Improving Probability Education Through Statistical Experiments

Saeed Al-Hajjar
Mathematical Sciences And Information Technology
Ahlia University
Manama, P.O.Box : 10878
KINGDOM OF BAHRAIN
Saidth2000@yahoo.com , hajjar@ahliauniversity.edu.bh

Abstract: This study analyze some hypothesis on the difficulties facing the teaching of probabilities, and see how models of probability are useful to solve any confusion. This will be clarified by formulating questions that can be addressed with data, collect , organize, and display relevant data to answer them. In this part we will see the solution of the problem ” who will win the million ”, which is a good example of improving probability education through statistical experiments.

Key–Words: Hazard , Confirmed , Change , Probability , Modeling , Random Experiment , Equiprobable

1 Introduction

The probabilities consist of one of the mathematical branches which has an important field of applications in different sciences such as Physics, Biology, Chemistry, Social sciences, Political sciences, Economics,etc. In addition, it plays an important role in our modern society in the fields of Health, Insurance, Assurance, Game etc. Their places in mathematical teaching are becoming more and more important, and the recent reforming programs assure the importance of the utilization of probabilities in the modern societies. The later brings into attention the concern of teaching this field. It is clear that in dealing with the subject of probabilities through schools, teachers and students alike suffer and get confused. They all feel the deception. Thinking of the fundamental probability notions, the process of teaching, the exchange of information with colleagues, the steps for solving the problems are major issues of the personal development and professional teaching.

2 The Bets On The Modeling Of The Probabilities (BERNARD DANTAL[1])

During the TV interview on the works of Paul-Emile VICTOR in Greenland, the ethnologist Claude Levi-Strauss[2] declared: ”the difficulty for an ethnologist is to be once inside and outside. He should live plenty with people when he wants to study the mode of life, but he should also behave as an exterior observer to them, otherwise, he never does ethnology” It seems that the mathematics teacher of the elementary classes is facing difficulties similar to ethnologists, while teaching probabilities. Well, beginning from an observation, then a description of the reality, with successive steps, the teacher should aid his students to be extracted gradually from this reality to construct progressively a mathematical model.

Therefore, this gradual extraction may give rise to many problems to this instructor: firstly, because he already has a footstep in the observation of reality, secondly, in the construction of model.

He does not often perceive what the students of each class might come up with, so he has to be always on toes to simulate a mathematical model depending on the situation he is in. so in this case the following statement holds true: ”probabilities and statistics in high school are not actual mathematics”.

This raises some educational questions: Is it necessary to perceive the fact to generate or construct the model? Isn’t possible to directly state the axioms then construct the model and later on describe the reality that corresponds to it.?

Nowadays one can answer, partly at least these questions. Well, in the program of 1972, the intermediate classes of which is the construction of a model through the observation of reality was tried to be avoided. This trial did not succeed and to a large number of students the contents of learning have no more sense.

However, the concept of probability itself has lost its sense and it is applied on realities, which achieve nothing constructive.

The actual start suggested by the program of probabilities on the intermediate classes, could be translated by Fig.1 that synthesizes the theoretical development presented in the last three parts of this chapter, through the comments observed from a large number of students:

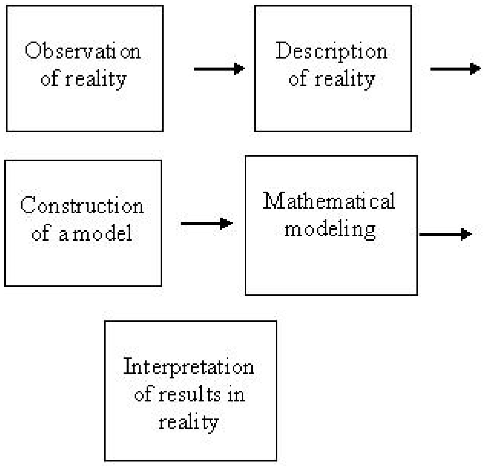

Figure 1. Difficulty in construction between observation of reality and mathematical model

The above schematic strategy was effectively used; the main difficulty is to construct the intermediate step between the observation of reality and the elaborate construction of the mathematical model.

Through these intermediate steps, Can the adopted concept be defined from the mathematical point of view? This question was supposed to be asked during the inter IREM commission concerning the people who works on probability concepts of random variable.

This concept then has raised questions of how to teach:

- If this concept does not simulate the reality, then it is an experiment where the observer cannot predict the result surely before hand. For example, for an individual to be told to go and pass an exam; is it a random variable? One usually says that the result of such experiment is due to hazard in another word he has the chance. However, the use of the word random in a mathematical course sometimes means something different for students. Well, until now, one doesn't establish a mathematics course unless it is concise and clear and proved to be effective. So a random experiment appears as such thing of little importance and does not mean a lot for students. For students, all experiments are considered as a random situation such as the conditions of production, product distribution.
- Nowadays, the process of teaching probabilities, first the model should construct objectively depending on the observed reality (then the question arises who is the observer anyway) two cases might be possible:

First case: For an observer to describe a real experiment under the term random experiment, he should do it after stating the conditions of the experiment:

- The results are not to be presumed.
- Roughly speaking, the possible outcome should be known perceived before hand. The process of eliminating during the modeling process of assumptions that might be proved irrelevant. All models should simulate reality as close as possible.
- He decides a prior, either by observing conditions, or by a total contingency (approach of Pascal and Fermat in the problem of going, subjective position of Laplace), that all possible results are equally probable.

Remarks On The First Case:

1. It is a Modeling, which depends so much on the reality in the case producing games of ”Hazard” easily where the issues are symmetric, so from this remark Pascal introduced the term ”geometry of Hazard”.
2. Unfortunately, this model is restrictive, because the set of possible outcomes should be foreseen a prior, it is on the other hand Inadaptable to describe the reality when the issues of the experiment do not bring a system with equiprobable cases.
3. This approach is conducted to the definition of Fermat and Laplace of the probability, with the formula:

$$P = \frac{\text{number of favorable cases}}{\text{number of possible cases}} \tag{1}$$

Second Case: If an observer decides to describe a real experiment under the term random experiment, what should be done after the conditions of the experiment? The experiment should be reproduced many times under similar circumstances.(who would reproduced the same experimentally exactly).

Remarks On The Second Case:

1. The definition of the term ”similar” should be precise.

2. For an observer, reproduction of the experiment under the same conditions holds true for the conditions he perceives but not for the ones which he does not recognize.
3. It is the historical work of Jacques Bernoulli Ars Conjectandi[3], who conducted to the law of big numbers. This law is some how one of the essential laws to validate the probability model, it is entirely used today to model all domains of reality, deeply the frequency approach of the notion of probability, one can regret that it is not going to be in high schools program any more.

Therefore, in this context one does not have to foresee results. Similarly if one repeats the experiment another time under the same conditions, then this is a random experiment.

However, what does ”under the same conditions” mean?

In reality, there is no way to reproduce physically the same conditions. For example, when one throws a die, one can't throw it again from the same location, with the same velocity, in the same direction and in all ways not at the same times (introducing the relativity theory of Einstien), unless one has two dice and then the conditions are not the same for the two dice, equilibrium, symmetry, weight, velocity, direction, even the movement of electron from the quantum point of viewetc. There exist some differences even imperfections between the two dice. It is the manner where Poincare[4] defines, in certain cases, the hazard is the big sensitivity to initial conditions.

In this sense, the success in an exam, the time needed tomorrow, the determination of sex of a baby are phenomena (real) randomly, that is where the result is due in part of hazard the same if the hazard that intervenes is not always identified properly.

3 WHAT IS A RANDOM EXPERIMENT ?(Jean-Claude GIRARD)

3.1 The Idea of Hazard

Random, synonymously: the uncertain in the future, the intervention of Hazard. (Robert Dictionary)

In common sense, a random proof is simply an experiment along which one can neither foresee nor calculate the result. It is the contrary of a determined experiment, as one faces in physics if the initial conditions were clearly stated then the solution could be reasonably found.

3.1.1 How to apprehend this hazard?

Is it due to the complexity of a system, which is very sensible to initial conditions (in sense of Poincaré).

Isn't that the reflex of our ignorance (in sense of the Laplace)?

Some real experiments could be described and studied using mathematical development with functions, setsetc. for example; results obtained by the laws of Mendel (genetic model) are in agreement with the relative frequencies of appearance of certain genes.

3.2 Random Experiment Real or Model

A mathematical random experiment that is a model of random experiment should verify the following three points:

1. The description of the experiment conditions determine it in a precise and sufficient way to guarantee the unity (we take in consideration the study of the object). One will not have then any problem for repeating this fictions experiment under the same conditions. One would then consider producing the experiments under the precise conditions taken of the same description.

   Most of the times, some hypothesis extracted through the observation of real experiments remain implicit. When one throws a die, one assumes to that the six faces are in equilibrium, without deceiving etc This would permit adopting a model, which attributes the same probability for each face (otherwise, one doesn't know which model to take).

   What does the phrase ”choose randomly” mean? One should often consider this expression as a synonym of drawing out with the same probability for all individuals of the population but there are other ways to ”choose randomly”, for example the angle of approach of picking the two dice with equal or unequal probabilities etc.

   Hypotheses should be explicitly stated as possible in a mathematical random experiment.

   As in every model, this idealistic experiment represents well more or less the reality.

2. One can determine the set of possible issues. This set E is related to the use of the chosen model that is adopted. For example, if one throws a coin, take in general E= $\{Tail, Face\}$ by eliminating the possibility that the coin would be thrown on its edge, to be imperfect or flown! The set E= $\{Tail, Face\}$ is sufficient to model the reality.

Throwing a coin is not considered as a random proof. The consideration of random proof allows us to know about what is the interest in such experiment.

3. One can neither predict nor calculate the result of the experiment.

   One would say that it is the intervention of hazard (with no need to have defined it in advance) which prevents it to be determined, among the possible outcomes; this will be realized during the execution of the experiment in similar conditions. Under the same initial conditions, one does not always obtain the same results.

4 AN EXAMPLE OF CONFUSION MODEL- REALIZATION (Jean-Claude GIRARD[5])

A mathematical model which seems indispensable to talk about is a probability. In practice, model and reality are related in a way that it is difficult to separate one from the other. Here is an example, the national subject of Bacc 1995, scientific series:

A baker produces breads, which should weigh, in theory, 600 grams.Dessignate by X the random variable, which takes the value of possible weights of breads, expressed in grams and rounded to the nearest 10 grams. Table 1 indicates the probability of event $X = Xi$:

Table 1. The probability of the random variable $X = Xi$

$X = Xi$	580	590	600	620
Pi	0.12	0.25	0.27	0.04

Read, for example, the probability that a bread chosen randomly weight 590 grams is 0.25.

1. Calculate the mathematical estimation of X and the standard deviation of X.

2. A customer buys bread. What is the probability that his bread weighs at least 600 grams?

3. A controller of deceived repression service enter the bakery and take, randomly, ten loaves of bread.

   (a) What is the probability to have exactly three loaves of bread of 580 grams?

   (b) What is the probability to have at least a bread of 580 grams?

   (c) What is the probability to have at most a bread of 580 grams?

Give the exact values, then the approximate decimal values to the nearest 10^{-4}.From the first sentence, A real experience is introduced (A baker produces breads) is an idealization of the results (bread should weigh, ”in theory”, 600 grams). Next mathematical representation was introduced through the random variable X .

Mathematically, a random variable could be defined without going back to some reality But here, the only possible definition depends on going back to concrete experiment. Which is it exactly? How did they determine the given probabilities? One can think that it is through statistical analysis of past production and on a large quantity of bread. Are these conclusions still available for future and in particular for the production of the considered day?

The hypothesis will be postulated and this will determines the model. It is then a reality (a large quantity of bread is produced and the distribution of relative frequencies of weights was obtained) which define a model (the law of probability of a random variable x that represents the weight of a bread).

If the first question in the model is clear, in contrary, the second question is taking care of the confusion between reality (a customer buys bread) and Modeling (to calculate a probability, should be in the model). The choice of bread is done ”randomly” (it is not mentioned but one can assume it, and this is already out of hypothesis) and the calculation should be done inside the model that is already given.

This brings us a more accurate problem: Is the experiment which is consisted to choose a bread ”by hazard” in the bakery, similar to that which is considered to produce a bread? The hypothesis should be answered as yes. It is not then the same reality. The baker adopted the first weigh and he can produce bread as much as he likes. When it reaches the customer who is unable to choose except in from what is being offered to him at the moment of buying. Isn’t it dangerous to mix probabilities of different values of the random variable X (which is the model) and the relative frequencies of the weight of the bread in the basket (which is the reality)?

The third question is again held on a reality: a customer chose ten loaves of bread randomly.

What is meant, by this statement, ”chose randomly”? It should often consider this expression as a synonym of election with the same probability to all elements of population.

In reality, one can enumerate bread and draw randomly a number through a table of random number or with the touch RUN of a calculator. One will get an equiprobable model .

Similarly, when one throws a die, one tend to see: on six faces, in equilibrium, without deceiving etc, which suggest to take a model that attributes the same probability for every face (otherwise, one does not know which model to take).

However, there are other cases to "choose randomly", see for example the different methods of the selection of second degree (or more), with equal or unequal probabilities, etc.

In contrary choosing "randomly", a point on a straight line doesn't have a physical reality and any model possible.

The hypothesis of model will become then explicit (as much as possible) in the description of a random variable.

However, this time, in the third question, the model is not given. It is left to the student to model; the only model that the student knows in Bacc II (the binomial law) is not applied here! Well, if caught in such a situation in real life, withdrawing two times the same bread is an not possible.

5 Brief description of the surrogate models)

The exact function of interest (exit total pressure for the application) is $J(\alpha)$. The description of the surrogate models is limited to the case of a two-component design vector α. The number of available exact evaluations of function $J(\alpha)$ is noted n_s. The mean square error (MSE) on the sampling between the exact function $J(\alpha)$ and the surrogate model $J(\alpha)$ is denoted by ε[6].

$$\varepsilon_i = \sum_{i=1}^{n^s} (J(\alpha_1^i, \alpha_2^i) - \overline{J}(\alpha_1^i, \alpha_2^i)^2) \tag{2}$$

The vector of the exact evaluations is J_s.

$$J_s = [J(\alpha_1), ..., J(\alpha_{n_s})]^T \tag{3}$$

6 MODELING CONCEPTION)

Through previous consideration, one can conceptualize the modeling of a real phenomenon that permits us to introduce some important variables, in the following manner:

6.1 Process of Modeling Conception

6.1.1 Step of the modelling :

Reality.

Object of the action: Study of a real phenomenon, or of an experimental processing.Ex.: binomial situation, geometric design of "English flag".

Attended activity : Simplified description of pertinent elements for the posed question Application of an experimental protocol. This description is stated theoretical in appropriate Manner .

6.1.2 Step of the modelling :

Pseudo - concrete model .

Object of the action : Generic situation, decontextualising, removing the abstract properties of the study.

Hypothesis of model:

1. implicit in general

2. Explicit for the particular context

Ex.: Bernoulli's urn and binomial selection, known geometrical figures: parallelogram.

Attended activity : Presentation of model in current terms or conception, rhetoric validation of the analogy with the preceding description. Confrontation of hypothesizes of model with the corresponding elements of the description. Conjectures on properties of the model responding to the question.

6.1.3 Step of the modelling :

Mathematical Model.

Object of the action : Set of equations or mathematical formulation representing the properties of model and the controlled hypothesis.

Ex.: Universe $\Omega = [0; n]$. Binomial variable N and probability $P_k = C_n^k P^k q^{n-k}$ Known configuration and theorem in geometry .

Attended activity : Release equation or formulation: through the understanding of the theoretical pseudo-concrete model and the pertaining laws., presenting mathematical equations between variables in a determined theoretical quadrant. Formulation of the posed question in this quadrant

6.1.4 Step of the modelling :

Mathematical study.

Object of the action : Deriving or stating mathematical theories and deriving appropriate model hypothesis. Ex.: E[N] = np. Medians pass through the center of parallelogram.

Attended activity : Demonstration of main theoretical results pertaining to the mathematical model Stating a preconceived statement to be posed as an objective.

6.1.5 Step of the modelling :

Confrontation Model-reality.

Object of the action : Formulation in concrete terms of obtained results Recontextualization Comparison of the results d-obtained from the model with those of the reality Ex.: the average of success obtained in a large number of comparable binomial selection at np.

Attended activity : Comparison of numerical results or qualitative with the correspondent experimental measures . Evaluation of the error margin and acceptability of model .

6.1.6 Step of the modelling :

Generalization and provisions.

Object of the action : Extension of the model to be used in other similar situations, conditions of generalization. Provisions of results resulted from the new situations. Ex.: Controlling the real value of the probability p Interval of confidence and test of p hypothesis for a real percentage in a sample of population.

Attended activity : The appreciation of the validity and the generalization of model assumes a special knowledge of the case under study. Specialists are able to correlate the conclusions, explications, and generalization resulted from the hypothetical mathematical model study without the need of mathematicians. In a modelling , there are always two simple steps that come from a special knowledge of studied phenomenon:

1. The identification that is the choice between many possible models and the experimental determinations of parameters is considered one way of hypothesizing a model.

2. The validation, that is the evaluation of approximating degree of theoretical results obtained with the corresponding experimental values and the decision that the model whether or not is well suited to the problem at hand.

7 Prevention of understanding related to the Modelling of the reality (or of a fact)

The introduction of probabilities, whatever the level is, creates a higher level of difficulties in comparison to other mathematical disciplines. According to David Ruelle , "a theory of physics is consisted to fit a mathematical theory on a piece of reality. For some of these pieces of reality, there exist idealizations that may be effected directly by probability. The interest of these idealizations is because of their usefulness.

It should be known that one is touching on a ground of mathematics where reality is interesting! One has to find the best model being applied in reality or on what one perceives of the reality before moving to the application. One is never sure about a model to be true. A theory is not applied except in a given limit (hence the invention of different geometric of vague logic, or of non-standard calculus) and a theory is not good unless there is no loop hole in it, and to test such a theory a smaller model of the supper theory is tested; for example the relativity theory when simplified should yield the Newtonian gravitation.

So we are facing this problem when we want to let students comprehend the model related to the experiment which consists to throw two dice and to sum up the appeared numbers of their faces: think to consider the outcomes 6+5 and 5+6 distinct, others identics. The trouble is more deep for students who try to think about many realities (according to the case that the dice have the same colour or not, but this does not change the sum of faces!)

The reason is that one believes to work on the reality, then a model is constructed! There are many possible models but there is only one reality. Maybe such an experiment or a simulation (on computer or with a table of numbers randomly), would give an idea of the model that should be chosen. So the work is not to be carried on solely on the probability but also on the very delicate process of modelling itself.

During the teaching of probability, the modelling process is often not given a great importance. Conditions of the experiments should be defined clearly and without ambiguity while establishing the model, a lot of implicit assumptions are introduced. Due to this fact, exercises seem to be the same to students while in fact they are not.

Finally, the need arises to talk about the physics experiment, the random proof, and the mathematical model of the problem at hand.

The sciences don't try to explain, especially if they try to interpret; They make essentially the models. With a model, a mathematical construction is to be devised to describe the observed phenomenon.

The justification of such mathematical construction occupy uniquely and precisely in the work where they have to be functioned. "John Von Neamam"[7]

In conclusion facing real difficulties in teaching probabilities process, isn't due to the reason that random is difficult to be learnt, and clearly not difficult more than the geometry of the universe that surround us. Why disproof of an approach which has difficulty of logic, even though the concept of probability is constructed as synthesis of different aspects. Moreover, Do not under estimate the conceptual difficulties related to the nature of random and probabilities.

Many philosophers and mathematicians, from Aristote to Rene Tom, had thought about these notions and the debate between random and determinism is not attained. This is not surely a random (in fact!) if it attains the twentieth century cycle to see the appearance of the first bubble of probability calculations and 1930 for the axiomatization of the theory. In comparison, the geometry was already studied largely before our generation and the elements of Euclide dated in fourth century before J.C.

8 The problem: Who would win the million?

This problem is raised as a game along which one would say that luck mostly interfere in choosing an outcome. However, it will be obvious after the following analysis, depending on statistical experiments and trials, that the probability of an interested outcome is somehow independent of luck or coincidence.

8.1 The following procedure should be adopted:

- A question is asked
- Four multiple choices are given, one answer is correct
- The choices are to be proposed
- No idea, which one is right
- Select one of the choices
- Two of the remaining choices will be eliminated
- Now, stay with the one chosen or change to the alternative
- How does the play go after the elimination of the two expressions?
- Confirm the selection
- Change the selection
- Pick randomly (head or tail)
- If the final chosen expression is the right answer, win a high price or etc.

8.2 Random Experiment (Hazard)

- A person has chosen randomly (by hazard) one answer among four expressions along which one corresponds to the right answer.
- Two out of the four expressions that do not correspond to the right answer are eliminated.
- The person rechooses by ”Hazard” his previous choice.

8.2.1 Universe of Possible Equiprobable

Figure 2 describe all possible outcomes of events rechooses by ”Hazard” his previous choice.

T : True and F : False

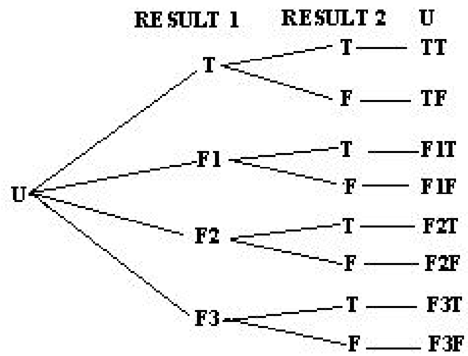

Figure 2. possible outcomes of rechoose by ”Hazard”

Probability to win : 4 / 8 = 1 / 2.

8.3 Random Experiment(Confirmed)

- A person has chosen randomly (by Hazard) one answer among four expressions along which one corresponds to the right answer.
- Two out of the four expressions that do not correspond to the right answer are eliminated.
- The person ”Confirmed ” his choice.

8.3.1 Universe of Possible Equiprobable

Figure 3 describes all possible outcomes of the event ”confirmed” his choice. Probability to win : 1 / 4 .

8.4 Random Experiment (Change)

- A person has chosen randomly (by Hazard) one answer among four expressions along which one corresponds to the right answer.

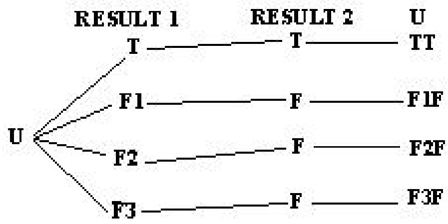

Figure 3. possible outcomes of "confirmed" his choice.

- Two out of the four expressions that do not correspond to the right answer are eliminated.
- The person "Changes" his choice.

8.4.1 Universe of Possible Equiprobable

Figure 4 describes all possible outcomes of the event "changes" his choice.

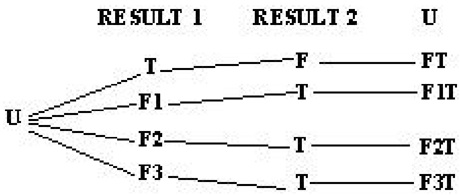

Figure 4. possible outcomes of "changes" his choice.

Probability to win : 3 / 4.

9 Distribution of the results through: Bar graph

figure 5 represents the data distribution in a Bar graph. It is obvious that the probability of the event "change" is more remarkable than that of the events "Hazard" or "confirmed".

10 Conclusion

Probabilities and statistics are twins. One can't separate them in many different fields. When statistical analysis is done, the observation is to be done through the concepts of medium, mean, mode, standard deviation, and so on. These concepts are drawn from a sample but using probabilities can be widely

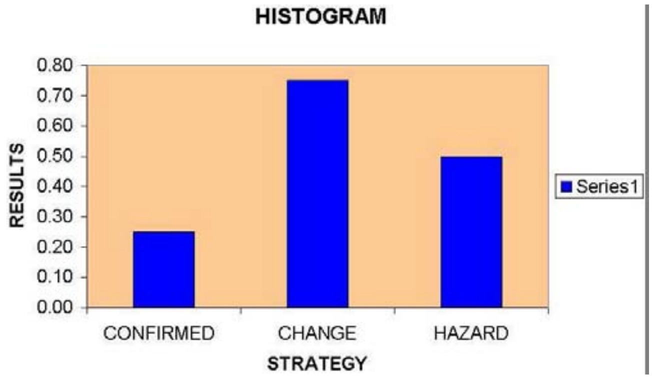

Figure 5. Bar graph (strategy V_s Results)

applied to the whole population as shown in this research "who will win the million".

Each case is done 15 times with different persons. By the end, it is found "statistically" that "change" the answer was the best and it has a probability $\simeq \frac{3}{4}$, and then the "hazard" case has a probability $\simeq \frac{1}{2}$, while the "confirm " case has a probability $\simeq \frac{1}{4}$. Well, statistically this is true and if repeated more than 15 times one will approximately reach the same results. The question that confuses us is "why is this happening?" Mathematics, as all know, is logic, reason, pattern and if probability is math, then one has to find a theoretical reason to clarify this fact. Unfortunately, there is not any theoretical proof. Therefore, one doubts that probability could be a mathematical concept! However, one can prove in other situations that probability is mathematics, especially with the model of favourable outcomes to possible outcomes. This illustrates many definitions that "probability is the mystery of math". One can feel it but not touch it.

References:

[1] Bernard Dantal, Autour de la modlisation en probabilités. publimath.irem.univmrs.fr/bibliocomp/IWV01001.htm-14k

[2] Claude Lévi-Strauss / Comment on deviant Ethnologue. Tristes Tropiques, Plon, 1955 , p.49

[3] Bernoulli , Jacques Ars conjectandi ... Accredit tractatus de seriebus infinitis , et epistola Gallice scripta de ludo pilae reticularis . Basle , Thurneisser Brothers, 1713 .

[4] Annales de l'institut Henri Poincaré (B) Probabilité et Statistques

[5] Derbyshire, J. Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics.New York: Penguin, 2004.36,266

[6] J. Peter, M. Marcelet, S. Burguburu, V. Pediroda, Comparison of Surrogate Models for the Actual Global Optimization of a 2D Turbomachinery Flow 7th WSEAS international conference on simulation, modeling and optimization ,Beijing ,China , september 15-17,2007 .

[7] Stated by James Cleik, to theorie du chaos, Edition Albin Michel, Paris, 1989