

A New Formula for Partitions in a Set of Entities into Empty and Nonempty Subsets, and Its Application to Stochastic and Agent-Based Computational Models

Ghennadii Gubceac, Roman Gutu, Florentin Paladi

Department of Theoretical Physics, State University of Moldova, Chisinau, Republic of Moldova

Email: fpaladi@usm.md

Received June 3, 2013; revised July 3, 2013; accepted July 10, 2013

Copyright © 2013 Ghennadii Gubceac *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

In combinatorics, a Stirling number of the second kind $S(n, k)$ is the number of ways to partition a set of n objects into k nonempty subsets. The empty subsets are also added in the models presented in the article in order to describe properly the absence of the corresponding type i of state in the system, *i.e.* when its “share” $p_i = 0$. Accordingly, a new equation for partitions $P(N, m)$ in a set of entities into both empty and nonempty subsets was derived. The indistinguishability of particles (N identical atoms or molecules) makes only sense within a cluster (subset) with the size $0 \leq n_i \leq N$. The first-order phase transition is indeed the case of transitions, for example in the simplest interpretation, from completely liquid state $\text{type}L = \{n_1 = N, n_2 = 0\}$ to the completely crystalline state $\text{type}C = \{n_1 = 0, n_2 = N\}$. These partitions are well distinguished from the physical point of view, so they are ‘typed’ differently in the model. Finally, the present developments in the physics of complex systems, in particular the structural relaxation of supercooled liquids and glasses, are discussed by using such stochastic cluster-based models.

Keywords: Partitions; Agent-Based Models; Stochastic Processes; Complex Systems

1. Introduction

In mathematics, particularly in combinatorics and the study of partitions, a Stirling numbers of the second kind

$S(n, k) \equiv \left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ count the number of ways to partition a

set of n labeled objects into k nonempty unlabelled subsets. By the way, Stirling numbers of the second kind show up more often than those of first and third kind (or Lah numbers), and James Stirling himself considered this kind first [1]. Equivalently, they count the number of different equivalence relations with precisely k equivalence classes that can be defined on an n element set, and they can be calculated using the explicit formula

$S(n, k) = \frac{1}{k!} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} j^n$, where $\binom{k}{j}$ is the bi-

nomial coefficient. The sum over the values for k of the Stirling numbers of the second kind, *i.e.* $\sum_{k=0}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\} = B_n$,

gives the n th Bell number, that is the total number of partitions of a cluster with n entities or agents [2]. The question arises when the empty subsets or clusters are needed to be used in the models, for example, with heterogeneous structure interactions and, subsequently, the partition by Stirling numbers of the second kind becomes inappropriate to count partitions in such complex systems [3].

In general, agent-based modeling is currently a technique widely used to simulate complex systems in computer science and social sciences. On the other hand, a Markovian process is a stochastic process whose future probabilities are determined by its most recent values. The agent-based computational models (ABM) fits well this description, except for the cases when decisions are dependent on the state of the systems of more than one steps ago, which is the case when ABM agents experience learning, adaptation, and reproduction [4].

In this study, a general equation that describes clustering process among interacting agents in heterogeneous

populations, *i.e.* the partition process in a set of entities into empty and nonempty clusters, is derived and used to study how different behavioral norms affect the individual and social welfare in a population with heterogeneous preferences. One can consider them as an idealization of an imperative and a more liberal approach to social norms or stylized behavioral rules studied by agent-based computational models [5,6]. Another application refers to the generic stochastic model for crystal nucleation which is useful to depict the impact of interface between the nucleus considered as a cluster of a certain number of molecules and the liquid phase for the enhancement of the overall nucleation process. It is generally known that first-order phase transitions occur by nucleation mechanism, and both the nucleus, a cluster of atoms or molecules, and the nucleation work, a energy barrier to the phase transition, are basic thermodynamic quantities in the theory of nucleation. However, the critical nucleus formation is statistically a random event with a probability largely determined by the nucleation work which increases with the subnuclei size [7]. The traditional differential equation modeling is not the alternative to agent-based models; only a set of differential equations, each describing the dynamics of one of the system's constituent units, is an agent-based model [8].

The general formulation is outlined in Section 2. In Section 3 a probabilistic approach to the crystal nucleation process is considered. The main conclusions are presented in Section 4.

2. The Model

There are N entities which can be in 3 different states (call them *left*, *center* and *right*), and can play 3 actions (again *left*, *center* and *right*). Interaction in this agent-based model involves always one *active* and one *passive* player, but agents can play both roles interchangeably. They have preferences over their states: *accept* one state, are *neutral* with respect to another state and *reject* the remaining state. When two agents meet, the active player sets the passive player's state according to his action,

which in turn is determined by one of the applied rule. This identifies only 6 possible combinations. Denote with $p_1 \cdots p_6$ the shares of the population characterized by each combination of preferences, as in **Table 1**. That is, drawing randomly one agent, it will be of type i with probability p_i . After each interaction, the passive player gets a payoff of +1 if it is in the accepted state, a payoff of 0 if it is in the neutral state, and a payoff of -1 if it is in the rejected state. The active player does not get any feedback. If the active player follows the first J -rule, it always plays the action corresponding to the accepted state. If it follows the second H -rule, it randomizes between actions corresponding to the accepted and neutral states. The further example will clarify. Suppose two individuals, A and B , meet. Player A is the active one, and rejects *left*, accepts *right*, and is thus neutral with respect to the *center*. Player B is the passive one. It accepts *left*, rejects *right*, and is neutral with respect to *center*, like player A . Suppose A follows the J -rule, and will play *right*, setting B 's state to *right*. B will then have a payoff of -1. Suppose, on the other hand, that A follows the H -rule, and will randomize between *center* and *right*. The payoff for B could then be either 0 or -1. Note that there is no strategic interaction in the model: the passive player's payoff depends on the active player's choice, but the active player's choice does not depend on the passive player in any way, so the game-theoretic solution concepts like Nash equilibrium become useless.

Aggregate results are defined in terms of both the mean π and the variance σ^2 of the payoffs which denote the stability and the heterogeneity of population, respectively. However, in order to avoid arbitrary choices we do not specify a particular functional form, and report separately the results for the mean and the variance. Let $N = 1, 2, \dots, \infty$ be the total number of entities in the model, and the $\{n_1, n_2, n_3, n_4, n_5, n_6\}$ is their partition into $m = 6$ subsets. Each subset can be called *cluster*, and the process itself—*clustering*. The size of each cluster can vary from 0 to N , $n_i = 0, N, i = 1, 6$, and $\sum_{i=1}^6 n_i = N$. The number of possible partitions P is a

Table 1. Distribution of preferences in the heterogeneous system with 6 types of states.

Type	Accepted State	Rejected State	Share	<i>left</i>	<i>center</i>	<i>right</i>
1	<i>left</i>	<i>center</i>	p_1	p_1	$-p_1$	0
2	<i>left</i>	<i>right</i>	p_2	p_2	0	$-p_2$
3	<i>center</i>	<i>left</i>	p_3	$-p_3$	p_3	0
4	<i>center</i>	<i>right</i>	p_4	0	p_4	$-p_4$
5	<i>right</i>	<i>left</i>	p_5	$-p_5$	0	p_5
6	<i>right</i>	<i>center</i>	p_6	0	$-p_6$	p_6
Coordinates				$p_1 + p_2 - p_3 - p_5$	$-p_1 + p_3 + p_4 - p_6$	$-p_2 - p_4 + p_5 + p_6$

function of N and m , and the explicit solution is

$$\begin{aligned}
 P(N, m = 6) &= {}_6C_1 + (N-1) {}_6C_2 + \left(1 + \sum_{i=2}^{N-2} i\right) {}_6C_3 \\
 &+ \left(N-3 + (N-2) \sum_{i=2}^{N-3} i - \sum_{i=2}^{N-3} i^2\right) {}_6C_4 \\
 &+ \left(1 + \sum_{i=2}^{N-4} i + \sum_{i=3}^{N-3} i \sum_{l=2}^{i-1} l - (N-3) \sum_{i=2}^{N-4} i^2 + \sum_{i=2}^{N-4} i^3\right) {}_6C_5 \\
 &+ (N-5) \sum_{i=2}^{N-5} (N-i-4) i + \sum_{i=3}^{N-4} (N-i-3) i \sum_{l=2}^{i-1} l \\
 &- \sum_{i=2}^{N-5} i^2 \sum_{l=i+1}^{N-4} l + \left(\sum_{i=2}^{N-5} (N-i-4) i^3\right) {}_6C_6 \\
 &= 2(74N - 213) + 2(11N - 13) \sum_{i=2}^{N-5} i \\
 &+ 2(1-3N) \sum_{i=2}^{N-5} i^2 + (2+N) \sum_{i=2}^{N-5} i^3 - \sum_{i=2}^{N-5} i^4 \\
 &+ \sum_{i=3}^{N-4} ((3+N)i - i^2) \sum_{l=2}^{i-1} l - \sum_{i=2}^{N-5} i^2 \sum_{l=i+1}^{N-4} l \\
 &= \frac{1}{5!} \prod_{i=1}^5 (N+i),
 \end{aligned}$$

where ${}_6C_1 = {}_6C_5 = 6$, ${}_6C_2 = {}_6C_4 = 15$, ${}_6C_3 = 20$, ${}_6C_6 = 1$ are combinations given by the binomial coefficient

$${}_mC_k \equiv \binom{m}{k} \equiv \frac{m!}{(m-k)!k!}.$$

In general, when the number of subsets $m > 1$, the number of possible partitions P of N agents into m subgroups or subsets is

$$P(N, m) = \frac{1}{(m-1)!} \prod_{i=1}^{m-1} (N+i). \tag{1}$$

The matrix is diagonally symmetric

Table 2. Bi-triangular array of values for the numbers of partitions $P(N, m)$.

N, m	1	2	3	4	5	6	7	8	9	10	Recurrence Relation
0	1	1	1	1	1	1	1	1	1	1	
1	1	2	3	4	5	6	7	8	9	10	
2	1	3	6	10	15	21	28	36	45	55	
3	1	4	10	20	35	56	84	120	165	220	
4	1	5	15	35	70	126	210	330	495	715	
5	1	6	21	56	126	252	462	792	1287	2002	$P(N, m) = P(N-1, m) + P(N, m-1)$
6	1	7	28	84	210	462	924	1716	3003	5005	
7	1	8	36	120	330	792	1716	3432	6435	11,440	
8	1	9	45	165	495	1287	3003	6435	12,870	24,310	
9	1	10	55	220	715	2002	5005	11,440	24,310	48,620	
10	1	11	66	286	1001	3003	8008	19,448	43,758	92,378	

$$P(N, i+1) = P(i, N+1)$$

for $i = 0, 1, \dots, N-1$, and it can be formed by arranging the partition numbers according to the parameters N and m . **Table 2** contains this array of values for the given numbers of partitions. The recurrence relation is

$$P(N, m) = P(N-1, m) + P(N, m-1)$$

for $m > 0$ with the initial condition $P(0, m > 0) = 1$. For instance, the number 330 in column $m = 5$ and row $N = 7$ is given by $330 = 210 + 120$, where 210 is the number above and 120 is the number on the left of 330. The diagonal elements of the bi-triangular array of values for the numbers of partitions are 1, 2, 6, 20, 70, 252, 924, 3432, 12,870, 48,620, ..., and it is represented by n th central binomial coefficient:

$$C(2n, n) \equiv \binom{2n}{n} = \frac{(2n)!}{(n!)^2} \text{ for all } n \geq 0.$$

They are called central since they show up exactly in the middle of the even-numbered rows in Pascal's triangle. These numbers have the generating function

$$\begin{aligned}
 \frac{1}{\sqrt{1-4x}} &= 1 + 2x + 6x^2 + 20x^3 + 70x^4 + 252x^5 \\
 &+ 924x^6 + 3432x^7 + \dots
 \end{aligned}$$

It is known that the asymptotics for the central binomial coefficient $C(2n, n)$ can be written in the form of a particular case of the Wallis formula, *i.e.*

$$\lim_{n \rightarrow \infty} \frac{2^{2n}}{n \binom{2n}{n}^2} = \pi \lim_{n \rightarrow \infty} \frac{n [\Gamma(n)]^2}{\left[\Gamma\left(\frac{1}{2} + n\right)\right]^2} = \pi,$$

where $\Gamma(x)$ is the gamma function, so

$$\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}, \quad n \rightarrow \infty.$$

By the way, this equation can also be used to determine the constant $\sqrt{2\pi}$ in front of the Stirling's formula.

Now it is straightforward to see that when all individuals share the same preferences (*polarization*) the first rule gives a higher payoff. In the other extreme case, when preferences are equally distributed in the ensemble (*dispersion*) and $p_1 = p_2 = \dots = p_6 = \frac{1}{6}$, it is again

straightforward to see that the rules are equivalent, and lead to an average payoff $\pi = 0$. Consider next an *active* player of type 1 (accepts *left* and rejects *center*) which meets in turn all other (*passive*) agents, including himself. If it follows the first rule, then it will play *left* causing a payoff of +1 in $(p_1 + p_2)N$ agents, and a payoff of -1 in $(p_3 + p_5)N$ agents. Note that there are $(p_1 + p_2)N$ similar entities in the ensemble. Suppose now that everybody meets everybody else both as *active* and as *passive* players. Coupling them randomly and randomly choosing *active* and *passive* players only adds some noise to the simulation results. So the average payoff when everybody plays according to the first *J*-rule is

$$\begin{aligned} \pi_1 = & (p_1 + p_2)(p_1 + p_2 - p_3 - p_5) \\ & + (p_3 + p_4)(-p_1 + p_3 + p_4 - p_6) \\ & + (p_5 + p_6)(-p_2 - p_4 + p_5 + p_6). \end{aligned} \quad (2)$$

Similarly, the average payoff with the second *H*-rule is

$$\begin{aligned} \pi_2 = & \frac{1}{2}(p_1(p_1 - p_3 - p_5 + p_6) \\ & + p_2(p_2 - p_3 + p_4 - p_5) + p_3(-p_1 + p_3 + p_5 - p_6) \\ & + p_4(-p_1 + p_2 + p_4 - p_6) + p_5(-p_2 + p_3 - p_4 + p_5) \\ & + p_6(p_1 - p_2 - p_4 + p_6)). \end{aligned} \quad (3)$$

To study the behavior of $\pi_1 - \pi_2$ based on Equations (2) and (3), it is convenient to set some of the probabilities to zero. It is straightforward to see that when there is just one probability different from 0 (and thus equal to 1)

we have $\pi_1 - \pi_2 = \frac{1}{2} > 0$. In general, the number of dif-

ferent possible combinations of non-zero probabilities is given by the above binomial coefficient ${}_m C_k$.

Thus for two non-zero probabilities we obtain a set of ${}_6 C_2 = 15$ equations which can be grouped in just 3 different functional forms shown in **Table 3**. **Figure 1** graphs these 3 curves for all values of p_i and $p_j = 1 - p_i$. The *J*-rule still performs better in all cases but one, when

Table 3. Expressions of $\pi_1 - \pi_2$ for different numbers of non-zero probabilities $p_i, i = 1, 2, \dots, 6$.

No. p_i	Non-zero Probabilities	Expressions for $\delta = \pi_1 - \pi_2$
	$(p_1, p_2) (p_3, p_4) (p_5, p_6)$	$\delta_1 = \frac{1}{2} + p_i - p_i^2$
2	$(p_1, p_3) (p_1, p_6) (p_2, p_4)$ $(p_2, p_5) (p_3, p_5) (p_4, p_6)$	$\delta_2 = \frac{1}{2}(1 - 2p_i)^2$
	$(p_1, p_4) (p_1, p_5) (p_2, p_3)$ $(p_2, p_6) (p_3, p_6) (p_4, p_5)$	$\delta_3 = \frac{1}{2}(1 + 3p_i(p_i - 1))$
	$(p_1, p_2, p_3) (p_1, p_2, p_4) (p_1, p_2, p_5) (p_1, p_2, p_6) (p_1, p_3, p_4) (p_2, p_3, p_4)$ $(p_3, p_4, p_5) (p_3, p_4, p_6) (p_1, p_5, p_6) (p_2, p_5, p_6) (p_3, p_5, p_6) (p_4, p_5, p_6)$	$\delta_1 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + 4p_h p_i - 2p_h p_j - p_i p_j)$
3	$(p_1, p_3, p_5) (p_1, p_3, p_6) (p_2, p_3, p_5) (p_2, p_4, p_5) (p_1, p_4, p_6) (p_2, p_4, p_6)$	$\delta_2 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 - 2(p_h p_i + p_i p_j) - p_h p_j)$
	$(p_2, p_3, p_6) (p_1, p_4, p_5)$	$\delta_3 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 - p_h p_i - p_h p_j - p_i p_j)$
	$(p_1, p_2, p_3, p_4) (p_1, p_2, p_5, p_6) (p_3, p_4, p_5, p_6)$	$\delta_1 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + p_k^2 + 4(p_h p_i + p_i p_j) - 2(p_h p_j + p_i p_k) - p_i p_j - p_h p_k)$
4	$(p_1, p_2, p_3, p_5) (p_1, p_2, p_4, p_6) (p_1, p_3, p_4, p_6) (p_1, p_3, p_5, p_6)$ $(p_2, p_3, p_4, p_5) (p_2, p_4, p_5, p_6)$	$\delta_2 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + p_k^2 + 4p_h p_i - 2(p_h p_j + p_i p_k + p_j p_k) - p_i p_j - p_i p_k)$
	$(p_1, p_2, p_3, p_6) (p_1, p_2, p_4, p_5) (p_1, p_3, p_4, p_5) (p_2, p_3, p_4, p_6)$ $(p_2, p_3, p_5, p_6) (p_1, p_4, p_5, p_6)$	$\delta_3 = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + p_k^2 + 4p_h p_i - 2(p_h p_j + p_h p_k) - p_i p_j - p_i p_k - p_j p_k)$
5	$\delta = \frac{1}{2}(p_h^2 + p_i^2 + p_j^2 + p_k^2 + p_l^2 + 4(p_h p_i + p_j p_k) - 2(p_h p_j + p_i p_k + p_l p_i + p_j p_l) - p_i p_j - p_h p_k - p_h p_l - p_k p_l)$	

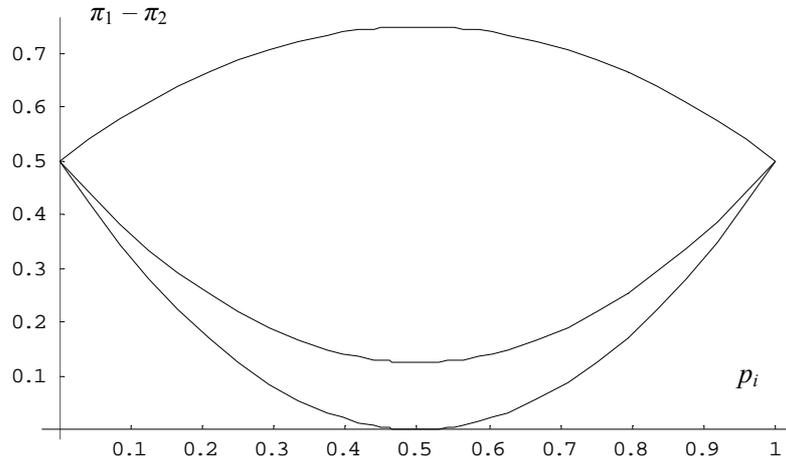


Figure 1. $\pi_1 - \pi_2$ in the case of two non-zero probabilities, p_i and $p_j = 1 - p_i$.

the two rules are equivalent. However, from three non-zero probabilities onward things start to look differently. For three non-zero probabilities we have ${}_6C_3 = 20$ equations, while for four non-zero probabilities there are ${}_6C_4 = 15$ equations, and for five non-zero probabilities ${}_6C_5 = 6$ equations. These equations reduce to just three different functional forms in case of three and four non-zero probabilities, and to just one expression in case of five non-zero probabilities.

A trend towards a better performance of the H -rule, as the distribution of preferences in the system becomes less polarized, is evident. However, in order to better investigate it, a definition of how much preferences are polarized is needed. We represent the distribution of states as a single point in a three dimensional space, where the axes are labeled l , c and r . The l coordinate is found by counting all agents who accept *left*, and subtracting all agents who reject *left*. The result is then normalized to the size of the population. Similarly for the other two coordinates. Hence,

$$\begin{aligned} l &= p_1 + p_2 - p_3 - p_5, \\ c &= p_3 + p_4 - p_1 - p_6, \\ r &= p_5 + p_6 - p_2 - p_4, \end{aligned} \tag{4}$$

where $l + c + r = 0$. Note that different distributions of states can lead to the same point in the sphere. For instance, the point in the origin is given not only by $p_1 = p_2 = \dots = p_6 = 1/6$, but by any combination of preferences such as $p_1 = p_3$, $p_2 = p_5$, and $p_4 = p_6$. Taking into account Equation (4), one can define now the polarization of states as the distance from the center of the sphere:

$$d(l, r, c) \equiv d(p_1, p_2, \dots, p_6) = \sqrt{l^2 + r^2 + c^2}. \tag{5}$$

Note that $d \in [0, \sqrt{2}]$: all points thus lie inside a sphere around the origin.

The variances σ_1^2 and σ_2^2 are defined for each discrete distribution $D \equiv 1, 2$ with the expectation (mean) value π_D as follows:

$$\sigma_D^2 = \sum_{i=1}^6 p_i (\pi_{i,D} - \pi_D)^2, \tag{6.1}$$

where

$$\begin{aligned} \pi_{1,1} &= p_1 + p_2 - p_3 - p_4, \\ \pi_{4,1} &= p_3 + p_4 - p_5 - p_6, \\ \pi_{2,1} &= p_1 + p_2 - p_5 - p_6, \\ \pi_{5,1} &= -p_1 - p_2 + p_5 + p_6, \\ \pi_{3,1} &= -p_1 - p_2 + p_3 + p_4, \\ \pi_{6,1} &= -p_3 - p_4 + p_5 + p_6 \end{aligned} \tag{6.2}$$

and

$$\begin{aligned} \pi_{1,2} &= (p_1 - p_3 - p_5 + p_6)/2, \\ \pi_{4,2} &= (-p_1 + p_2 + p_4 - p_6)/2, \\ \pi_{2,2} &= (p_2 - p_3 + p_4 - p_5)/2, \\ \pi_{5,2} &= (-p_2 + p_3 - p_4 + p_5)/2, \\ \pi_{3,2} &= (-p_1 + p_3 + p_5 - p_6)/2, \\ \pi_{6,2} &= (p_1 - p_2 - p_4 + p_6)/2. \end{aligned} \tag{6.3}$$

Figure 2 explores how outcomes vary as functions of the distance $d(l, r, c)$ defined by Equation (5). The whole range $[0, 1]$ is sampled, for all probabilities $p_1 \dots p_6$. The step considered for creating all combinations of probabilities is 0.025, i.e. the total number of agents is 40. The average values for π_1 and π_2 are shown in Figure 2(a), and for each value of the distance from the center of the sphere, $d(l, r, c)$, the frequency of wins with each rule is computed (see Figure 2(b)). When $\pi_1 - \pi_2 > 0$ a win is assigned to the first J -rule, and when $\pi_1 - \pi_2 < 0$ a win is assigned to the second

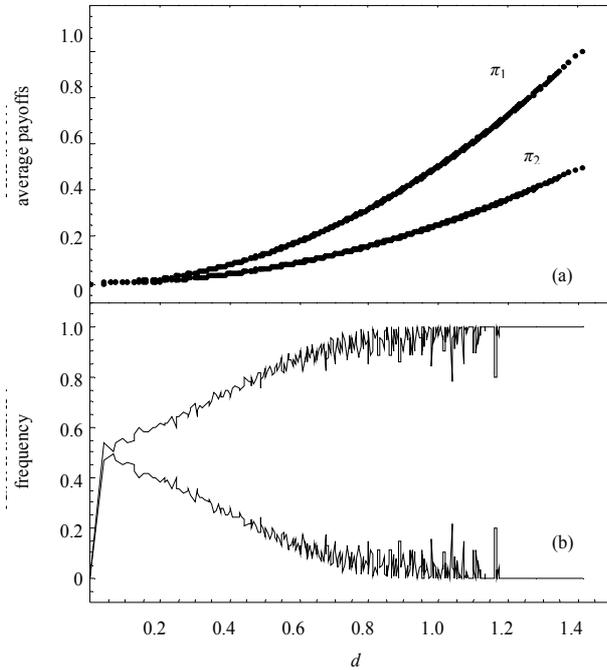


Figure 2. Average values for π_1 (upper line) and π_2 (a), and relative frequency of negative and positive values for $\pi_1 - \pi_2$ (b) as functions of the distance $d(p_1, p_2, \dots, p_6)$ from the center of sphere with a radius of $\sqrt{2} : 40$ entities, 6 clusters in a similar ABM model.

one. Exactly in the center of the sphere the two rules lead to the same payoff, independently of the underlying distribution of states. Close to the center, each rule wins in about 50% of the cases. Then, as we move away from the center, the first rule improves its performance, and it is always better when the states are totally polarized, but the total number of states for intermediate values of the distance d is much larger than for the dispersed and polarized states.

Meanwhile, the variance (6.1) with the J -rule (6.2) is generally higher than the variance with the H -rule (6.3), especially when the preferences are dispersed in the system, and when they are quite polarized. On average, however, when one rule is better in terms of higher expected payoffs it is also better in terms of lower heterogeneity, and there are the same average values of the scale-free coefficients of variation on the distance d for any level of fragmentation of states.

3. A Probabilistic Approach to the Crystal Nucleation Process

The second application refers to the nucleation process, a widely spread phenomenon in both nature and technology, which may be considered as a representative of the aggregation phenomena in complex systems. Let's consider N atoms which can be in 3 different states (*cluster*,

liquid and their *interface*), and can perform 4 possible moves: *liquid to interface*, *interface to liquid*, *interface to cluster*, and *cluster to interface*. One can identify 4 different combinations denoted with probabilities $p_1 \dots p_4$, as in **Table 4**. That is, drawing randomly one particle, it will be of type i with probability p_i . Let $N = 1, 2, \dots, \infty$ be the total number of atoms in the system, and $\{n_1, n_2, n_3, n_4\}$ are their partition into 4 subsets. Each subset can be called cluster, and the number of possible partitions (1) in this case is

$$P(N, m = 4) = \frac{1}{3!} \prod_{i=1}^3 (N + i),$$

where $n_i = \overline{N}, i = 1, 4$ and $\sum_{i=1}^4 n_i = N$. For example, in

a system of $N=1000$ atoms, $P(N, m = 4)$ equals to 167,668,501! Accordingly, the number of repeated computer runs in an ABM model would be very large due to different possible partitions. But we are able to overcome this problem, as already mentioned in the previous section, by developing similar stochastic mathematical models which can describe exactly the results of the agent-based computational models, and, finally, by bridging the gap between ABM modeling and stochastic processes.

Let's consider further that each particle interacts with the entire group both as an aggressor in terms of the Kolmogorov mathematical theory [9], and as a passive agent in terms of the ABM computational models as well. Then the mean π , namely the stability index, takes here the form

$$\pi = p_1(p_1 - p_3) + (p_2 + p_3)(-p_1 + p_2 + p_3 - p_4) + p_4(p_4 - p_2) \quad (7.1)$$

or, taking into account that $\sum_{i=1}^4 p_i = 1$, one can exclude

one probability, for example p_4 , from the above equation:

$$\pi = p_1(p_1 - p_3) + (1 - p_1 - 2p_2 - p_3)(-p_1 - p_2 - p_3) + (p_2 + p_3)(-1 + 2p_2 + 2p_3). \quad (7.2)$$

One can represent again the distribution of states as a three-dimensional point (l, r, c) inside a sphere, and the mean payoff π (Equation (7.1) or Equation (7.2)) can be obtained as a function of the distance

$$d(l, r, c) \equiv d(p_1, p_2, p_3, p_4) = \sqrt{l^2 + r^2 + c^2},$$

where the axes are labeled l, c and $r: l = p_1 - p_3, c = p_2 + p_3 - p_1 - p_4, r = p_4 - p_2$, and, as stated in the previous section, $l + c + r = 0$ and $d \in [0, \sqrt{2}]$. Thus different distributions of states can lead to the same point in the sphere, *i.e.* different microscopic partitions can

Table 4. Distribution of accepted and rejected states in the system.

Type	Accepted State	Rejected State	Share	<i>cluster</i>	<i>interface</i>	<i>liquid</i>
1	<i>cluster/nucleus</i>	<i>interface</i>	p_1	p_1	$-p_1$	0
2	<i>interface</i>	<i>liquid</i>	p_2	0	p_2	$-p_2$
3	<i>interface</i>	<i>cluster/nucleus</i>	p_3	$-p_3$	p_3	0
4	<i>liquid</i>	<i>interface</i>	p_4	0	$-p_4$	p_4
Coordinates				$p_1 - p_3$	$-p_1 + p_2 + p_3 - p_4$	$-p_2 + p_4$

generate the same result on aggregate inside a sphere around the origin. The results for the two limit cases are obviously: if all particles would show the same behavior, then $d = \sqrt{2}$ and there is a maximum stability of states in such a completely asymmetrical system, but $\pi = 0$ for a homogeneous system, $p_1 = p_2 = p_3 = p_4 = 1/4$, and for combinations such as $p_1 = p_3$ and $p_2 = p_4$ in the case of unstable states.

4. Conclusions

Mathematical models developed in this paper were used together with agent-based computations to study interactions in two quite different heterogeneous systems. The models are general ones and allow simulations for any size of the system. Our results also support the idea that general properties observed in the complex systems of any kind, which arise from the interplay between random interactions and their complex structures, could be successfully investigated by the same tool. For example, these properties are frequently found in the physical sciences, and are the domain of non-equilibrium statistical mechanics. Because of the real limitations in the use of analytical methods to study such problems, it is often necessary to resort to numerical ones, and the advent of computer simulations has led to an increase of scientific activity in this area that has emerged nowadays as a major subject of interdisciplinary research.

The recent discussions about the gap between agent-based computational models and stochastic analytical models have stimulated the research on this topic [3,10]. In such context, the article supports a unified probabilistic approach to simulation of multi-agent interactions in heterogeneous complex systems. This method appears phenomenological if it is not agent-based, and we prove analytically that the average outcomes of multiple agent-based computations can be described precisely. Furthermore, a useful aggregation procedure for representing the three-dimensional distribution of states, and a general formula that describes clustering process among interacting agents in heterogeneous populations, which is the Equation (1) for partitions in a set of entities into empty and nonempty subsets, are proposed. Bi-triangular array of values for the numbers of partitions is presented, and

its diagonal elements are represented by n th central binomial coefficient.

In particular, we obtained that different distributions of states can lead to the same point inside the sphere around the origin, *i.e.* different microscopic partitions can generate the same aggregate result. It would be a particular conclusion set for the analyzed situations, but one would expect this in different circumstances due to the universal aspects observed in the behavior of complex systems.

5. Acknowledgements

F.P. thanks Dr. Matteo Richiardi for his ABM contribution.

REFERENCES

- [1] R. L. Graham, D. E. Knuth and O. Patashnik, "Concrete Mathematics. A Foundation for Computer Science," 2nd Edition, Addison-Wesley Professional, Reading, 1994.
- [2] J. Sandor and B. Crstici, "Handbook of Number Theory II," Kluwer Academic Publishers, Dordrecht, 2004. <http://dx.doi.org/10.1007/1-4020-2547-5>
- [3] F. Paladi, "On the Probabilistic Approach to Heterogeneous Structure Interactions in Agent-Based Computational Models," *Applied Mathematics and Computation*, Vol. 219, No. 24, 2013, pp. 11430-11437. <http://dx.doi.org/10.1016/j.amc.2013.05.042>
- [4] E. Bonabeau, "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 99, No. 3, 2002, pp. 7280-7287. <http://dx.doi.org/10.1073/pnas.082080899>
- [5] M. Richiardi and F. Paladi, "Jesus, Hillel and the Man of the Street. Moral and Social Norms in Heterogeneous Populations," LABORatorio R. Revelli Working Paper 40, 2005. http://eco83.econ.unito.it/terna/swarmfest2005papers/richiardi_paladi.pdf
- [6] M. Richiardi, "Jesus vs Hillel. From Moral to Social Norms and Back," *European Journal of Economic and Social Systems*, Vol. 19, No. 2, 2006, pp. 171-190.
- [7] D. Kashchiev, "Nucleation. Basic Theory with Applications," Butterworth-Heinemann, Oxford, 2000.
- [8] C.W. Gardiner, "Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences," 2nd Edition

tion, Springer-Verlag, Berlin, 1985.

- [9] A. N. Kolmogorov, "On Statistical Theory of Metal Crystallisation (in Russian)," *Izvestiya Academy of Sciences, USSR, Mathematics*, Vol. 3, 1937, pp. 355-360.
- [10] L. Feng, B. Li, B. Podobnik, T. Preis and H. E. Stanley, "Linking Agent-Based Models and Stochastic Models of Financial Markets," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 109, No. 22, 2012, pp. 8388-8393.
<http://dx.doi.org/10.1073/pnas.1205013109>