UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2018:37

A Historical Survey of the Development of Classical Probability Theory

Özlem Kart

Examensarbete i matematik, 15 hp
Handledare: Ingemar Kaj
Examinator: Veronica Crispin Quinonez
Augusti 2018

Department of Mathematics
Uppsala University

Table of Contents

Abstract

This project involves a brief survey on classical probability theory. The early classical problems and topics on the games of chance provide a useful feedback; and let us to know how probability theory was developed. The three main approaches (the frequentist, the measure and the subjectivist theory) on probability theory, which we mention in the end of the work, cover the period especially after 1900. All these advances and discussions throughout the history of probability theory, paved the way for the development of modern probability theory as an independent field in mathematics.

Keywords**:** *Classical probability, the problem of points, dice problems, mathematical expectation, frequentist theory, measure theoretic approach, subjectivism.*

1. Introduction

This work presents a short review on the topics of classical probability by following them in their historical periods. Throughout the work we survey two questions; *how were early problems on probability discussed? And what were the contributions of these early discussions to the development of the main philosophical approaches?*

Gambling has been a popular activity for thousands of years ago. The first probability calculations were made in the 16th century by the Italian mathematician and physicist Gerolamo Cardano (1501-1576). He wrote the book *Liber de Ludo Aleae* (The Book of Games of Chance) (1565) about probabilities in games of chance like dice, which was published more than a hundred year after it was written. He introduced the concept of the set of outcomes of an experiment, and for cases in which all outcomes are equally probable. He defined the probability of any one event occurring as the ratio of the number of favourable outcomes to the total number of possible outcomes.

Nevertheless, according to many sources, probability theory had started in 17th century. The year 1654 is considered as the birth of the study of probability, since in that year French mathematicians Blaise Pascal (1623–1662) and Pierre de Fermat (1601– 1665) carried on a correspondence discussing mathematical problems dealing with problems posed by gamblers. Their correspondence concerned *the Problem of Points*. Throughout the correspondence they continued to exchange their solutions and methods on mathematical principles and problems through a series of letters. In 1657, Christiaan Huygens (1629–1695) learned of the Fermat–Pascal correspondence and shortly thereafter published the book *De Ratiociniis de Ludo Aleae* (On Reasoning in Games of Chance). He introduced the concept of expected value and dealt with the problems, in particular the problem of points that had been solved earlier by Fermat and Pascal. Huygens' work was discussed by the Swiss mathematician Jakob Bernoulli (1654–1705) in *Ars Conjectandi* (The Art of Conjecturing) which was published by his nephew in 1713. The first part of *Ars Conjectandi* includes the period between 1684 and 1689 related to the work of mathematicians such as Christiaan Huygens, Cardano, Pascal and Fermat. Apart from these subjects, we find the first version of the law of large numbers, which is known as the weak law of large numbers (WLLN).

During the 18th century, probability theory had developed and expanded in many various ways with the advances in combinatorial calculus and in statistics. Thomas Bayes (1702-1761) formulated his inverse theory, which is known as Bayes' Theorem that is an example of what we call statistical inference today. Inverse probability says that we can use conditional probability to make predictions in reverse. The theorem was developed by Pierre Simon Laplace (1749–1827). He worked out principles, developed them into axioms, and finally provide a theorem that is known as Laplace's rule. According to this rule ”... *the probability of a given event is to be inferred from the number of cases in which the same event has been observed to happen: if m is the number of observed positive cases, and n that of negative cases, the probability that the next case observed is positive equals (m + 1)/(m + n + 2). If no negative cases have been observed, the formula reduces to (m + 1)/ (m + 2). Laplace's rule applies to two alternatives, such as the occurrence and non-occurrence of some event, and assumes uniform distribution of priors and independence of trials.*" (Galavotti 2017, p.3-4).

In Bayesianism there are differences between prior and posterior probability. The prior probability represents the observer's degree of belief in the actual happening of the event before the evidence comes into play. On the other side, the posterior probability describes the observer's degree of belief in the actual happening of the event after the evidence has come into play. The two types of Bayesianism are commonly called for objective and subjective Bayesianism which led to the emergence of two common notions of probability; frequentist probability and subjective probability.The objectivists consider that the prior distribution of the parameters must be objective while the subjectivists mostly deal with decision in practical situations, and stand for the prior information on the distribution of the parameters in possession of the decision-maker.

The Laplace notion of probability has been identified as "subjectivism" because of the fact that probability is considered as relative to our beliefs and the expression of human impossibility of getting complete knowledge. The main founder of the subjective theory is known as Bruno de Finetti (1906–1985). He states shortly that probability does not exist. Finetti states that is there is no such thing as an objective measure of probability because all probabilities are subjective.

The second concept of probability is perhaps the most popular one and is based on relative frequency, which is known as an objective probability in so much as it refers to the physical world.The frequency theory is that a probability of an event, which is the relative frequency of occurrence of the event in infinite repeated trials of an experiment.

The frequentist notion of probability appeared around 1900-1930. The studies on laws of large numbers and other limiting theorems about the relative frequency led to the emergence of two different approaches within the axiomatic systems. One of them was developed by von Mises in 1929. Mises' approach does not contain Laplace's classical definition. His theory of probability is based explicitly on the frequentist idea. The other was the modern axiomatic theory of probability which was developed by Andrei Kolmogorov (1933), who was mostly influenced by von Mises's the frequentist theory. Kolmogorov defined his approach based on set theory and measure theory. His ideas was accepted by the majority of probability theorists from the later 1930s to our days.

By the late 1930s probability theory had become an independent part of mathematics. The developments of these early decades have been covered by the nearly universal acceptance of modern axiomatic and measure theoretic probability represented by the classic work of Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933).

We present Kolmogorov's list, which it states the status of probability. It can be an important source for researchers in this field. He divides the modern studies within probability in their research directions together with the principle names:

- ***Anology to the measure theory of real variables, with***
    - *general axiomatization of probability: Borel, Frechet, Kolmogorov, Hopf; and*
    - *Strong laws of large numbers: Borel, Cantelli, Slutsky, Frechet, Khintchine, Kolmogorov, Glivenko, Levy.*
- ***New schemes in physical and other applications, with***
    - *the theory of stochastic processes: de Finetti, Hostinsky, Hadamard, von Mises, Kolmogorov, Frechet, Khintchine, Levy; and*
    - *The theory of random functions: Wiener, Slutsky, Levy.*
- ***New analytical apparatus, with***
    - *equations for stochastic processes: Kolmogorov, Hostinsky, Frechet, Bernstein, Pontryagin;*
    - *characteristic functions and moments in infinite dimensional and functional spaces; Khintchine; and*
    - *New methods for proving limit theorems; Kolmogorov, Petrovski, Bernstein, Kintchine, Bavli."*(Plato 1994, p.233).

1.1. Sources and References

Some of sources we have used in this work are; some early topics from classical probability, like division problem, dice problems, Petersburg problem and Huygens's fifth problem from the book *Problems and Snapshots from the World of Probability* (1994) by G. Blom, L. Holst, and D. Sandell (1994), and from the book *Classic Problems of Probability* (2012) by Prakash Gorroochurn, who won the 2012 PROSE Award for Mathematics from The American Publishers Awards for Professional and Scholarly Excellence.

Another source Jan Von Plato's *Creating Modern Probability* (1994) focuses mainly on the changes from classical to modern probability in mathematics. It provides a wider view of points about three fundamental approaches.

The Unfinished Game: Pascal, Fermat and the Seventeenth-Century Letter that Made the World Modern (2008) by Keith Devlin gives a detailed examples from the letters over the correspondence between Fermat and Pascal. Another sources are In Ian Hacking's book *The Emergence of Probability* (1975), *Studies in the History of Probability and Statistics I. Dicing and Gaming* (2007) by Florence Nightingale David.

2. Classical probability

2.1. The Definition of Probability

The first probability calculations appeared in the 16th century in Italy by Cardano (1501-1576). He considered the word *probability* as "*chances*." For the first time he defined the probability of an event as a fraction: "*the number of ways the event can occur divided by the total number of possible outcomes.*" He considered *circuit* as *the latter*. Today this word is known as *the sample space*. It means that the set of all possible outcomes when an experiment is performed.

Pascal and Fermat didn't actually use the word probability throughout their correspondence, although they were interested in calculating probabilities. Instead they mentioned the division ratios and used such terms as "*value of the stake*" or "*value of a throw*" to express a player's probability of winning. In his book Gorroochurn gives some quotes about the definition of probability from Cardano, Leibniz, Bernoulli, de Moivre and Laplace. They are presented as follows:

The definition given by Cardano (In Chapter 14 of the Liber):

"*So there is one general rule, namely, that we should consider the whole circuit, and the number of those casts which represents in how many ways the favorable result can occur, and compare that number to the rest of the circuit, and according to that proportion should the mutual wagers be laid so that one may contend on equal terms."*

The definition given by Leibniz (1646–1716) in 1710 (Leibniz, 1969, p. 161):

"*If a situation can lead to different advantageous results ruling out each other, the estimation of the expectation will be the sum of the possible advantages for these to fall these results, divided into the total number of results*."

Jacob Bernoulli's (1654–1705) statement from *the Ars Conjectandi* (Bernoulli, 1713, p. 211

"... if the integral and absolute certainty, which we design ateby letter a or b yunity1, will be thought to consist, for example, of five probabilities, as though of five parts, three of which favor

the existence or realization of some events, with the other ones, however, being against it, we will say that this event has 3/5a, or 3/5, of certainty."

De Moivre's (1667–1754) definition from t*he De Mensura Sortis* (de Moivre, 1711; Hald, 1984):

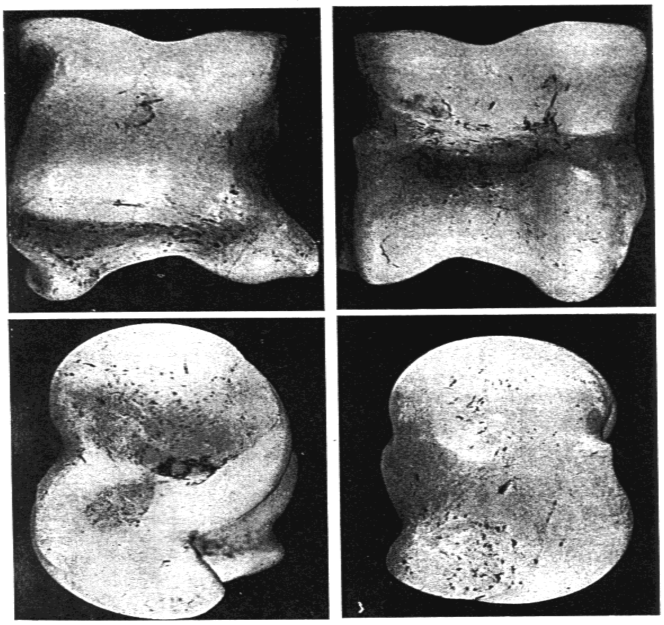
"*If p is the number of chances by which a certain event may happen, and q is the number of chances by which it may fail, the happenings as much as the failings have their degree of probability; but if all the chances by which the event may happen or fail were equally easy, the probability of happening will be to the probability of failing as p to q*."

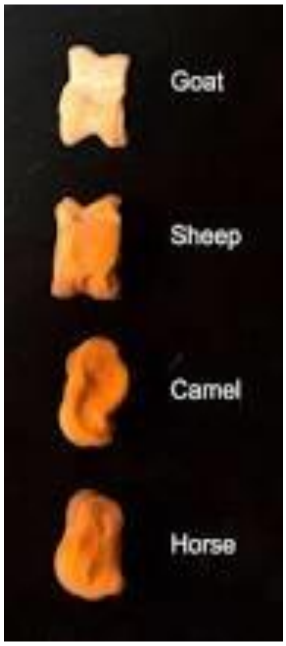
The definition given in 1774 by Laplace (1749–1827), with whom the formal definition of classical probability is usually associated. In his first probability paper, Laplace states:

"*The probability of an event is the ratio of the number of cases favorable to it, to the number of possible cases, when there is nothing to make us believe that one case should occur rather than any other, so that these cases are, for us, equally possible*." (Gorroochurn 2012, p.5).

2.2. A Brief History of Gambling

Humans have been playing games for all of recorded history. Archeological evidence shows that people used special materials which were made of the bones from various animals "*as deer, horse, oxen, sheep and hartebeeste." These* dice-like objects were called the *astragalus* or *talus*. *"This bone is so formed that when it is thrown to land on a level surface it can come to rest in only four ways." (*Hacking 1975, p. 1). The astragalus was not properly shaped in the regular ways. That means that the *astragalus* is not symmetrical as modern dice as we see below:





https://telescoper.wordpress.com/2009/09/05/game-theory/ https://www.pinterest.se

In the classical Greek, Rome and Egypt, it is possible to find a few examples of astragalus. Besides the astragali, *throwing sticks* were also used for games of chance in those times. The sticks were generally made of wood or ivory. Such throwing sticks had been also used by the ancient societies like Greeks, Romans, Egyptians and Maya Indians. In the European throwing sticks were marked by small engraved circles. The actual numbers, which can be obtain, are mostly 1, 2, 5, 6, but 3 and 4 (David, 2007).

Some early examples about *randomness* have appeared also in Indian. Hacking (1975) notices a story that is called for *'Nala story*', which is known as the first piece of Sanskrit writing in modern Europé and in this story he founds a passage which V.P. Godambe emphasized in the Indian epic Mahabarata and H.H. Milman translated. Hacking asserts that no one paid any attention to this connection between dicing and sampling untill he noticed that. The passage is as follows:

I of dice possess the science

And in numbers thus am skilled

This story is about two men who were called Nala and Kali. ”*Kali, a demigod of dicing, has his eye on a delicious princess and is dismayed when Nala wins her hand. In revenge Kali takes possession of the body and soul of Nala, who in a sudden frenzy of gambling loses his kingdom. He wanders demented for many years. But then, following the advice of a snake-king whom he meets in the forest, he takes a job as charioteer to the foreign potentate Rtuparna. On a journey the latter flaunts his mathematical skill by estimating the number of leaves and of fruit on two great branches of a spreading tree. Apparently he does this on the basis of a single twig that he examines. There are, he avers, 2095 fruit. Nala counts all night and is duly amazed by the accuracy of this guess…*” This story shows that dicing was considered as estimitation of “*the number of leaves on a tree*” (Hacking 1975, p. 1).

We do not know how long ago people began to play with throwing astragalus. It was probably thousands of years ago. ”*It is possible that the first primitive dice were made by rubbing the round sides of the astragalus until they were approximately flat*.” (David 1962, p.10).

The usage of the numbers in games of chance appeared in around 960. It is known that first time Bishop Wibold of Cambrai introduced the 56 outcomes in the case of three dice. These numbers thrown simultaneously could be in the sequence as: 1, 1, 1; 1, 1, 2; 2, 3, 5; and so on. Another person from 13th century is Latin poem, De Vetula. He gives several outcomes by listing them as; 216 (= 6 x 6 x 6).

During the 14th century, istead of throwing dice, playing with cards became more popular. In Europe in Italy, we find early examples of various cards games. Because of the banning by the Church, this kind of games did not have any chance to grow up. The emergence of the games of chance at the gaming tables enabled direct observations for the frequencies of various hands.

In 1494, for the first time, Luca Pacioli intruduced a problem about games of chance. It was printed in his book *Summa de arithmetica, geometrica, proportioni et proportionalita* (Everything About Arithmetic, Geometry, and Proportions). This problem is known as *The Problem of Points* and two centuries later after Pacioli, it would be solved by Pascal and Fermat.

3. Problems from Classical Probability

3.1. The Problem of Points

Problem: ”*Two persons A and B participate in a game. In each round, A wins with probability p and B with probability q = 1- p. The game ends when one of the players has won r rounds; he is the winner. The winner receives an amount 1, which we call the stake. For some reason, the play stops in advance. At this moment, A has won (r – a) rounds and B has won (r – b) rounds. How should the stake be divided between the two players?”*

Solution: *If the (a+b-1) additional rounds, A wins at least a rounds and hence B less than b rounds. Clearly, A then wins the game. In the second case B wins at least b rounds and hence A less than a rounds. Player B then wins the game*.” (Blom, Holst, Sandell 1994, p.36-37).

There are two solitions, using modern notation. The game must terminate at round no. *(2r – 1)*, for then certainly one player has won *r* rounds. It therefore suffices to consider what would happen if *(2r – 1) – (r – a) – (r – b) = a + b – 1* further rounds are played. There are two cases which can be distinguished.

1. Of the *(a + b – 1)* additional rounds. *A* wins at least *a* rounds and hence *B* less than *b* rounds. Clearly, *A* then wins the game.
2. *B* wins at least *b* rounds and hence *A* less than *a* rounds. Player *B* then wins the game.

Therefore, it seems correct to give *A* and *B* the parts *p(a,b)* an *q(a,b)* of the stake, respectively, where *p(a,b)* is the probability that *A* wins at least *a* of the *(a + b – 1)* rounds and *q(a,b)* is the probability that *B* wins at least *b* of the games. Clearly, these two probabilities have the sum *1*.

First solution: The probability that *A* wins *i* of *n* rounds is equal to

$$\binom{n}{i} p^i q^{n-i}$$

Hence we obtain $p(a,b) = \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} p^i \cdot q^{a+b-1-i}$

Second solution: It is seen that *A* wins *a* of the *(a + b – 1)* games if he wins *(a – 1) of the first (a – 1 + i)* rounds and also the (*a+i)* th round. Here i is an integer which can assume the values $0,1, \dots, (b-1)$. Hence by the negative binomial distribution.

$$p(a,b) = \sum_{i=0}^{b-1} \binom{a-1+i}{i} p^a q^i$$

3.2. Fermat-Pascal Correspondence

In 1494, Luca Pacioli (1447–1517) introduced a problem known as *The Problem of Points* which was discussed by Niccolo Tartaglia (1499-1557) and Cardano (1501-1576) but they all made wrong estimations and they could not give correct solutions to the problem. Pascal was the first to answer this question correctly, and Pierre Fermat was the second. Christian Huygens learned of this correspondence, rediscovered their reasoning, and shortly wrote it up in *De Ratiociniis in Ludo Aleae* (1657), which was a treatise on problems associated with gambling.

Cardano in *the Practica arithmetice* (1539) discussed the problem. Cardano had stated that the division of stakes was based on "*how many rounds each player had yet to win, not on how many rounds they had already won*", but in the case of the division ratio he could not come to the correct solution. The similar observation was made by Fermat. His solution depended not on "*how many rounds each player had already won but on how many each player must still win to win the prize*" (Gorroochurn 2012, p.24). To get correct division ratio Fermat simply used enumeration as a method.

The problem was to decide whether or not to bet even money on the occurrence of at least one "double six" during the 24 throws. A seemingly well-established gambling rule led de Méré to believe that betting on a double six in 24 throws would be profitable, but his own calculations showed the opposite.

Fermat succeed by using the existing probability principles of the time and knew how to apply them to the problem. On the other hand, Pascal's method was more innovative and general, and his ability lies in that he could create new ways to handle old problems. Moreover, Pascal was the first to have solved the initial significant problem in the calculus of probabilities.

This correspondence between them consisted of five letters and started in 1654 and Pascal closed this process before his second conversion on November 23rd, 1654 (Devlin, 2008). Fermat's initial reply to Pascal showed that the two men had different approaches to solving the problem. Actually both methods are correct. However, Pascal gives an algebraic solution to the problem while Fermat gives a combinatorial one.

The problem took more than half a century before it was taken over by Jacob Bernoulli (1654-1705). It is obvious that these letters paved the way to that many of the great mathematicians further developed the theory of probability by applying it to a growing range of applications.

3.3. Dice Problems

3.3.1. Cardano's Games of Chance (1564)

Problem: *How many throws of a fair die do we need in order to have an even chance of at least one six?* (Gorroochurn 2012, p.1).

Solution: The question is asking about how likely it is that we will get a six in any of the rolls we make. The chance of not getting a 6 is 5/6. The chance of not getting any sixes in two rolls is

$$\left(\frac{5}{6}\right) \cdot \left(\frac{5}{6}\right) = \left(\frac{5}{6}\right)^2$$

Similarly, the chance of not getting any sixes in *n* rolls is

$$\left(\frac{5}{6}\right)^n$$

So the chance of actually getting at least one six in *n* rolls is$1 - \left(\frac{5}{6}\right)^n$; so we want to find the *n* for which this number exceeds $\frac{1}{2}$.

A little trial-and-error shows that if *n*=4, then the probability of getting at least one six is

$$1 - \left(\frac{5}{6}\right)^4 \approx 0.52$$

Cardano's works on probability were published in 1663, in the 15-page *Liber de ludo aleae*. Related to the problem, he could not get the correct solution due to the fact that he considered that the number of throws should be three.

Cardano's conflict resulted from a conflict between the concepts of probability and expectation. He could not rightly recognize the variations of outcomes from their expectations.

3.3.2. Galileo on Dice (1620)

Problem. *Suppose three dice are thrown and the three numbers obtained added. Thetotalscores of 9, 10, 11, and12 can all be obtained in six different combinations. Why then is a total score of 10 or 11 more likely than a total score of 9 or 12?*

Solution. *Table 2.1 shows each of the six possible combinations (unordered arrangements) for the scores of 9–12. Also shown is the number of ways (permutations or ordered arrangements) in which each combination can occur.*

	Score							
	12		11		10		9	
	6-5-1	6	6-4-1	6	6-3-1	6	6-2-1	6
	6-4-2	6	6-3-2	6	6-2-2	3	5-3-1	6
	6-3-3	3	5-5-1	3	5-4-1	6	5-2-2	3
	5-5-2	3	5-4-2	6	5-3-2	6	4-4-1	3
	5-4-3	6	5-3-3	3	4-4-2	3	4-3-2	6
	4-4-4	1	4-4-3	3	4-3-3	3	3-3-3	1
Total no. of ways		25		27		27		25

This problem was discussed by the mathematician Galileo Galilei (1564–1642). The throwing of three dice was part of the game of passadieci, which involved adding up the three numbers and getting at least 11 points to win. Galileo gave the solution in his probability paper *Sopra le scoperte dei dadi* (1620) follows as:

”*But because the numbers in the combinations in three-dice throws are only 16,that is, 3,4, 5, etc. up to 18, among which one must divide the said 216 throws, it is necessary that to some of these numbers many throws must belong; and if we can find how many belong to each, we shall have prepared the way to find out what we want to know, and it will be enough to make such an investigation from 3 to 10, because what pertains to one of these numbers, will also pertain to that which is the one immediately greater.”* (Gorroochurn 2012, p.11).

Galileo then continues to use a method, which is similar to the one in the solution provided previously. He counts the number of favorable cases from the total number of equally possible outcomes in the causal way. That indicates that the use of the classical definition of probability was common at that time.

3.3.3. De Méré's Problem

Problem. *When a die is thrown four times, the probability of obtaining at least one six is a little more than 1/2. However, when two dice are thrown 24 times, the probability of getting at least one double-six is a little less than 1/2. Why are the two probabilities not the same, given the fact that Pr {double-six for a pair of dice} =1/36=1/6 Pr {a six for a single die}, and you compensate for the factor of 1/6 by throwing* $6\cdot 4=24$ *times when using two dice?*

Solution. *Both probabilities can be calculated by using the multiplication rule of probability. In the first case, the probability of no six in one throw is 1 – 1/6=5/6. Therefore, assuming independence between the throws,*

$$\begin{aligned}\Pr\{\text{at least one six in 4 throws}\} &= 1-\Pr\{\text{no six in all 4 throws}\}\\ &=1-(5/6)^4\\ &=0.518.\end{aligned}$$

In the second case, the probability of no double-six in one throw of two dice is

$1-(1/6)^2 = 35/36$. Therefore, again assuming independence,

$$\begin{aligned}\Pr\{\text{at least one double-six in 24 throws}\} &= \Pr\{\text{no double-six in all 24 throws}\}\\ &=1-(35/36)^{24}\\ &=0.491.\end{aligned}$$

Chevalier de Méré (1607 – 1684) posed a question quite similar to Pacioli's original problem. Previous "*solutions*" to these "*points*" problems seemed incorrect, perhaps because the players had to receive proportional amounts reflecting their respective chances of winning the series before the normal or expected time. Thus, the problem greatly confused de Méré and he shared the problem with Pascal and other mathematicians. Pascal proposed possible solutions, while also discussing the problem in multiple correspondences with Fermat. In the end, both Pascal and Fermat solved the problem, by using different methods of reasoning (Gorroochurn, 2012).

3.4. The St. Petersburg Paradox

Problem. *A player plays a coin-tossing game with a fair coin in a casino. The casino agrees to pay the player 1 dollar if heads appears on the initial throw, 2 dollars if head first appears on the second throw, and in general 2^{n-1} dollars if he ads first appears on the nth throw. How much should the player theoretically give the casino as an initial down-payment if the game is to be fair (i.e., the expected profit of the casino or player is zero)?*

Solution. *The player wins on the nth throw if all previous (n -1) throws are tails and the nth thrown is a head. This occurs with probability $\left(\frac{1}{2}\right)^{n-1}\left(\frac{1}{2}\right) = \frac{1}{2^n}$ and the player is then paid 2^{n-1} dollars by the casino. The casino is therefore expected to pay the player the amount*

$$\sum_{n=1}^{\infty} \frac{1}{2^n} 2^{n-1} = \sum_{n=1}^{\infty} \frac{1}{2} = \infty$$

In *Problems and Snapshots from the World of Probability*, the solution to the problem can be seen in details as follows:

"A fair coin is tossed until heads appears for the first time. If this happens on the kth toss, the player receives 2k dollars, where k=1, 2. Determine the expectation of the amount that the player receives."

Let Y be this amount. The number X of tosses has the probability function $P(x = k) = \left(\frac{1}{2}\right)^k$ where $k = 1, 2, \ldots$ We find

$$E(Y) = E(2^X) = 2\left(\frac{1}{2}\right) + 2^2\left(\frac{1}{2}\right)^2 + \ldots$$

The sum diverges and hence *E(Y)* is infinite. The game seems to be exceptionally favourable to the player, since his rich friend apparently loses 'an infinite amount of money'. This paradox deserves a closer examination.

First, suppose that the game is played only once. It is seen that $x \leq 6$ with the probability

$$\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^6}=\frac{63}{64}.$$

This tells us that there is a large probability that the rich friend escapes with paying at most $2^6 = 64$ dollars. Similarly, the change is 1,023/1,024, that is, almost 1, that the friend has to pay at most $2^{10} = 1{,}024$ dollars. To let the person play once thus seems rather safe for the rich friend.

It is worse for the friend if the player makes use of the offer many times. Then it is no longer possible to disregard the fact that the mean is infinite, for the means shows, as we know, what will be paid on the average. The wealthy friend may perhaps demand a stake *s* which makes the game fair. Then *E(Y)-s* should be zero, and, paradoxically enough, the stake must be infinite.

It may be prescribed that the amount 2^k dollars is paid to the player only if the number *k* of tosses is at most equal to a given number *r*. then expectation of the payment becomes

$$2\left(\frac{1}{2}\right)+2^2\left(\frac{1}{2}\right)^2 \ldots.+2^r\left(\frac{1}{2}\right)^r = r$$

And it is possible to make the game fair (Blom et al., 1994).

In 1738, the problem was presented from by Daniel Bernoulli (1700–1782) and he published his solution in the *Commentaries of the Imperial Academy of Science of Saint Petersburg*. However, the problem was firstly discussed by Daniel's cousin Nicolas Bernoulli. By most of mathematicians of the time tried to solve the problem.

3.4.1. Huygens's expectation

In the case of a gamble we need to know that ”a gamble is worth the mathematical expectation of that gamble.” Huygens tried to justify this mathematical expectation. He used the word ”expectatio” in his first tentative translation of his Dutch manuscript. According to Hacking (1975) Huygens didn't want to justify expectation, but he tried ”to justify a method for pricing gambles which happens to be the same as what we call mathematical expectation.”

The notion ”expectation” of a gamble appears in the 16th century. Cardano's notions of *'equality'* and *'the circuit'* in games of dice can be considered as expressions for mathematical expectation. The use of similar word can be found in Huygens' work *calculating in Games of Chance*. The famous example for this case is *the St. Petersburg problem*, which is a first case on record for the question as a practice of pricing by expectation.

This concept was not clear in Pascal and Fermat's problem of the unfinished game. For the first time, Huygen made it explicit by recognizing its significance. Expected gain is generally regarded as the correct objective measure of the value of a particular gamble to the person who makes it. To compute it, you multiply the probability of each outcome by the amount that will be won and add all the results together.

3.4.2. Huygens's fifth problem

Problem. *Two players A and B have 12 counters each. They play in each round with three dice until either 11 points or 14 points are obtained. In the first case, A wins the round and gives a counter to B; in the latter case, B wins and gives a counter to A. The game goes on until one player obtains all the counters; he is the winner. Find the probability that A wins the game.*

Solution: *For the solution we use Struyck's method. After the first round, A has either a + b or a – b counters. Hence we obtain the difference equation*

$$p(a,b) = p^p(a+1,b-1) + q^p(a-b,b+1)$$

There are two boundary conditions, $P(0,a+b) = 1$ *and* $P(a+b,0) = 0$

The solution is

$$p(a,b) = \frac{(q/p)^{a+b} - (q/p)^a}{(q/p)^{a+b} - 1}$$

Interchanging p and q, and a and b, we further obtain

$$Q(a,b) = \frac{(p/q)^{a+b} - (\frac{p}{q})^b}{(p/q)^{a+b} - 1}$$

It is found that these expressions add up to 1, which implies that an unending game is impossible. When a=b, we find after a reduction the following quotient between A's and B's probabilities of losing the game:

$$\frac{P(a,a)}{Q(a,a)} = (q/p)^0$$

We now return to Huygens's fifth problem. We then have a=12, and a simple calculation shows that p=9/14 and q=5/14. Hence

$$\frac{p(12,12)}{Q(12,12)} = \left(\frac{5}{9}\right)^{12} \approx 0.0009.$$

A's chance of winning all counters divided by B's chance is (9/5)12 ≈1,157 23 (Blom, Holst, Sandell 1994, p.38).

4. The Main Approaches of Probability Theory

After 1900, probability theory developed and grew up also within science and physics. In those years, many important science notions were developed by James Clerk Maxwell, Ludwig Boltzmann, and Josiah W. Gibbs. Ludwig Boltzmann had developed the kinetic theory of gases. Marian von Smoluchowski and Albert Einstein worked out Brownian motion. The analysis of radiation led Einstein and others such as Max Planck, Erwin Schrödinger, Louis de Broglie, Paul Dirac, Werner Heisenberg, Max Born, and Niels Bohr to formulate quantum mechanics. It made probability theory a basic element of the description of matter.

In the 1929, Richard von Mises formulated a theory of probability based explicitly on the frequentist idea. It was followed by the more abstract approach based on set theory and measure theory developed by Kolmogorov, who mostly influenced by von Mises's the frequentist theory. His ideas accepted by the majority of probability theorists from the later 1930s to our days.

A completely different approach to probability theory is *subjectivism*, it is known as *Bayesians* especially in statistical field. Probability is understood as a measure of 'degree of belief' and probability theory as a systematic guide to 'behaviour under uncertainty.' The subjectivist theory became important since the 1950s, which is created by Bruno de Finetti in the late 1920s and early 1930s.

By the late 1930s probability theory had become an independent part of mathematics. The developments of these early decades have been covered by the nearly universal acceptance of modern axiomatic and measure theoretic probability represented by the classic work of Andrei Kolmogorov of 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung.*

4.2. Measure Theoretic Approach

This theory was almost as old as measure theory. Measure theory took its name because of the measure theoretic probabilities studied by Kolmogorov. He provides the two mathematical essentials of *Grundbegriffe der Wahrscheinlichkeitsrechnung* follows as:

- The theory of conditional probabilities,
- The general theory of random or stochastic processes.

In elementary probability a conditional probability is defined as:

$If\ P(B) > 0$, the conditional probability $p(A \setminus B)$ of event A given event B is the number $P(A \cap B)\ /\ P(B)$. A conditional probability measure is a probability measure over a new space. We can call it as *a conditional probability space* (Plato, 1994).

Kolmogorov published several informal analysis of his frequentist ideas of probability in 1938–1959. In 1963 he formalized his philosophy on tables of random numbers. In his article he states his reasons follows as:

- *The infinitary frequency approach based on limiting frequency (as the number of trial goes to infinity) cannot tell us anything about real applications, where we always deal with finitely many trials.*

- *The frequentist approach in the case of a large but finite number of trials cannot be developed purely mathematically* (Galavotti 2017, p.3-4).

Kolmogorov was allways sure about point 1, but with the complexity of algorithms he changed his mind about point 2. He accepted the fact that there are few simple algorithms to define a finitary version of von Mises's collectives.

4.3. The Frequentist Theory

According to the frequentist theory, probability is defined as the limit of the relative frequency of a given attribute, observed in the initial part of an indefinitely long sequence of repeatable events, such as the observations obtained by experimentation. The basic assumption underlying this definition is that the experiments generating frequencies can be reproduced in identical conditions, and generate independent results.

The frequency approach started in the 19th century thanks to two Cambridge mathematicians, namely Robert Leslie Ellis and John Venn, and reached its climax with Richard von Mises.

The existence of a collective is taken to be a necessary condition for probability, in the sense that without a collective there cannot be meaningful probability assignments. As von Mises pointed out, in order to qualify as a collective a sequence of observations must have two features. According to Von Mises, the resulting sequence of numbers is called a collective if the following two postulates are satisfied:

- *Limits of relative frequencies in a collective exist.*
- *These limits remain the same in subsequence formed from the original sequence.*

In addition, a collective must be random. Randomness is defined in an operational fashion as insensitivity to place selection. Insensitivity to place selection obtains when the limiting values of the relative frequencies in a collective are not affected by any of all the selections that can be made on it. The limiting values of the relative frequencies observed in the sub-sequences obtained by place selection equal those of the original sequence.

This randomness condition is also called by von Mises as the 'principle of the impossibility of a gambling system' because it excludes contriving a system leading to a sure win in any hypothetical game of chance. The failure of all attempts to devise a gambling system is meant to secure an empirical foundation to the notion of a collective.

According to von Mises probability is a concept that applies only to collectives. He re-states the theory of probability in terms of collectives, by means of four operations which are based on the following ideas:

- **Subsequence selection.**
- **Mixing.** A collective, that is, the ordering of labels to a sequence of experiments, is produced by a noninvertible function of the label space. A simple example is offered by the sum of the faces from a toss of a pair of dice.
- ***Division*** (Teilung or Aussonderung). A collective K' is formed from a given collective K by taking only those elements ei∈K whose labels belong to a subset M' of the original label space M. M', like any proper label space, has to have at least two members, both with an infinity of labels ordered to the elements. Division does the job conditional probability in von Mises' theory.
- ***Combination*** (Verbindung) of two collectives. The elements of the new collective are pairs of elements of the given collectives. The distribution of the new collective is a product formed by the distribution of the given collectives. This multiplication rule of probabilities is proved from the requirement that combination produces a collective, in particular, fulfils condition II. Combination generalizes inductively to any finite number of given collectives (Plato, 1994).

The obvious objection to the operational character of this theory is that infinite sequences are never to be obtained. Von Mises answers that probability as an idealized limit can be compared to other limiting notions used in science, such as velocity or density.

Since probability can only refer to collectives, under von Mises' approach it makes no sense to talk of the probability of single occurrences of events, like the death of a particular person, or the behavior of a single gas molecule. To stress this feature of the frequency theory, von Mises says that talking of the probability of single events "has no meaning".

This originates the so-called single case problem affecting frequentism. Also debatable are the basic assumptions supporting the frequency theory that is the independence of outcomes of observations, and the absolute similarity of the experimental conditions. Nonetheless, after von Mises'work, frequentism became as popular with physicists and natural scientists as to become the official interpretation of probability in science, and was also accepted by orthodox statisticians.

4.4. The Subjectivist Theory

Subjectivism is a theory of knowledge, which is generated from the mind, without reference to reality, which is reflections from outside world and affects our minds. The base of the theorem stems from Bayes' formula, which is combined with simple criteria of rationality, allows a direct link from experience and prior information to decisions. Bayes' Theorem is a simple mathematical formula used for calculating conditional probabilities. It figures prominently in subjectivist or Bayesian approaches to epistemology, statistics, and inductive logic. Subjectivists, who maintain that rational belief is governed by the laws of probability, lean heavily on conditional probabilities in their theories of evidence and their models of empirical learning. Bayes' Theorem is central to these operations because it simplifies the calculation of conditional probabilities and clarifies significant features of subjectivist position.

Italian mathematician Bruno de Finetti made a decisive step towards a mature subjectivism by showing that the adoption of Bayes' rule taken in conjunction with exchangeability leads to a convergence between degrees of belief and observed frequencies. This result, often called 'de Finetti's representation theorem', ensures the applicability of subjective probability to statistical inference. Unlike Bayesian, de Finetti considers the change from prior to posterior probabilities as the cornerstone of statistical inference, and interprets it from a subjective perspektiv, in the sense that moving from prior to posterior probabilities always contains personal judgment.

The subjectivist theory is a completely different approach to probability, which was developed by Bruno de Finetti in the late 1920s and early 1930s. In the twentieth century, objective probability has been connected to the development of physics. An epistemic notion of probability takes its grounds from the Laplacian doctrine of mechanical determinism. In that time probabilities were classified into epistemic and objective. Finetti's probability was related to first one. Firstly he dealt with usual conception of epistemic probability, which was based on a deterministic view of science, was formed by modern physics.

De Finetti's theory of probability can be divided into two parts; quantitative and qualitative one. The first is related to '*the measure of the subjective probability'* and transforming to '*degree of uncertainty into a determination of a number'*. The second part of Finetti's theory can be replaced by *'axiomatic theory of qualitative probability*." (Plato 1994, p.166).

In his theorem of exchangeability, de Finetti states that exchangeable observations are conditionally independent relative to some latent variable. The theorem holds for infinite sequences of exchangeable events.

In order to describe the theorem we make the following construction: Consider a sequence of *n* independent trials, where, at a given trial, an event *H* occurs with probability *p*. Let *A*, be the event that *H* occurs at the *i*th trial. Then $A_1, \ldots, A_n$ are independent events. Let *X* be number of events that occur among $A_1, \ldots, A_n$. The *rv X* has a binomial distribution:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}\,, \qquad (1)$$

Where $k = 0,1, \ldots, n$.

We will now give a the description from independent to exchangeable events. Suppose that *p* is no longer constant but an *rv* taking values over the interval *(0, 1)*. Its distribution can be continuous or discrete; for brevity we assume that it is continuous, but the discrete case is entirely analogous. What happens now to the probabilities involved?

Let f*(p)* be the density function of *p*. We find

$$P(A_i) = \int_0^1 pf(p)\,dp, \qquad (2)$$

$$P\left(A_i A_j\right) = \int_0^1 P^2 f(p)\,dp, \qquad (j \neq j) \qquad (3)$$

and so on. Evidently, the A_i\`s constitute a finite sequence of exchangeable events, for any choice of density function.

The density function *f(p)* is often called the *prior distribution* or, more briefly, the *prior*.

We obtain from *(1)* the important expression:

$$P(X = k) = \int_0^1 \binom{n}{k} P^k (1-p)^{n-k} f(p)\,dp, \qquad (4)$$

Where $k = 0,1, \ldots, n$. This is the probability function of *X*.

After these preparations we are ready for de Finetti's theorem. Consider any infinite sequence $A_1, A_2, \ldots$ of exchangeable events. According to the theorem, the finite subsequences $A_{1,}\ A_2, \ldots, A_n$ of such a sequence can always be obtained in the way described by *(4)*: Start with independent events and integrate over the prior. Any infinite sequence has its own prior, which has to be determined in each special case.

On the other hand, if we assign a fixed value to *p*. the A_i's become independent and *(1)* holds. Expressed otherwise, exchangeable *A*'s taken from an infinite sequence are conditionally independent, given *p* (Blom, Holst, Sandell 1994, p.16-17).

5. Conclusions/Discussions

It can shortly be said that there are two essential movements that define the history of probability. Firstly, there is a period of application of probability from Pascal's and Fermat's gambling games to limit theories. Secondly, it is characterized through process of axioms based on different approaches on the probability. In this period two radically different systems were formulated by von Mises and by Kolmogorov. Von Mises tried to define randomness from a frequentist point of view while Kolmogorov focused on creating foundations of measure theory.

In generally Pascal and Fermat dealt with analyses which were based on the symmetric situations in games of chance. On the other side, Bernoulli extended the concept of probability to asymmetric situations applicable to diverse real-life problems.

Philosophical interpretations of probability are related to the mathematical theory. The efforts to apply probability theory practically lies beyond the mathematicians's work. Kolmogorov's probability theory can thus be seen as the alternative explanation for a mathematician to formalize complex objects. The contributions of Kolmogorov to probability theory are very important. His research lead to an impressive area of research which flourishes to this day.

Undeniably, mathematics is the background of most important philosophical debates in the foundation of the probability theory. The philosophical and historical concept of probability play a crucial role in arousing the success of probability theory. In that case we can discuss two questions for the later studies:

- Which philosophical approaches can be argued in the foundation of the probability theory?
- What is the crucial role of Kolmogorov in forming of modern probability theory?

6. Bibliography

[1] Blom, G., Holst, L., Sandell, D. (1994) *Problems and snapshots from the world of probability*. New York: Springer-Verlag.

[2] David, F. N. (2007). *Studies in the History of Probability and Statistics I. Dicing and Gaming.* Biometrika Trust.

[3] Devlin, K. (2010). *The Unfinished Game: Pascal, Fermat, and the Seventeenth- Letter that Made the World Modern. Basic Books*. A Member of the Perseus Books Group.

[4] Galavotti, M. C. (2014): *The Interpretation of Probability. Still an Open Issue?* University of Bologna.

[5] Gorroochurn, P. (2012). *Classic Problems of Probability*, John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

[6] Hacking, Ian (1975). *The Emergence of Probability*. Cambridge Univ. Press.

[7] Von Plato, Jan. (1994). *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective.* New York: Cambridge University Press.