A gentle introduction to combinatorial stochastic processes I

Enrico Scalas

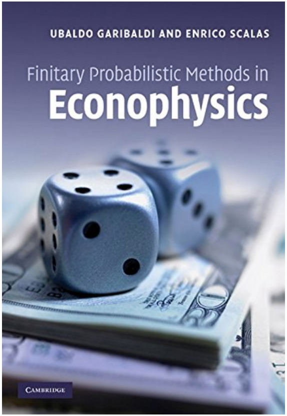
Department of Mathematics, University of Sussex, UK

Stochastic Models for Complex Systems
10 June - 7 July, 2021

Outline

Monograph

Ubaldo Garibaldi and Enrico Scalas, *Finitary Probabilistic Methods in Econophysics*, Cambridge University Press, 2010.



The genealogy of ideas

Here are some chains of influencers (direct influencers in red):

A Direct path (learning live):
(Hume) $\rightarrow$ Johnson $\rightarrow$ (Keynes) $\rightarrow$ Carnap $\rightarrow$ Costantini and Garibaldi $\rightarrow$ Scalas

Two Indirect paths (reading):
Hume $\rightarrow$ Laplace $\rightarrow$ Johnson $\rightarrow$ Carnap $\rightarrow$ Zabell $\rightarrow$ Scalas
Ramsey and de Finetti $\rightarrow$ $\rightarrow$ Scalas

A *Physical* path:
Boltzmann $\rightarrow$ $\rightarrow$ Bach $\rightarrow$ Costantini and Garibaldi $\rightarrow$ Scalas

An *Economics* path:
(Keynes) $\rightarrow$ $\rightarrow$ Aoki and Yoshikawa $\rightarrow$ Scalas

The Wittgenstein of *Tractatus* is in the background.

The problem

We have n objects divided into g categories (or classes). How do we describe this state of affairs? The answer is that we can use *individual descriptions*, *statistical descriptions* and *partitions*.

Facts and propositions

We want to explain how to describe the state of affairs in which for every object listed “alphabetically” or in a sampling order, its category is given. We can consider the descriptions below as *facts* (taking place or not), and the *events* of probability theory as propositions (true or not) about facts (taking place or not).

Objects and categories I

Let us consider a set of n elements $U = \{u_1, u_2, \ldots, u_n\}$ representing a finite population of n physical entities, or physical elements. The symbol $\#$ is used for the *cardinality* of a set, that is $\#U = n$ denotes that the set U contains n distinct elements. A *sample* of U is any subset $S \subseteq U$, i.e. $S = \{u_{i_1}, u_{i_2}, \ldots, u_{i_m}\}$. The size of S is its cardinality, i.e. $\#S = \#\{u_{i_1}, u_{i_2}, \ldots, u_{i_m}\} = m \leq n$.

Objects and categories II

It is possible to use quite complicated notations. For the sake of simplicity, let us denote each object with a numerical label from 1 to n and each category with a numerical label from 1 to g. We are considering finite (or denumerable) collections of objects and categories, therefore this mapping is always possible.

Individual descriptions

Let us introduce a set of variables $X_1, \ldots, X_n$. With the notation $X_2 = 3$ we mean the fact that the object labeled by 2 belongs to the category labeled by 3. The corresponding proposition is denoted by $\{X_2 = 3\}$. If we know to which category each object belongs (i.e. we know the value of $X_1, \ldots, X_n$), we have full information on the system of n objects.

Statistical descriptions

Let us introduce a set of variables $Y_1, \ldots, Y_g$. With the notation $Y_3 = 1$ we mean the fact that the category labeled by 3 contains 1 object. The corresponding proposition is denoted by $\{Y_3 = 1\}$. If we know the values of $Y_1, \ldots, Y_g$, we know how many objects belong to each category, but we do not know exactly which objects belong to which category. We have the constraint

$$\sum_{i=1}^{g} Y_i = n. \tag{1}$$

Partitions

Let us introduce a set of variables $Z_0, \ldots, Z_n$. With the notation $Z_1 = 2$ we mean the fact that there are 2 categories with 1 object. The corresponding event is denoted by $\{Z_1 = 2\}$. If we know the values of $Z_0, \ldots, Z_n$, we know the number of categories with 0 objects, 1 object, etc., but we lose information on exactly which categories have 0 objects, 1 object, etc.. We have the constraints

$$\sum_{i=0}^{n} Z_i = g, \tag{2}$$

$$\sum_{i=0}^{n} i Z_i = \sum_{i=1}^{n} i Z_i = n. \tag{3}$$

Number of descriptions

The number of possible individual description is

$$W(\mathbf{X}|n, g) = g^n.$$

The number of possible statistical description is

$$W(\mathbf{Y}|n, g) = \binom{n+g-1}{n} = \binom{n+g-1}{g-1}.$$

There is no closed formula for the number of possible partitions $W(\mathbf{Z}|n, g)$.
The number of individual descriptions compatible with a given statistical description $\mathbf{Y} = (Y_1 = n_1, \ldots, Y_g = n_g) = \mathbf{n}$ is

$$W(\mathbf{X}|\mathbf{n}, n, g) = \frac{n!}{n_1! \cdots n_g!}.$$

The number of statistical descriptions compatible with a given partition $\mathbf{Z} = (Z_0 = z_0, \ldots, Z_n = z_n) = \mathbf{z}$ is

$$W(\mathbf{Y}|\mathbf{z}, n, g) = \frac{g!}{z_0! \cdots z_n!}.$$

Five firms in a small town

Assume we have five firms in a small town belonging to three economic sectors. Let $X_1 = 1, X_2 = 3, X_3 = 1, X_4 = 2, X_5 = 3$ denote the following state of affairs: firms 1 and 3 belong to sector 1, firms 2 and 5 belong to sector 3 and firm 4 belongs to sector 2. What can we say on statistical descriptions and partitions? It is immediate to see that $Y_1 = 2, Y_2 = 1, Y_3 = 2$. As for partitions, one has $Z_0 = 0, Z_1 = 1, Z_2 = 2, Z_3 = 0, Z_4 = 0, Z_5 = 0$. One can easily verify that the constraints are satisfied.

Five particles and three energy levels

Assume that we have five particles that can be in three energy levels. Let Let $X_1 = 1, X_2 = 3, X_3 = 1, X_4 = 2, X_5 = 3$ denote the following state of affairs: particles 1 and 3 are in energy level 1, particles 2 and 5 are in energy level 3 and particle 4 in energy level 2. Again one has that that $Y_1 = 2, Y_2 = 1, Y_3 = 2$. and $Z_0 = 0, Z_1 = 1, Z_2 = 2, Z_3 = 0, Z_4 = 0, Z_5 = 0$.

We can use this language to describe facts (and propositions) in all disciplines.

Samples I

Consider $n = 4$ agents and $g = 3$ strategies (e.g. 1:= "bull", optimistic; 2:="bear", pessimistic; 3:="neutral", normal), and assume that the individual description is $\mathbf{X} = (X_1 = 1, X_2 = 2, X_3 = 3, X_4 = 2)$ and the corresponding statistical description is $\mathbf{Y} = \mathbf{n} = (n_1 = 1, n_2 = 2, n_3 = 1) = (1, 2, 1)$. Now two agents are drawn without replacement (for a pizza with a friend).

Samples II

i) *List all sample descriptions* $(\varsigma_i)_{i=1,\dots,4}$.
They are:

$$\{(1,2,2,3),(1,2,3,2),(1,3,2,2),(2,1,2,3),(2,1,3,2),(2,2,1,3),(2,2,3,1),(2,3,1,2),(2,3,2,1),(3,1,2,2),(3,2,1,2),(3,2,2,1)\};$$

Their number is $W(\varsigma|\mathbf{n}) = \frac{n!}{n_1!n_2!n_3!} = \frac{4!}{1!2!1!} = 12.$

Samples III

ii) *Write all possible descriptions of the reduced sample* $(\varsigma_i)_{i=1,2}$. If the rule is applied and the first two digits of each permutation are taken, one gets some repetitions: {(1, 2), (1, 2), (1, 3), (2, 1), (2, 1), (2, 2), (2, 2), (2, 3), (2, 3), (3, 1), (3, 2), (3, 2)}, so that the possible description are {(1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2)}, and this is due to the fact that only “bears” can appear together, as the draw is without replacement. It is useful to note that (1, 2), (2, 1), (2, 2), (2, 3), (3, 2) can appear twice as the initial part of the complete sequences, while (1, 3), (3, 1) can appear only once.

Samples IV

iii) *write all possible frequency vectors for the reduced sample and count their number.*
$\mathbf{n} = (Y_1 = n_1 = 1, Y_2 = n_2 = 2, Y_3 = n_3 = 1)$ as there are 1 bull, 2 bears and 1 neutral. The possible sample descriptions are (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), and (3, 2).
Here (1, 2), (2, 1) belong to $\mathbf{m} = (1, 1, 0)$, and

$$W(\varsigma|\mathbf{m},\mathbf{n}) = \frac{m!}{m_1!\cdots m_g!}\frac{(n-m)!}{(n_1-m_1)!\cdots(n_g-m_g)!} = \frac{2!}{1!1!0!}\frac{2!}{1!1!1!} = 4; \qquad (4)$$

(1, 3), (3, 1) belong to $\mathbf{m} = (1, 0, 1)$, and $W(\varsigma|\mathbf{m},\mathbf{n}) = \frac{2!}{1!0!1!}\frac{2!}{0!2!0!} = 2$;
(2, 2) belongs to $\mathbf{m} = (0, 2, 0)$, and $W(\varsigma|\mathbf{m},\mathbf{n}) = \frac{2!}{0!2!0!}\frac{2!}{1!0!1!} = 2$;
(2, 3), (3, 2) belong to $\mathbf{m} = (0, 1, 1)$, and $W(\varsigma|\mathbf{m},\mathbf{n}) = \frac{2!}{0!1!1!}\frac{2!}{1!1!0!} = 4$. Note that $\sum_{\mathbf{m}} W(\varsigma|\mathbf{m},\mathbf{n}) = W(\varsigma|\mathbf{n}) = 12$, where the sum is on the set $\{\mathbf{m} : m_i \le n_i, \sum m_i = 2\}$.

n coins, distributed over g agents

Consider a system of n coins, distributed over g agents. Supposing that each coin is labelled, for a given coin, the individual description $\mathbf{X} = (j_1, j_2, \ldots, j_n)$, with $j_i \in \{1, \ldots, g\}$ tells one to which agent the i-th coin belongs; the frequency vector of the elements, $\mathbf{n} = (n_1, \ldots, n_g)$, is the *agent description* and gives the number of coins (the wealth) of each agent; finally, the partition vector $\mathbf{z} = (z_0, \ldots, z_n)$ describes the number of agents with $0, 1, \ldots, n$ coins and is commonly referred to as the *wealth distribution*; however, it is a description (a fact or event) and it should not be confused with a probability distribution.

Exercise 1

Consider a system of five "energy elements" (whose value is ε) allocated into three molecules.

1. Is it consistent to assume that the partition vector is $(Z_0 = 1, Z_1 = 0, Z_2 = 1, Z_3 = 1, Z_4 = 0, Z_5 = 0)$?
2. What is its physical meaning?

Exercise 2

Repeat the analysis of the Example on sampling for $n = 3$ and $g = 3$, the following individual description, $\mathbf{X} = (1, 2, 3)$ and a sample of 1 element.