do not identify agent numbers with specific types.

**Densities of the large fractions**

Watterson (1976) and Watterson and Guess (1977) derived the expression for the densities of the $r$ largest fractions of clusters. In the case where the largest fraction $x$ is greater than 1/2, its density is given by

$$p(x) = \frac{\theta}{x}(1-x)^{\theta-1}.$$ 

For the largest two fractions $x$ and $y$ such that $y \geq (1-x)/2$, the joint density is

$$f(x, y) = \frac{\theta^2}{xy}(1-x-y)^{\theta-1}.$$ 

These have been used in Aoki (2002b) to discuss asset returns in an asset market in which there are two dominant groups of agents, that is two largest clusters such that $x + y$ is about 0.7 or larger.

**The Pitman two-parameter model**

Pitman (1992) generalized the Ewens' distribution by using the transition rates

$$w(n, n + e_j) = \frac{n_j - \alpha}{n + \theta},$$

where $\theta + \alpha > 0$.

---

8This $\alpha$ is not the same as the one in the size-biased sampling.
In terms of the rule of succession it becomes

$$\Pr(X_{n+1}|n) = \frac{n_i - \alpha}{n + \theta - K_n \alpha}$$

for $\alpha$ between 0 and 1, and $\theta$ positive, the conditional probability for a new type is

$$\Pr(X_{n+1} = \text{new}) = \frac{\theta}{n + \theta - K_n \alpha},$$


With this, the conditional probability that a new type enters in the next $\Delta$ time interval is approximately given by $\frac{K_n \alpha + \theta}{n + \theta} \Delta$. Pitman also derived the equilibrium distribution for this two-parameter version.

The two-parameter Poisson-Dirichlet distribution, $PD(\alpha, \theta)$, for some $\alpha$ between 0 and 1 and $\theta > -\alpha$ is a probability distribution on the sequence of fractions $V_n$, with $V_1 > V_2 > \cdots$, and $\sum_n V_n = 1$. Let $X_n, n = 1, 2, \ldots$ be independent random variables with Beta$(1 - \alpha, \theta + n\alpha)$ distribution. Let $U_n$ be the random variables with residual allocation, that is, $U_1 = X_1, U_2 = (1 - X_1)X_2, \ldots$. Let $V_n$ be the decreasing order statistics. This is Pitman’s $PD(\alpha, \theta)$ distribution. When $\alpha$ is zero, it reduces to the Kingman $PD(\theta)$.

This corresponds to the conditional probability

$$\Pr(\text{new type}|X_1, X_2, \ldots, X_n) = \frac{\alpha + \theta}{n + \theta},$$

and

$$\Pr(\text{existing type}|X_1, X_2, \ldots, X_n) = \frac{n - \alpha}{n + \theta}.$$

In other words, when there are $k$ clusters of types in the data, the probability of a new type appearing as the next observation is increased from the $\theta/(\theta + n)$ to $(\theta + k\alpha)/(\theta + n)$, and correspondingly the probability of observing type $i$ next is reduced to $(n_i - \alpha)/(\theta + n)$. With the partition vector $a$, the probability is

$$\Pr(a : \alpha, \theta) = \frac{1}{\theta^n} \prod_{i=1}^{K_n} (\theta + (i - 1)\alpha) \prod_{j=1}^{n} [(1 - \alpha)^{j-1}]^{a_j},$$

where $K_n = \sum a_j$, and $\sum ja_j = n$.

Sibuya and Yamato discuss urn models to generate Pitman’s two-parameter generalization of the Ewens model. See and Yamato (1995).
Other structures

There are other classes of combinatorial structures. An example which may describe a niche phenomenon in industrial organization, Sutton (1997), is as follows: Suppose that there are $K$ clusters, with the entry rate

$$w(n, n + e_j) = \lambda(c_j - n),$$

for $n_j$, non-negative integer, not greater than $c_j$, $j = 1, \ldots, K$. Here $c_j$ is the capacity of category or type $j$.

Posit the exit rate by $w(n, n - e_j) = \mu n_j$. The detailed balance condition leads to the expression of the equilibrium distribution. Given a partition vector $a$, the number of possible configuration of selection is

$$N(a) = I_{\sum_i a_i = n} \prod_{i=1}^{n} C_{w_j, a_j},$$

where $C_{a, b} = a! / b!(a - b)!$ is the binomial coefficient.

In the next section we discuss Gibbs distribution. For others see Pitman (2002) and Arratia, Barbour and Tavaré (2002). van Lint and Wilson (2001) as well as Pitman (2002) describe adding "color" as additional attribute or property, that is, internal property that might correspond to "energy" in statistical mechanics, to a given structure.

5 Economic applications and interpretations

Gibbs distribution in policy analysis

In binary choice models agents have two alternative choices. In such models, there are at most two clusters, and agents are partitioned into two clusters, called type 1 and type 2 clusters. When dynamics allow the use of detailed balance condition, as in the case where states are trees, then, Gibbs distribution may be used to model the effects of attractiveness of joining a large group, such as fashion effects, group pressures, reduced entry costs or learning costs, or crowding effect as the negative aspects of joining a large group. A simple birth-and-death model may be modified to have the transition rates

$$w(k, k + 1) = \lambda N(1 - x) \eta(x),$$

and

$$w(k, k - 1) = \mu x \{1 - \eta(x)\},$$
where $x = k/n$ is the size of cluster of type 1, and the Gibbs distribution is given by
\[
\eta(x) = \frac{\exp\{\beta g(x)\}}{\exp\{\beta g(x)\} + \exp\{-\beta g(x)\}},
\]
where $\beta$ is a non-negative parameter, and $g(x)$ may be thought of the expected value of the advantage of type 1 over type 2 when the fraction of agents of type 1 is $x$. See Aoki (1996, 2002a). Parameter $\beta$ may then be interpreted as the uncertainty or imprecision of information in deciding that choice 1 is superior to choice 2. The smaller the value of $\beta$, the less certain that choice 1 is better than choice 2. The limiting case of $\beta = 0$ corresponds to situations where no information is available on the relative merits of choices. The value of $\eta$ is equal to $1/2$ for all possible $x$ values, and the two choices are equal in the eyes of all the agents. See Aoki (2002, Chapt.6) for example.

This model has the equilibrium distribution $\pi(x) = \exp\{-\beta n U(x)\}$, where $U(x)$ is called potential, and is given by
\[
U(x) = -2 \int^x g(u)du - \beta H(x) + o(n),
\]
where $H(x) = -x \ln x - (1 - x) \ln(1 - x)$ is the Shannon entropy term arising from the combinatorial terms in the transition rates.

The maximum probability corresponds to minimum value of the potential. It is known\textsuperscript{9} that the mean $\phi = E(x)$ is goverend by the ordinary differential equation
\[
\frac{d\phi}{dt} = (1 - x)\eta(x) - x\{1 - \eta(x)\}.
\]

The critical point of this differential equation, $x^*$, corresponds to the local minima of the potential. Aoki, Yoshikawa, and Shimizu (2002) use this framework to discuss ineffectiveness of monetary policy when $\beta$ is small, that is, when there is a large degree of uncertainties about the economic future of Japan. Policy changes affecting $g(x)$ moves $x^*$ little when $\beta$ is small and the potential well is steep, that is $U''(x^*)$ is large.

Analysis of situations with many small clusters are in Aoki (1996, Sec.6.3) where information exchange equilibrium is discussed and in K. Nishimura (1999) which discusses land prices as the result of interactions of many agents with different opinions on price behavior.

\textsuperscript{9}See Aoki (1996, Sec. 5.10) for example.
In Aoki (2003) ultrametric distance is introduced between sectors of economy to describe the dynamics of labor market, because workers cannot easily move from one sector to any other sector with equal probability distributions due to the differentiated in required human capitals and job experience. The probability of move from one sector to another depends on the ultrametric distance between sectors that are organized as trees.

**Who jumps first?: A traditional approach to model two-sector economy**

In 1989 Dixit has analyzed several economic problems, such as that of how to optimally allocate capital stocks among two sectors, and of assessing the effects of exchange rate changes to induce entries or exits of firms in some export industry.

In a setting of a two-sector economy what Dixit derives is the price schedule, that is, the price as a function of the number of firms existing in one sector. When the relative price of the two goods crosses the price schedule from above or below, a move by one more firm into or from one sector to the other is triggered.

What are some of the objections to this analysis? Note that his answer is useful only to a central planner of the economy, not to individual firm managers. He does not say how a firm manager knows that it is his turn to enter or switch sectors. Also, there is apparently no uncertainty as to which firm switches. Problems of imperfect or incomplete information and externalities among firms (agents)decision processes are cleverly hidden or abstracted away in his analysis.

In our approach to business cycles in Aoki (2002 a Sec. 8.6), and in Aoki and Yoshikawa (2001) we drop the assumption that marginal factors equalize across sectors, because allocation of factor of production are not instantaneous and are sluggish in our view. This is of course quite different from what is practiced in the main line economics. We also have a stochastic process for deciding which firm moves or changes its production level.

**Alternative approach**

Jump Markov processes stay at each state it visits for a while, called sojourn or holding time, before it jumps to another state. Holding times are exponentially distributed. Given a number of agents wishing to jump, one with the minimum holding time actually can jump. This notion is applied
in Aoki (2002a, Chapt. 8) to a model of economy with several sectors. Each sector faces a fraction of the aggregate outputs of the economy as its demand. In his model some agents are in excess supplies and others are in excess demands. Those in the excess supply conditions wish to reduce their production, and those in excess demands wish to expand the production. The aggregate outputs affect demand conditions each agent faces. Therefore, as soon as one agent adjusts its output first, that changes the aggregate output, and possibly the demand conditions each agent faces. Consequently, sets of agents with positive or negative aggregate demand conditions generally change as one agent actually jump to a new state. The notion of holding time is therefore useful as a conceptual device in choosing which of the agents actually can carry out their intended decisions in conditions with externalities.

**An Alternative Approach: Basic Setup**

Aoki (1996, 2002a) presents several examples of some alternative approaches to that sketched above. Basically, our approach focuses on the random partitions of the set of firms into clusters induced by subsets formed by firms of the same types, and utilizes the conditional probability specifications for new entries and exits to derive equilibrium distributions, when they exist, for cluster sizes. We use the master equation (backward Chapman-Kolomorov equation) as the dynamic equation for the probabilities of state vectors.\(^{10}\)

Given the total number of agents, \(n\), and the number of possible types, \(K\), both of which are assumed in this paper to be known and finite for ease of explanation, we examine how the \(N\)-set, that is, the set \(\{1, 2, \ldots, n\}\) is partitioned into \(K\) clusters, or subsets. This partition is treated as a random exchangeable partition in the sense of Zabell (1992). See Kingman (1993, 1978a,b) who used the order statistics of the fractions of agents by types, and invented what is known as the paint-box process and the resultant Poisson-Dirichlet distribution to solve this problem.

**Entry of new types**

Our view on economic growth is that growth is sustained by continual introduction of goods of new types which stimulate demands for these new goods,\(^{10}\)

\(^{10}\)In a closed two-sector model the scalar variable of the number of firms in sector one, say, serves as the state variable. In an open model with \(K\) sectors, a \(K\)-dimensional vector is used.
not by R & D activities which refine existing goods, Aoki and Yoshikawa (2002).

New entries could be newly invented or improved goods, new business models, new behavioral patterns and so on. Law of succession in the statistical literature address these questions as conditional probabilities of agents entering models from outside being new or one of existing types in the model. Here we rely on recent works by Kingman and Pitman. Their models can be approximated as birth-immigration models in the context of continuous time branching processes and we introduce their results into our models. See Feng and Hoppe (2001) for the mathematical set-up.

Power laws and non-self-averaging phenomena

Aoki (2002a,b) give one potential applications of the Ewens sampling formula in finance in which stocks of a holding company is traded by a large number of agents. There are two trading rules being used by two dominant clusters of agents trade. The majority of traders are either chartists or fundamentalists. With $\theta = .3$, the share of the largest cluster is larger than 0.5, and the sum of the fractions of the two largest groups are shown to capture nearly 80 per cent of the market shares. Hence the two groups dominate the market excess demands for the shares, which in turn determine the behavior of returns. This market exhibits power laws for returns. When two types of agents switch between two strategies, that is the largest cluster of agents switch from being fundamentalists to chartists or vice versa, returns exhibit switchings of volatility as well. In this way it is also possible to relate the tail distribution of the market clearing prices with entry and exit assumptions.

6 Concluding Remarks

Although we have not mentioned the use of hierarchically structured state space in the main body, they may be useful in explaining sluggish responses of some economic phenomena. Sluggish responses of dynamics as power laws, that is, decay of the form $t^{-\alpha}$ for some positive $\alpha$ rather than exponential decay $e^{-\lambda t}$ are also found in models in which states are arranged as trees and transition rates between states are the functions of ultrametric distance as in Aoki (1996, Sec. 7.1). As the number of layers of tree nodes increase, reduce transition rates appropriately. In the limit we obtain power law decays, not exponential decays. See Ogilovski and Stein (1985), or Aoki (1996, p. 157).
In this chapter we have focussed on replacing the assumption of representative agents so commonly used in economic models by exchangeable agents. See Kirman (1992) why the assumption of representative agents is undesirable.

We propose a finitary approach to economic modeling, that is to start with a finite number of agents with discrete choice sets, and with explicit transition rates. We use exchangeable random partitionings of a finite number of exchangeable agents.

This chapter discusses several entry and exit transition rates in economic models. In particular, it presented Ewens and related distributions as candidates for distributions of cluster sizes formed by a large number of economic agents who interact in a market. This distribution seems to be very useful in economic modelings, although we have only a few economic or financial applications so far. However, Arratia, Barbour and Tavaré (1992), and Kingman (1980) strongly suggest that the Ewens' and related distributions are robust and ubiquitous.

In our approach, clusters of agents by types are the basic ingredients for modeling and analysis. There are two extreme cases of interest: Models with a large number of small clusters, and models with a few clusters of large sizes. Let $x_i, i = 1, 2, \ldots$ be the sizes of clusters of type $i$. By definition $\sum_i x_i = 1$. We examine the question if $\langle Y^2_k \rangle - \langle Y_k \rangle^2$ goes to zero as the number of agents goes to infinity, where $Y_k = \sum_i x_i^k$. Here $Y_2$ has been used in several context. In the older literature on industrial organizations this is called Herfindahl index of concentration. See Aoki (2002a, p. 173). Similar expressions are used in genetics and in physics. See Higgs (1995), Mekjian (1991) or Derrida (1997). With the share generated as the RAM using beta distributions, they do not have self-averaging property. This means that significant fluctuations are observed on these moments from sample to sample, and the usual statistical or econometric methods may not be appropriate for those phenomena with non-self-averaging properties. Carlton (1999) discusses some estimation issues of two-parameter Poisson-Dirichlet distribution.

References


Los Angeles.


————, and ———— (2002). "Demand saturation-creation and economic growth", J. Econ. Behav. Org. 48


———, (1996) "Remarks on the law of succession", in Athens Conf. on Applied Probab. on Time Series Analysis
———. (1980), Mathematics of Genetic Diversity, SIAM, Philadelphia


Watterson, G.A., "The stationary distribution of the infinitely-many neutral alleles diffusion


