9. Modelul Ewens

Vom considera un sistem dinamic, format din n elementele microscopice (unități, agenți) şi d categorii (celule, poziții sau *site*-uri), a cărui stare este descrisă de un vector cu numere de ocupare intregi non-negative $\overline{n} = (n_1,...,n_i,...,n_d)$, $n_i \geq 0$, $\sum_{i=1}^{d} n_i = n$. Starea sistemului variază cu timpul la schimbarea celulelor de către agenți. Această dinamică probabilistică este modelată ca o extracție a unor agenți care abandonează temporal sistemul iar apoi are loc o redistribuție a aceloraşi unități, dar, de obicei, în celule diferite de cele inițiale. În intervalul de timp (t, t+1) dimensiunea populției se reduce pe durata selecției, iar apoi revine gradual la dimensiunea inițială după realizarea tuturor repartițiilor. Astfel $\overline{n}(t)$ şi $\overline{n}(t+1)$ posedă aceiaşi dimensiune n, iar sistemul, prin urmare, este unul de tip închis. Setul de vectori $\overline{n}(0), \overline{n}(1), ..., \overline{n}(t)$ este o descriere temporală a numerelor de ocupare a d *site*-uri. Fie k reprezintă numărul total de clusteri în sistem, iar g este numărul finit de locuri existente (g>n). Atunci g-k este numărul locurilor neocupate. Setul $\overline{s}(0), \overline{s}(1),...,\overline{s}(t)$ caracterizează numărul de ocupare în diferite momente de timp a g poziții, iar $\pi(\overline{s})$ este distribuția corespunzătoare de echilibru. Fie în continuare vectorul $\overline{z} = (z_1,..., z_n)$ reprezintă descrierea statistică a pozițiilor ocupate, unde z_i este numărul de clusteri care conțin i agenți. Prin urmare, $\sum_{i=1}^{n} iz_i = n$ este o constantă a modelului, iar $\sum_{i=1}^{n} z_i = k$ este o variabilă aleatoare, adică numărul de clusteri, $1 \leq k \leq n$. Pe baza definiției lui z_i urmează, că, pentru orice distribuție de probabilitate $P(.)$, valoarea medie a numărul de clusteri care conțin i agenți

$$E(z_i) = \sum_{j=1}^{g} P(s_j = i), \tag{9.1}$$

iar în cazul când $P(s_j = i) = P(i)$ se obține $E(z_i) = gP(i)$.

Procesele de formare a clusterilor [11, 35, 36] sunt deseori descrise cu ajutorul formulei lui Ewens, *Ewens Sampling Formula* (ESF) [37−39]. Această distribuție a fost indrodusă de către Ewens cu peste 30 de ani în urmă în contextul unor studii dedicate geneticii populației:

$$P(z) = \frac{n!}{\theta^{[n]}} \prod_{i=1}^{n} \left(\frac{\theta}{i}\right)^{z_i} \frac{1}{z_i!}, \tag{9.2}$$

unde $\theta > 0, \theta^{[n]} = \theta(\theta+1)...(\theta+n-1)$ este simbolul Pochhammer (vezi **Anexa 3**) şi $\bar{z} = (z_1,...,z_n)$, $\sum_{i=1}^{n} iz_i = n$, reprezintă un vector de distribuție, unde n este numărul total de unități în sistem.

Garibaldi et al [39] au studiat trăsăturile esențiale ale acestui model: 1. Crearea sau nimicirea unui cluster este rezultatul redistribuției unităților elementare. Un cluster moare dacă este părăsit de către toate unitățile sale. 2. Fiecare unitate sau agent are opțiunile de inovație *vs* anexare, adică poate respectiv produce apariția unui nou cluster sau se poate anexa unui cluster deja existent cu o probabilitate dependentă de dimensiunea clusterului. “Probabilitatea de inovație” este funcție de doi parametri, ponderea inovației θ şi dimensiunea clusterului ν care corespundea numărului instantaneu de unități în cluster în momentul de timp precedent. Dimensiunea (virtuală) a populației la fiecare pas corespunde secvenței (n, n-1, …, n-m, n-m+1, …, n-1, n). În caz general $n-m \le v \le n-1$. Probabilitatea că un agent va forma un cluster nou este $u = \dfrac{\theta}{\theta + v}$. Astfel în cazul limită $m=n$, când la fiecare pas toate unitățile părăsesc clusterii prezenți şi apoi se regrupează în secvență, prima apariția a unui nou cluster are loc când v=0, astfel încât toți agenții “acționează rațional”, iar ultimul cluster este format pentru v=n-1, când ponderea "efectului de grup" (*herding*) este maximală. Doar în cazul tranzițiilor unitare, m=1, când toți clusterii nou-formați apar la v=n-1, dependența lui u de v este constantă. 3. Distribuția de echilibru este independentă de m. 4. Lanțul marginal este proiecția dinamicii colective din sistem pe un singur *site*. Notând cu $X_s=i$ numărul de ocupare pe site după s pasuri, iar tranzițiile unitare prin $\{w(i,j) \equiv P(X_{s+1}=j \mid X_s=i), i, j=0,1,...,n\}$, elementele diferite de zero ale matricii probabilității de tranziție sunt [39]:

$$w(i,i+1)=\frac{n-i}{n}\frac{i}{\theta+n-1}, \; i=1,...,n, \tag{9.3a}$$

$$w(i,i-1)=\frac{i}{n}\frac{\theta+n-i}{\theta+n-1}, \; i=1,...,n, \tag{9.3b}$$

$$w(i,i)=1-w(i,i+1)-w(i,i-1), \; i=1,...,n \tag{9.3c}$$

$$w(0,1)=\frac{1}{g-E_n(k)}\frac{\theta}{\theta+n-1} \tag{9.3d}$$

$$w(0,0)=1-w(0,1) \tag{9.3e}$$

Ecuațiile (9.3a)−(9.3c) caracterizează site-urile ocupate, adică pentru numărul de ocupare i>0. Toate site-urile exterioare sunt comasate în unul singur (termostat) cu ponderea inițială θ şi numărul de ocupare n-i. Ecuațiile (9.3d) şi (9.3e) sunt pentru cazul când site-ul este inițial neocupat (i=0). $E_n(k)$ reprezintă numărul mediu de clusteri în sistem. Distribuția de echilibru este descrisă de $P(.)$ [39]:

$$\begin{cases} P(i)=\dfrac{\theta}{g\,i}\dfrac{\theta^{[n-i]}/(n-i)!}{\theta^{[n]}/n!}, \; i=1,...,n \\ P(0)=1-\displaystyle\sum_{i=1}^{n}P(i) \end{cases}, \tag{9.4}$$

iar în conformitate cu ecuația (9.1) numărul mediu de clusteri cu dimensiunea i când numărul populației este n:

$$E(z_i)=\frac{\theta}{i}\frac{\theta^{[n-i]}/(n-i)!}{\theta^{[n]}/n!}. \tag{9.5}$$

Pentru n>>1, folosind aproximația Stirling, se obține:

$$E_n(z_i) \approx \frac{\theta}{i}(1-\frac{i}{n})^{\theta-1}, \tag{9.6}$$

iar numărul mediu de clusteri $E_n(k)=\sum_{i=1}^{n}E_n(z_i)=\sum_{i=0}^{n-1}\frac{\theta}{\theta+i}\approx\theta\ln\frac{n-1+\theta}{\theta}+\gamma,$ unde γ=0.5772... este constanta Euler. În particular, raportul $E_n(z_i)/E_n(k)$ ar putea fi interpretat ca fracțiunea de timp petrecută de un cluster în dimensiunea *i*. Când θ=1, $E(z_i)=1/i$ şi se obține o dependență la puterea -1 a distribuției. În sisteme reale, însă, θ>>1 [40, 41], iar numărul mediu de clusteri se aproximează conform ecuației $\theta\ln\frac{n}{\theta}$, de unde reiese că dependența respectivă se păstrează doar pentru sisteme de dimensiuni mici, dar nu şi în cazul unui număr mare de unități *n* [42].

În cazul limită, se va nota în relațiile (9.3) $w(i,i+1)=\theta(i)$ şi $w(i,i-1)=\phi(i)$, unde $\theta(i)=\frac{n-i}{n}\frac{i}{n-1+\theta}\approx\frac{i(n-i)}{n^2}$ şi $\phi(i)=\frac{i}{n}\frac{n-i+\theta}{n-1+\theta}\approx\frac{i(n-i)}{n^2}+\frac{i\theta}{n^2}$. În continuare se poate estima $E(i)=\theta(i)-\phi(i)\approx-\frac{i\theta}{n^2}$,

$\sigma^2(i)=\theta(i)+\phi(i)-(\theta(i)-\phi(i))^2\approx 2\frac{i(n-i)}{n^2}+\frac{i\theta}{n^2}\approx 2\frac{i(n-i)}{n^2}$.

Introducând notația $x\equiv\frac{i}{n}$, se obține $E(i)\approx-\frac{\theta}{n}x$ şi $\sigma^2(i)\approx 2x(1-x)$ sau $E(x)=\frac{E(i)}{n}\approx-\frac{\theta}{n^2}x$ şi $\sigma^2(x)\approx\frac{2x(1-x)}{n^2}$. Definind intervalul de timp $\tau=\frac{1}{n^2}$, parametri noi ai problemei devin $\mu(x) = -\theta x$ şi $\sigma^2(x) = 2x(1-x)$. Calculând integrala $\int \frac{2\mu(x)}{\sigma^2(x)} dx = -\int \frac{\theta}{1-x} dx = \theta \ln(1-x)$ şi notând $s(x) = Exp[-\theta \ln(1-x)] = (1-x)^{-\theta}$, $\left(s(x)\sigma^2(x)\right)^{-1} = (2x(1-x)(1-x)^{-\theta})^{-1} \equiv k\, x^{-1}(1-x)^{\theta-1}$, se obține distribuția de echilibru $P(x) = \theta\, x^{-1}(1-x)^{\theta-1}$ care corespunde cazului limită (9.6).