How the probability of a sequence of independent and equally probable events can be dependent of the order of the events

Abstract

It is commonly understood among mathematically skilled individuals that a uniform sequence of independent random events are just as probable to occur as a varying sequence of the same length, if the probabilities of the individual events are equal. However, this is here demonstrated to be a misconception under some circumstances.

1 Introduction

Many mathematically skilled individuals would probably say that any specific sequence of equally likely independent events has the same probability to appear in a certain time interval, independent of the order of the events. This opinion may be related to the fact that a uniform sequence is just as likely to occur as a varying sequence at a specific point in time. For example, the uniform sequence of heads (H) and tails (T) HTHTHTHTHT is just as probable as the varying sequence HTTHHHTHTT if a fair coin is flipped ten times.

On the contrary, many mathematically naive persons would erroneously guess that it is less likely to experience the uniform sequence than the varying sequence. This mistake is presumably based on the fact that orderly *types* of sequences are less probable than disordered ones, and that most people therefore would be more surprised if they experienced the former type. The mathematically skilled individual still knows that the order of the events do not have any relevance in this case.

Nevertheless, in some cases the probability that a certain sequence of independent events occur *actually is* dependent of the order of the events. The aim of the present paper is to investigate under which circumstances this is the case.

2 Probability of a random sequence as dependent on the order of equally likely events

The following problem will be the subject of the analysis:

> Assume that you sit at a roulette table one evening and only keep track of the red (R) och black (B) outcomes. Then, which of the following specified sequences of fifteen consecutive games do you have the greatest probability to experience at least once during that evening – RRRRRRRRRRRRRRR or BRBBRBRRRBBRBRR (subsequently named s_{homo} and s_{vary}, respectively). If one of them is more likely, how much more likely is it than the other one?

It is important to see what separates this problem from the seemingly similar coin flipping problem in section 1. On a closer inspection, the crucial difference seems to be that the roulette case concerns sequences that appear "at least once during that evening" while the coin flipping case concerns sequences that appear "at a specific point in time".

At any point in time (after the first fourteen games) there is a specific fifteen-games sequence of red (R) and black (B) that most recently has been completed. Let p denote the probability of an arbitrary specific fifteen-games sequence to be this most recently completed sequence. Then p is equal for all possible sequences and has the value $(18/37)^{15}$ or $(18/38)^{15}$, dependent on if there are one or two zeros. If we assume that N is the average number of games during one evening at the casino, the average number of any specific combination of fifteen black and red outcomes during a casino evening is $p \cdot (N-14)$.[1] This average number is consequently equal for the uniform and the varying sequence. However, it must also be taken into account how many sequences of each kind that can expected to appear *at the same evening.*

First, the average number for the varying sequence to appear during an evening will be calculated. Let n denote the number of s_{vary} sequences, let z denote the number of zeros (which is either one or two) and let

$$f(n, N, z) = p(\text{at least } n \; s_{\text{vary}} | N \text{ games}) \; .$$

The probability that at least n such sequences will appear during a night with N consecutive games then can be expressed as

$$\begin{aligned} f(n, N, z) &= \frac{\binom{N+1-n\cdot 15+n-1}{n} \cdot 18^{15\cdot n} \cdot (36+z)^{N-15\cdot n}}{(36+z)^N} \\ &= \binom{N-n\cdot 14}{n} \cdot \left(\frac{18}{36+z}\right)^{15\cdot n} \; , \end{aligned}$$

where $\binom{N+1-n\cdot 15+n-1}{n}$ denotes how many ways n objects can be placed in $N+1-n\cdot 15$ different positions, where this number multiplied by

[1] There are only $N-14$ possible fifteen-games sequences during an evening with N games.

$18^{15\cdot n} \cdot (36+z)^{N-15\cdot n}$ equals the total number of sequences where the varying sequence appears at least n times during N games, and where $(36+z)^N$ is the total number of possible sequences.

It is now possible to calculate an upper limit for the average number of s_{vary} on the evenings when this varying sequence appears at least one time (the lower limit is of course unity). For this purpose, it is assumed that $N = 1000$, since the average time for a roulette game is more than a half minute, and since a greater N will result in a greater $f(n, N, z)$. Furthermore, it is assumed that $z = 1$, again to maximise the upper limit. The upper limit in question then is

$$\frac{f(1,1000,1)\cdot 1 + f(2,1000,1)\cdot 2 + \cdots + f(66,1000,1)\cdot 66}{f(1,1000,1)}$$
$$= 1 + \frac{\binom{972}{2}\cdot\left(\frac{18}{37}\right)^{15}\cdot 2 + \binom{958}{3}\cdot\left(\frac{18}{37}\right)^{30}\cdot 3 + \cdots + \binom{76}{66}\cdot\left(\frac{18}{37}\right)^{975}\cdot 66}{\binom{986}{1}},$$

since 66 is the maximum number of s_{vary} sequences that can appear during an evening with 1000 consecutive games. In plain numbers this equals approximately term by term

$$1 + 0.01936 + 0.0001819 + \cdots \approx 1.01955\ .$$

Taken together, this implies that a s_{vary} sequence appears on average about 1.02 times on an evening where it appears. This is a consequence of the fact that it is a rare event in the first place, and that one appearance does not facilitate another appearance of the same kind.

In contrast, the uniform sequence, RRRRRRRRRRRRRRR, has the probability 1/2 to appear once more after the first appearance because one more immediately following red outcome is enough to bring about another sequence of fifteen red outcomes after the first fifteen has appeared. In other words, if the first uniform sequence appears from game n to $n+14$, an additional red outcome at game $n+15$ would result in a new uniform sequence from game $n+1$ to $n+15$.

On occasions when two uniform sequences have appeared in the same evening in this way, there is again a probability of 1/2 that it will appear a third time. Thus, the probability of a third uniform sequences to appear at the same evening, given that it has appeared once, is $1/2 \cdot 1/2 = 1/4$. In accordance with a similar line of reasoning, the probability of a fourth uniform sequence at the same evening is 1/8, and so forth. Consequently, given that at least one such sequence has appeared, the total number of uniform sequences to appear during that evening will, on average, approximately be

$$1 + 1/2 + 1/4 + \ldots = 2\ ,$$

since

$$\begin{aligned} s &= 1 + 1/2 + 1/4 + \ldots \\ \Rightarrow 2s &= 2 + 1 + 1/2 + 1/4 + \ldots \\ \Rightarrow s &= 2s - s = 2 \ . \end{aligned}$$

It has already been concluded that the average number of the uniform sequence of fifteen red outcomes during a casino evening is $p \cdot (N - 14)$. Since they appear roughly twice every evening, $p \cdot (N - 14)/2$ is the approximate relative frequency for the uniform sequence RRRRRRRRRRRRRRR to occur on an arbitrary casino evening.

The probability of a sequence to occur on an arbitrary evening equals the relative frequency of its occurance. Therefore, the probability of the sequence BRBBRBRRRBBRBRR is approximately twice as high as the probability of the sequence RRRRRRRRRRRRRRR to be experienced at least once during an evening at a roulette table.

The above line of reasoning can also be applied to the less homogeneous sequence RBRBRBRBRBRBRBR, which only differs from the varying sequence with respect to the order of the events. Given that at least one such sequence appears during an evening at the casino, the total number of such sequences that evening will, on average, approximately be (according to a similar calculation as above):

$$1 + 1/4 + 1/16 + \ldots = 4/3 \ .$$

Since the probability to experience this sequence at least once during an arbitrary casino evening therefore approximately is $p \cdot (N - 14)/(4/3)$, the varying sequence is approximately 4/3 as probable as a the sequence RBRBRBRBRBRBRBR to be experienced at such an event.

In conclusion, the probability to experience a sequence of random events with equal probabilities on a certain occasion is under some circumstances dependent on the order of the events.

3 Discussion

The conjecture that the probability of a sequence of independent and equally probable events is independent of the order of the events seems at first sight to be self-evident. The question is how something that appears to be such an undeniable truth can be false.

At a closer inspection, the reason is that while the probabilities of the individual events, red and black, surely are independent, the probabilities of the sequences are *not* independent on occasions where there can be more than one sequence in the *time interval* in question. A uniform sequence facilitates the emergence of another uniform sequence in the same time interval, while a varying sequence does not do it to the same degree (nevertheless, the sequence RBRBBRBRRRBBBRB, where the "RB" in the beginning also appears at the end, makes another identical sequence a little more likely to appear twice on the same evening).

In section 1 it was presumed that a "mathematically naive person" would guess that a uniform sequence of equally likely events is less probable than a variable sequence of the same length. If this turns out to be true, a mathematically naive person will probably have a better chance to give a correct answer to the first part of the question posed in the beginning of this section 2 than a mathematically skilled person would. At least, this it at risk to happen if the latter person do not abandon one of the commonly occurring prejudices concerning randomness.