

Ireneusz Krech

Awaiting the series of colors – stochastic graph as a means of mathematical treatment and argumentation

Abstract. This paper presents methods of calculating the probability of events related to Markov's homogeneous chains theory. The main means of argumentation being the stochastic graph.

The method of some events probability calculations connected with a particular sampling experiment having a random number of stages is presented in this work. The presented argumentation is based on a stochastic graph, but the reasoning is carried out in the probabilistic space generated by this graph.

The construction of a probabilistic space as a model of a certain multi-stage sampling experiment (see [7], pages 24–25) is a form of mathematical treatment. The argumentation apply to probabilities in this space.

Let us consider a box U_n containing r balls of n different colors: $k_1, k_2, \dots, k_n$ ($n \leq r$). The results of sampling a balls from box U_n create a set $\Omega = \{k_1, k_2, \dots, k_n\}$. Let u_j denote the probability of sampling a ball of color k_j from box U_n ($j \in \{1, 2, \dots, n\}$). Function p meets the following conditions: $p(k_1) = u_1 > 0$, $p(k_2) = u_2 > 0$, $\dots$, $p(k_n) = u_n > 0$ is the probability distribution on the set Ω , and therefore the pair (Ω, p) is the probabilistic space. This is the probabilistic model of sampling a ball from box U_n . Sampling a ball from box U_n we call a *trial* and denote it by $d_{u_1, \dots, u_n}$ or by d_n if $u_1 = u_2 = \dots = u_n = \frac{1}{n}$. The elements of the set Ω are called *colors*. A trial is defined unequivocally by the n -terms sequence $(u_1, u_2, \dots, u_n)$ of rational numbers from the interval $(0, 1)$, the sum of which is equal to 1, where $u_j = p(k_j)$.

The result of m -time recurrence of trial $d_{u_1, \dots, u_n}$, i.e. any m -terms arrangement of the set $\{k_1, k_2, \dots, k_n\}$, is called *series of colors* or *flag* and is denoted by α . Number m is called *the length of series of colors* and is denoted by $|\alpha|$. If $t < m$, the sub-sequence t of subsequent ending elements of series of colors α is called its *ending of length t* .

We say that flag α_1 is contained in flag α_2 , what is denoted by $\alpha_1 \subset \alpha_2$, if α_1 is a sub-sequence of the consecutive terms of the sequence α_2 . If flag α_1 is

not a sub-sequence of the consecutive terms of sequence α_2 we say that flag α_1 is not contained in flag α_2 , which is denoted by $\alpha_1 \not\subset \alpha_2$.

Let $\alpha_1, \alpha_2, \dots, \alpha_k$ be fixed flags of $|\alpha_j| = n_j$ so that for any $l, j \in \{1, 2, \dots, k\}$ and $l \neq j$ the following condition is satisfied:

$$\alpha_l \neq \alpha_j \wedge [(|\alpha_l| \neq |\alpha_j|) \Rightarrow (\alpha_l \not\subset \alpha_j \wedge \alpha_j \not\subset \alpha_l)].$$

Repeating trial $d_{u_1, \dots, u_n}$ for such a long time that:

- the results of n_1 last trials create flag α_1 ,
- or the results of n_2 last trials create flag α_2 ,

⋮

- or the results of n_k last trials create flag α_k ,

is called *awaiting k flags* and is denoted by $d_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$ (or by $d_{\alpha_1, \dots, \alpha_k}^n$, if $u_1 = u_2 = \dots = u_n = \frac{1}{n}$).

Any result of the experiment $d_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$ (as an elementary event) is such a sequence ω , that:

- $\omega \in \{1, 2, \dots, n\}^m$, where $m \geq \min\{n_1, n_2, \dots, n_k\}$ and
- either the ending of length n_1 of sequence ω creates flag α_1 ,
- or the ending of length n_2 of sequence ω creates flag α_2 ,

⋮

- or the ending of length n_k of sequence ω creates flag α_k ,

and none other sub-sequence of the previous terms of sequence ω creates any flag of $\alpha_1, \alpha_2, \dots, \alpha_k$.

The set of all possible results of the experiment $d_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$ is denoted by $\Omega_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$.

Let $J_j(\omega)$ denotes the number of terms equal to k_j in the sequence ω , where $j = 1, 2, \dots, n$. We define the function $p : \Omega_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n} \rightarrow R$ by the formula:

$$p_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}(\omega) = u_1^{J_1(\omega)} \cdot u_2^{J_2(\omega)} \cdot \dots \cdot u_n^{J_n(\omega)}.$$

Function $p_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$ is the probability distribution on the set $\Omega_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$. The pair $(\Omega_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}, p_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n})$ is the probabilistic space that we consider as a probabilistic model of the sampling experiment $d_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$.

Let us associate the below events with the sampling experiment $d_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$: $A_j = \{\text{awaiting } d_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n} \text{ will result with flag } \alpha_j\}$, for $j = 1, 2, \dots, n$. The event A_j we denote by $\{\dots \alpha_j\}$ and its probability — by $P(\dots \alpha_j)$.

Let us consider box U_3 containing three balls of colors: white, black and green. There are only three possible results of sampling a ball from this box:

- b : the sampled ball will be white,
- c : the sampled ball will be black,
- z : the sampled ball will be green.

Sampling of a ball from box U_3 is the trial d_3 . Composed sampling with return of the ball back to box U_3 until black ball will be sampled two times running or until black ball will be sampled after the white one is awaiting one of two flags: $\alpha_1 = cc$ and $\alpha_2 = bc$ or in other words: it is the sampling experiment d_{cc-bc}^3 .

Let us distinguish the following states in the experiment d_{cc-bc}^3 : the initial state s and the states c, b, cc, bc . The procedures described below lead up to the *stochastic graph of the sampling experiment* d_{cc-bc}^3 . Every state of the experiment is interpreted as a point of the plane and is called a *node of the graph*. The node of a graph representing state s is denoted by $\textcircled{s}$. The node of a graph representing state j different from s is denoted by $\boxed{j}$. Let $p_{j \rightarrow k}$ denotes the probability that the experiment d_{cc-bc}^3 would be in state k after a given sampling if after the previous sampling it was in state j . If $p_{j \rightarrow k} > 0$ then we join node $\boxed{j}$ with node $\boxed{k}$ by a straight or curved arrow. This arrow is called *edge of the graph* and it is denoted by $j \rightarrow k$. Close to each edge of the graph we write down:

— the result of the trial after which awaiting would be in state k if after the previous trial it was in state j , if $u_1 = u_2 = \dots = u_n = \frac{1}{n}$ or

— number $p_{j \rightarrow k}$ if $j, k \in \{1, 2, \dots, n\}$ exist such that $u_j \neq u_k$.

If $p_{j \rightarrow j} > 0$ then the edge $j \rightarrow j$ is called a *loop*. If $p_{j \rightarrow j} = 1$ then we neglect loop $j \rightarrow j$ and node $\boxed{j}$ is called a *peripheral node*. The set of all peripheral nodes is called the *periphery of the graph*.

Any sequence of edges such that the beginning of the first one is node $\textcircled{s}$, the end of the last edge is a peripheral node and in case of any two other sequential edges the beginning of it is at the same time the end of the next one (excluding the last edge) is called a *trace*. The *weight of a trace* is the product of numbers associated to its sequential edges.

Any sequence of edges such that the beginning of the first one is node $\boxed{j}$, the end of the last edge is a peripheral node and in case of any two other sequential edges the beginning of it is at the same time the end of the next one is called a *cycle* and we denote it by c_k (k is an index ordering the set of cycles). The number of edges creating cycle c_k is called the *length of the cycle* and is denoted by $r(c_k)$. A loop is considered as a cycle of length 1. The *weight of a cycle* is the product of numbers associated to its edges. The nodes being the beginnings or endings of a cycle's edges are called the *inner nodes of the cycle*. If $j, k_1, \dots, k_n$ are the inner nodes of a cycle and $p_{j \rightarrow k_1} > 0$, $p_{k_n \rightarrow j} > 0$ and $p_{k_l \rightarrow k_{l+1}} > 0$ for $l = 1, 2, \dots, n-1$, such a cycle we denote by $j \rightarrow k_1 \rightarrow \dots \rightarrow k_n \rightarrow j$. If a sequence of edges of a trace creates a cycle we say that this trace *contains a cycle*. Otherwise we say that the trace *does not contain a cycle*.

Figure 1 presents the stochastic graph of the sampling experiment d_{cc-bc}^3 (see [7] and [1]).

Let us consider set Ω_{cc-bc}^* of all traces on the stochastic graph of the sampling experiment d_{cc-bc}^3 . Function p_{cc-bc}^* assigning the weight to each trace from the set Ω_{cc-bc}^* is the probabilistic distribution on set Ω_{cc-bc}^* . The pair $(\Omega_{cc-bc}^*, p_{cc-bc}^*)$ is the probabilistic space called *the space induced by the stochastic graph*.

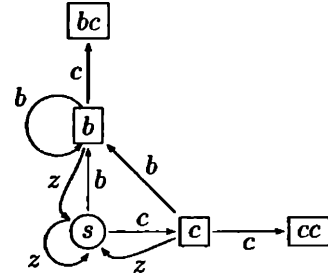


Figure 1.

Probabilistic spaces $(\Omega_{cc-bc}^*, p_{cc-bc}^*)$ and $(\Omega_{cc-bc}^3, p_{cc-bc}^3)$ are isomorphic (see [8], page 32). Therefore we assume probabilistic space $(\Omega_{cc-bc}^*, p_{cc-bc}^*)$ as a probabilistic model of sampling experiment d_{cc-bc}^3 . The following deductions and calculations are carried out in such probabilistic space.

In the probabilistic space $(\Omega_{cc-bc}^*, p_{cc-bc}^*)$ event $\{\dots\alpha\}$ is a set of all traces leading to the node $\boxed{\alpha}$ on the graph in Figure 1, $P(\dots\alpha)$ is the sum of all weights of traces leading to node $\boxed{\alpha}$ ($\alpha \in \{cc, bc\}$).

Let A be a set of traces leading to node $\boxed{bc}$ that do not contain any cycle. By $w(A)$ we denote the weight of traces belonging to set A . Let B be a set of traces leading to node $\boxed{bc}$ that contain at least one cycle. By $w(B)$ we denote the weight of traces belonging to set B . Here we have:

$$\{\dots bc\} = A \cup B \quad \text{and} \quad A \cap B = \emptyset.$$

Because $P(\dots bc)$ is the sum of weights of traces leading to node $\boxed{bc}$, therefore:

$$P(\dots bc) = P(A) + P(B) = w(A) + w(B). \quad (1)$$

Let C denote a set of stochastic graph cycles of sampling experiment d_{cc-bc}^{cc-bc} containing the starting node $\textcircled{s}$. Here we have:

$$C = \{s \rightarrow s, s \rightarrow b \rightarrow s, s \rightarrow c \rightarrow s, s \rightarrow c \rightarrow b \rightarrow s, s \rightarrow b \rightarrow b \rightarrow s, s \rightarrow b \rightarrow b \rightarrow b \rightarrow s, \\ \dots, s \rightarrow c \rightarrow b \rightarrow b \rightarrow s, s \rightarrow c \rightarrow b \rightarrow b \rightarrow b \rightarrow s, \dots\}.$$

C is an infinite but countable set, therefore we can enumerate its elements. Let $C = \{c_1, c_2, \dots, c_j, \dots\}$. Let c_n be any fixed cycle and let $w(c_n)$ be its weight. Now let us consider a set B_{c_n} made up of those traces leading to node $\boxed{bc}$, that have $r(c_n)$ of the beginning edges creating cycle c_n . Let $w(B_{c_n})$ denote

a sum of weights of all traces belonging to set B_{c_n} . Set B_{c_n} is equinumerous to the set $\{\dots bc\}$. Let $x = P(\dots bc)$. Therefore:

$$w(B_{c_n}) = w(c_n) \cdot x. \quad (2)$$

From

$$\bigcup_{c_n \in C} B_{c_n} = B \quad \text{and} \quad \forall c_j, c_k \in C [B_{c_j} \cap B_{c_k} = \emptyset], \quad (3)$$

it follows that

$$P(B) = \sum_{c_n \in C} w(B_{c_n}). \quad (4)$$

From equations (1) and (4) it results that

$$P(\dots bc) = w(A) + \sum_{c_n \in C} w(B_{c_n}).$$

If we take into consideration equation (2) in the above equation then we have

$$x = w(A) + \sum_{c_n \in C} w(c_n) \cdot x = w(A) + x \cdot \sum_{c_n \in C} w(c_n).$$

So we get the equation

$$x = w(A) + x \cdot \sum_{c_n \in C} w(c_n).$$

After some transformations we have

$$x = \frac{w(A)}{1 - \sum_{c_n \in C} w(c_n)}. \quad (5)$$

Because

$$\sum_{c_n \in C} w(c_n) = \frac{5}{6} \quad \text{and} \quad w(A) = \frac{1}{9},$$

therefore

$$P(\dots bc) = \frac{\frac{1}{9}}{1 - \frac{5}{6}}.$$

Finally we get: $P(\dots bc) = \frac{2}{3}$ and $P(\dots cc) = 1 - P(\dots bc) = \frac{1}{3}$.

The method presented in this work allows for probability calculation of events of type $\{\dots \alpha_j\}$ in the case the graph of the sampling experiment $d_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$ has cycles with node ⑤ being one of inner nodes. This method enables us to find the probability of an event of type $\{\dots \alpha_j\}$ on the basis of

sampling experiment's stochastic graph. Thereby this method may be used in case of any sampling experiment being the Markov chain and satisfying the above assumption.

The problem of drawing up a method of calculating the probability of events of type $\{\dots\alpha_j\}$ in the case of any sampling experiment $d_{\alpha_1, \dots, \alpha_k}^{u_1, \dots, u_n}$ is still open.

References

- [1] A. Engel, *Wahrscheinlichkeitsrechnung und Statistik*, Band 1, Ernst Klett Verlag, Stuttgart 1980.
- [2] M. Gardner, *Time Travel and Other Mathematical Bewilderments*, W.H. Freeman and Company, New York 1988.
- [3] R. Graham, D. Knuth, O. Patashnik, *Matematyka konkretna*, Wydawnictwo Naukowe PWN, Warszawa 1996.
- [4] I. Krech, *Probability in probability spaces connected with generalised Penney's games*, Acta Universitatis Purkynianae 42, Usti nad Labem 1999, 71-77.
- [5] I. Krech, *Osobliwe własności modeli probabilistycznych czekania na serie sukcesów i porażek*, Ann. Acad. Paed. Cracov. 5 Studia Ad Calculum Probabilitatis Eiusque Didacticam Perinentia 1(2002), 39-55.
- [6] A. Płocki, *Gry Penneya i paradoksy stochastyczne*, Matematyka 1 (1999), 12-18.
- [7] A. Płocki, *Stochastyka 1. Rachunek prawdopodobieństwa i statystyka matematyczna jako matematyka in statu nascendi*, Wydawnictwo Naukowe WSP, Kraków 1997.
- [8] M. Major, B. Nawolska, *Matematyzacja, rachunki, dedukcja i interpretacja w zadaniach stochastycznych*, Wydawnictwo Naukowe WSP, Kraków 1999.
- [9] H. Humenberger, *Ein Paradoxon bei Münzwurfserien und bedingte Erwartungswerte*, Beiträge der 32.Tagung für Didaktik der Mathematik, München 1998.
- [10] H. Humenberger, *Kopf-Adler-Muster in Münzwurfserien, unendliche Reihen und Fibonacci-Zahlen*, Beiträge zum Mathematikunterricht, Verlag Franzbecker, Hildesheim, Berlin 1999, 245-248.

Pedagogical University
 Institute of Mathematics
 Podchorążych 2
 30-084 Kraków
 Poland
 E-mail: ikrech@wsp.krakow.pl