韓國數學教育學會誌 <數學教育>
1981. 12. Vol. XX No. 1

Journal of the Korea Society of Mathematical Education
Dec. 1981. Vol. XX No. 1

# A Sketch of the History of Probability Theory

By Richard Dahlke, Robert A. Fakler, Ronald P. Morash

*University of Michigan, Michigan, U.S.A.*

## Introduction

Ranking alongside death and taxes as certainties in our world is the existence of the uncertain. Uncertainty has always been with us and will surely continue to be. Recognizing this fact, man has long been engaged, albeit often haphazardly, in trying to predict the uncertain. But in recent centuries, while much of the informal and imprecise guessing has continued to go on among many, a formal science, the theory of probability, has come into existence, and has had a great influence on many aspects of modern life. The main purpose of this paper is to provide a brief sketch of the historical development of this area, and of its metamorphosis from a pseudo-science, applied mainly to games of chance, to a full-fledged branch of mathematics, with wide-spread applications in fields as diverse as physics and political science. In the course of our sketch, we will consider such questions as when did a formal or systematic study of the uncertain (i.e. of probability theory) begin? What provided the impetus for such a study? Who were the persons primarily responsible for its early development and what were their significant contributions toward this end? Were there misconceptions that hindered its development and application?

A secondary objective of the article is to convey some of the spirit of the past. We will see that, in many instances this spirit is not unlike that of the present. To help attain this secondary objective, we will occasionally relate personal anecdotes on the lives of some of the very early principals involved. A motley group they were, lawyers, physicians, astrologers, and philosophers among them.

This sketch of the history of probability theory should be of interest to various and diverse groups: student-teachers enrolled in mathematics education programs, instructors at various levels giving courses in probability and their students, historians (amateur or otherwise) of mathematics, and scientists in various fields who use probability in their work. It is not intended as an introduction to the theory itself, and in fact presupposes some familiarity with the basic concepts of probability, or at least the willingness on the part of the reader to consult other sources for that information.

**Origins of the Study of Probability - Gambling**

It is impossible to state with absolute certainty exactly which force served as the primary impetus for man's earliest dabblings with probability. Indeed, in the inconclusive world of prehistory, it is difficult to amass more than just indications (as opposed to conclusive evidence) of the types of interests, goals, and ideas of mankind that led to the evolvement of probability. Some such indicators suggest the possibility of religious and economic factors, among others. But, in truth, the one factor which seems to emerge more frequently than any other in this development is man's love of gambling. Not only is it logical to believe that gambling games would have helped generate man's interest in probability, given how convenient such games are as a device for describing chance phenomena, but it is actually fairly well-documented. In the next two sections, you will see that the origin of the study of chance was stimulated, if not actually caused, by man's interest in problems associated with games of chance. One might say to users of probability who are opposed to gambling on moral grounds that its key role in the early development of probability theory is certainly a redeeming feature, questionable though the activity itself may be.

**Early Gambling Devices**

Examples of gambling are recorded in the earliest annals of human history. Several millennia ago, small bones of animals called **astragali** (heel bone) were tossed and the side landing upward observed. One of the games played in ancient Greece was the simultaneous tossing of four astragali with the best throw being the one in which the four sides landing face-upward were all different. The board games of today had a forerunner in an ancient Egyptian game (ca. 3500 B.C.) now labeled as "Hounds and Jackals." In this game, astragali were tossed and then the hounds and jackals were moved on a board according to how the astragali had landed.

Another ancient gambling device was a rectangular-shaped stick made of wood or bone whose lateral faces were marked with dots from one to four. The stick was tossed and the number of dots on the upper side was noted.

Eventually, the astragali and sticks gave way to the use of dice (made of clay, ivory, wood, sandstone, etc.) with the most popular being six-sided cubical dice, where the faces were marked with from one to six dots. In some instances, the sum of the number of dots on opposite faces was not seven, unlike the standard dice of today. The earliest known dice are dated about 3000 B.C.

It is believed that playing cards were in existence in ancient times, with modern cards first appearing in France in the 14th century. Cards slowly replaced dice as the most popular form of gambling device. Besides providing recreation, cards and dice were used for divination purposes, that is, for determining the intentions of a deity.

**Prehistory of Probability Theory: Cardano and Galileo** (This period terminated early in the 17th century.)

It might be appropriate at this point to give two examples of types of problems in probability whose correct solution or interpretation characterizes the progress which marked the very beginnings of probability as an organized discipline. This earliest progress is the topic of this section of our paper; the time period in question ends by the early 17th century, and the key figures are the Italians Cardano and Galileo. At the close of this section, we will point out a third major new direction whose beginnings can be associated with the two names mentioned above.

The first type of problem whose solution was necessary for the theory's early development had to do with **counting**, and more specifically, distinguishing among different outcomes of an experiment (using modern terminology). For example, a problem that arose in connection with gambling was the calculation of the possible number of outcomes in the tossing of three dice. There are recorded examples, from the 10th and 11th centuries, of incorrect calculations being made of this number. Specifically, it was often not recognized that the event corresponding to the description "one 'six' and two 'fours'" actually corresponds to three different possible outcomes of a toss, rather than just one. Poems appeared in which each verse corresponded to one of the 56 outcomes (instead of the correct 216) that arise when the calculation is done this way. On the other hand, the calculation which, it was eventually agreed, is a correct one, also was carried out as early as the 11th century. The earliest known poem which included the correct method of calculating the ways in which three dice can occur (counting repetitions) is the Latin poem **De Vetula**, written in the 11th century. The following is a free translation of an excerpt from this poem:

*If all three numbers are alike there are six possibilities; if two are alike and the other different there are* 30 *cases, because the pair can be chosen in six ways and the other in five; and if all three are different there are* 20 *ways, because* 30 *times* 4 *is* 120 *but each possibility arises in six ways. There are* 56 *possibilities.*

*But if all three are alike there is only one way for each number; if two are alike and one different there are three ways; and if all are different there are six ways.* (7, *pp.* 5-6)

Putting these two paragraphs together, we see that the total number of ways is

$$6 \cdot 1+30 \cdot 3+20 \cdot 6=216.$$

For a long period of time, with no agreed-upon rules of procedure for solving such problems, many errors were made in attempted solutions to counting problems, such as the dice problem just described.

Another type of problem, perhaps a more subtle one, which only began to be recognized as a problem and handled properly in the late 16th century, is illustrated by the **fair division of stakes** problem. It is primarily a problem of **formulation and interpreting**

**what the meaning of a problem should be.** The specific problem we pose here was studied by Luca dal Borgo (among others), also known as Paccioli (1445-1509), a professor of mathematics at Milan. In his 1487 treatise entitled "Summa de Arithmetria, Geometria, Proportioni et Proportionalita," he surveyed the mathematical learning in existence at the time. The following example of a "fair division of stakes" problem appears in this treatise:

*A and B are playing a fair game of balla. They agree to continue until one has won six rounds. The game actually stops when A has won five and B three. How should the stakes be divided?*

Paccioli argues that the stakes should be divided in a manner proportional to the rounds won; that is, in the ratio of 5 to 3. By modern interpretations his solution is in error since one should actually take into consideration the number of rounds yet to be won by each participant and from this the likelihood that each would win the match, if it were played out to its agreed-upon length. In this case, by the way, the stakes would actually be divided in a ratio of 7 to 1, in favor of A. Obviously, in any progress that might be made in relating a developing mathematical theory with "real-world" problems, the overcoming of misconceptions of the type just described would be crucial.

A notable and important first step in a direction toward some general principles on which a science of probability would be based was the **first book on probability**. In 1526, the Italian Gerolamo Cardano (1501-1576) wrote a small volume (15 folio pages, each containing two columns) entitled **Liber de ludo aloae** (Book on Games of Chance), which was not actually published until it appeared with his collected works in 1663. Cardano was a professor of medicine, a mathematics teacher, and a habitual gambler who was often in dire financial straits. He seems to have been one of the true eccentrics of all time. It is said that he predicted the day when he would die and starved himself for three weeks before to make it come true. At times he was shown to be dishonest. As an example, some historical accounts state that he plagiarized work of the mathematician Tartaglia concerning the solving of cubic equations.

Although his book on probability has been debunked by many as a "badly printed gambler's manual", it did contribute to the beginning of the formulation of basic notions and concepts. For example, Cardano wrote of the notion "circuit," which was a forerunner of the modern concept **sample space**. With reference to the problem of tossing two dice, this term represented the 36 ways in which two tossed dice can land. Using this concept, he comes close to a proper method for calculating the **probability of an event.** For example, he talks about the number of ways in which the sum of the faces of two dice can be 10 as follows:

*The point* 10 *consists of* (5, 5) *and of* (6, 4), *but the latter can occur in two ways, so that the whole number of ways of obtaining* 10 *will be* 1/12 *of the circuit*...(13, *p.* 192).

In essence, he is determining the "probability", but only as a fractional part of the circuit, not as a ratio of equally likely events. More on this distinction shortly.

It might also be mentioned that, in the above example, in which Cardano clearly distinguishes between the two ways of arriving at one six and one four in the toss of two dice, he is correctly dealing with a counting problem of the type described earlier. He also has a correct notion of what is a **fair game**, and presents examples to this effect. He then states a general rule for calculating the amount of stakes to be bet on a game, and in this rule he is essentially defining probability as a **ratio of equally-likely events** via the concept of **odds.** These two ideas are aptly illustrated by one of his examples:

*If, therefore, someone should say, I want an ace, a deuce, or a trey, you know that there are 27 favorable throws, and since the circuit is 36, the rest of the throws in which these points will not turn up will be 9; the odds will therefore be 3 to 1. Therefore, in four throws, if fortune be equal, an ace, deuce, or trey will turn up three times and only one throw will be without any of them; if, therefore, the player who wants an ace, deuce, or trey were to wager three ducats and the other player one, then the former would win three times and would gain three ducats; and the other once and would win three ducats; therefore in the circuit of four throws they would always be equal. So this is the rationale of contending on equal terms; if, therefore, one of them were to wager more, he would strive under an unfair condition and with loss; but if less, then with gain.* (13, *p.* 200)

Other contributions of Cardano in the area of the origins of statistical theory will be mentioned shortly.

Another Italian, Galileo Galilei (1564-1642) followed Cardano in contributing to the advent of probability. A professor of mathematics, he was nonetheless better known for inventing the telescope and for his work in the physics of falling bodies. He wrote a minor book entitled **Sopra le Scoperta a dei Dadi** (Thoughts about Dice Games) that was first published in Florence in 1718. One example of his work, as exhibited in this book, is his solution to the problem of determining the number of possible outcomes when three dice are tossed. This was the most complete solution given up to that time, and furthermore generalized easily to a larger number of dice. For the two dice, he obtained the solution 36 by determining the product 6·6 and for three dice the solution 216 by calculating 6·6·6. In the process, he discovered why the sum ten (of three dice) is more likely than the sum nine, thus answering a longstanding question.

A factor which contributed to certain developments in probability theory in this period was the just-mentioned invention of the telescope. This new tool contributed to an increase in the need for the development of a third aspect of early probability and statistical theory which had its origin in the period under discussion, the **estimating of the frequency and distribution of observational errors.** This is related to the question of the relationship between the mathematical probability of an event and the occurrence of that event, in practice, when the experiment is carried out. Cardano dealt with ideas related to this in his work. In modern terms he had touched upon **statistical regularity** and the

**law of large numbers.** For example, his writings show him to be aware that, for a particular event, the conjectured portion of a circuit of which it consists, can deviate substantially from the result obtained from a small number of observations, whereas it will deviate only slightly from the result obtained by a large number of observations. This is evidenced when he discusses his conclusion that, in the toss of two dice, there are 30 outcomes with unlike faces and 6 with like faces:

> *So the whole set of circuits is not inaccurate, except insofar as there can be repetitions ... in one of them. Accordingly, this knowledge is based on conjecture which yields only an approximation...; yet it happens in the case of many circuits that matter falls out very close to conjecture.* (13, *p.* 196)

Galileo, in addition to contributing in the area of statistics by his invention, pioneered theoretical work in this area as well. Maistrov writes:

> *Thus Galileo arrived at the conclusion that errors in measurement are inevitable, that the errors are symmetrically distributed, that the probability of error increases with the decrease of the error size, and that the majority of observations cluster around the true value. Moreover, errors obtained in observation are not to be compared with the final errors, which arise as a result of computations based on the observed values.* (9, *p.* 33)

In these conclusions Galileo revealed features of the important **normal probability distribution** law which was yet to come.

Before moving to the next period, we make some final comments about Cardano and Galileo, the two key figures of this period. Even though Cardano's book was the first in which probability theory is discussed, some historians do not classify him as the founder of probability theory (Since his problems were not subject to mathematical analysis, no rigorous rules for their solution were available at that period and possibly also because his book did not appear until long after his death). Nonetheless, when one considers the many aspects of the probability theory to come which his examples touched upon, one must concede that he was well ahead of his time and deserved more recognition. Everything considered, Galileo's work in probability was not as far reaching as Cadano's, but his style of writing and ability to think scientifically were far superior. Although he did not achieve many far-reaching solutions to problems in statistics, he contributed by formulating many problems and suggesting avenues for attacking them, which in turn influenced later workers.

**Origins of Probability Theory as a Science** (A period extending from the middle of the 17th century to the beginning of the 18th).

The birth of the theory of probability, occuring around 1650, is generally attributed to two Frenchmen, Blaise Pascal and Pierre de Fermat. As we shall see, however, a third figure, the Dutchman Christiaan Huygens, made contributions so significant as to justify his being placed very nearly in a category with these two as an originator of the theory.

Blaise Pascal (1623-1662) was a child prodigy who was burdened by poor health. At the age of 18, he invented the first mechanical calculator. A mathematical triangle of numbers which he used and popularized (but was not the first to construct), and a law of physics on fluid pressure both bear his name. At the age of about 30, he essentially gave up his scientific work and concentrated on religions and philosophic writing until his death at age 39. He is of course one of the most famous French literary figures, in addition to his important role in our history. Pascal's interest in problems related to probability was apparently initiated by his association with Chevalier de Méré (1607–1684), a philosopher and member of the French aristocracy, whose writings make him also a well-known figure in 17th century French literature. Evidently de Méré was a very charming man who placed a high premium on sociability and good conversation. He also placed possibly a higher value on his own role in the achievements of Pascal (as we shall see shortly, de Méré essentially posed some key questions to Pascal) then we are able to do here. In a letter to Pascal, he writes:

> *You know that I have discovered such rare things in mathematics that the most learned among the ancients have never discussed them and they have surprised the best mathematicians in Europe. You have written on my inventions, as well as Monsieur Huygens, Monsieur de Fermat, and among others who have admired them.*

De Méré's initial contact with Pascal came as a result of a pleasure trip he made with a duke who happened to bring Pascal along, and they soon became good friends.

We now turn to two probability problems which de Méré posed to Pascal, thus beginning a chain of events which led to the developments of several basic concepts and theorems of probability. Being a gambler of sorts, de Méré was familiar with the first of these problems, an old dice problem: What is the smallest number of throws of two dice that must be made so that there is a better than even chance of obtaining two sixes at least once? De Méré, basing his reasoning on an erroneous ancient gambling rule, believed that there should be 24 tosses. He turned to Pascal with this problem who rightly showed that the number of tosses is 25. (The usual solution to this problem shows that, in $n$ throws, the probability of getting two sixes at least once is $\left(1-\frac{35}{36}\right)^n$. For $n=24$, the probability is 0.4914, and for $n=25$ it is 0.5055.)

The second problem which de Méré posed to Pascal was a problem (referred to earlier) that was several centuries old, and one which Paccioli and Cardano had worked on (but not solved), namely, the **division of stakes problem**: How should the stakes in a game be divided if the game is interrupted before it is completed and the points each participant needs to win the game are known? Pascal solved this problem for a particular case and also for the general case by making use of the arithmetic triangle of binomial coefficients that bears his name. His solution recognized the fact that the stakes had to be divided proportionally to the probabilities of each participant winning, provided the game is continued. This was a problem of such difficulty for that period that Pascal's solution can be considered a major breakthrough in the history of probability theory, and a major point of departure between this period and the earlier period covered in the previous section.

Pascal's solution to the division problem received some criticism, thus he corresponded with Pierre de Fermat (1601-1665), perhaps the most influential mathematician in France at that time, to get his opinion. Fermat was a lawyer who studied literature and mathematics His accomplishments were many. He not only has the distinction of being considered a co-founder of probability theory, but he also did important work in the early development of calculus, and in fact influenced Newton in this area. In addition he is credited with being the founder of modern number theory. Pascal and Fermat frequently exchanged letters during 1654 on the division problem and on other mathematical problems including the dice problem. Fermat solved the division problem and Pascal was delighted to find that the latter's results were in complete agreement with his own. Pascal writes "I see that the truth is the same in Toulouse and Paris." Fermat's solution is not as elegant as Pascal's but their methods are basically the same.

The generalities of Pascal's and Fermat's work were well known in Paris. To the extent that, even then, they were being credited with discovering a new branch of mathematics.

Being aware of the work of Fermat and Pascal, other mathematicians entered the area of probability and made contributions to it. Most notable for his time was Christiaan Huygens (1629-1695), a Dutchman who was born into a family of wealth and position. He received a doctorate in law, and besides working in mathematics, was also a physicist. Numbered among his accomplishments in physics were the discovery of the rings of Saturn, the use of a pendulum in clocks, and a principle in optics which carries his name. His treatise on the theory of probability titled "De Rationciniis in Ludo Aleae" ("Reasoning on Games of Chance") was published in 1657. It was the **first book actually published** on the subject of games of chance. (Recall that Cardano's book did not appear in print until 1663, long after his death).

Huygens's treatise on probability theory consisted of a short preface, 14 propositions (some of which are specific problems with detailed solutions-e. g., propositions 4 through 8 give solutions to specific "division of stakes" problems, and five specific problems left for the reader. His book systematized the propositions created by Pascal and Fermat, and was the source for an introduction to probability theory to students for half a century, going through several editions. The foreword to this book contained a letter written by him to another mathematician, and included this sentence:

> *I would like to believe that in considering these matters closely, the reader will observe that we are dealing not only with games but rather with the foundations of a new theory, both deep and interesting.* (6, *p.* 58).

This new attitude toward probability, together with the caliber of mathematicians now working on it, were characterizing features of the period we are discussing.

A major contribution of Huygens to probability theory was the introduction of the notion of **mathematical expectation** (expected value) and applications of it. During Huygen's time, Holland was a leader in commerce and so in the development of accounting procedures. In accounting the arithmetic mean is frequently used, and it appears that the notion of expected value surfaced as a generalization of the arithmetic mean. Huygens used commercial jargon in referring to the notion of mathematical expectation, calling it the "value of the chance." The first three propositions in his treatise were related to this concept, with the third defining it:

*To have p chances of obtaining a and q of b, chances being equal, is worth* $(pa+qb)/(p+q)$. (6, *p.*66)

Huygens was the first to use probability in studying vital statistics of humans. He used John Graunt's (London) new famous book displaying vital statistics to construct a mortality curve and to define the notions of mean and probable duration of life. Shortly thereafter, probability theory was being applied to annuities.

Huygens, unlike other early probabilists, did not use combinatorial methods when solving probability problems. Thus, his solutions are not considered elegant by some. The development of **combinatorics** had some influence on the early development of probability theory, since it was the major tool of probability theory. It became of lesser importance when calculus came into use in probability theory. Leibniz (1646-1716) extensively develops combinatorial methods somewhat later in his dissertation "Ars Combinatoria," which was published in 1666.

Including the work of Pascal, Fermat, and Huygens, we see that a substantial amount of probability knowledge had accumlated by the middle of the 17th century. The notion of probability itself was more meaningful, problems were being categorized according to the probabilistic notions needed to solve them, the addition and multiplication laws of probability were known, mathematical expectation was introduced, and probability was being applied to astronomy, physics and statistics. Nonetheless, as a mathematical theory it was still an infant. Further significant advancements were not forthcoming until the 18th century. Perhaps mathematicians were distracted from working in probability theory in the latter stages of the 17th century because of the invention of the absorbing subject of calculus in the middle of that century.

**The continuing development of probability as a mathematical discipline,** (early 18th to mid 19th centuries).

We now embark upon a period during which probability theory became solidly established as a mathematical discipline. The period was marked by the continuing increase in involvement in the field by mathematicians who were of great stature, independent of their work in probability. Indeed, Our exposition on this era will feature such names as Euler, Gauss, Laplace, Bernoulli and de Moivre. It was also during this period that advances were made both in the area of a continuing refinement and reformulation of basic concepts (e.g., a better-defined definition of probability became generally agreed upon), as well as in that of new major advances. In this section, we shall discuss a number of significant theorems that were spawned in this era and which are familiar to any student completing a junior or senior level probability and statistics course.

Proceeding in chronological order, we first consider the contributions of James Bernoulli (1654–1705), a member of the mathematically famous Bernoullifamily. James became a mathematician and mathematics teacher after taking a degree in theology; he also had a considerable interest in astronomy. He was well-known in Europe as an excellent teacher of mathematics and his services were sought by students far and wide.

Bernoulli was influenced greatly in probability by Huygens's book **Reasoning on Games of Chance**. He is best known among probabilists for his work titled **Ars Conjectandi (Art of Conjecturing)**. The book was published in 1713 and demonstrates his powerful mathematical attack on problems. In this book, he presents and proves the most decisive conceptual innovation in the early history of probability theory, namely the **first limit theorem.** His book consisted of 300 pages, divided into four parts. The first part was reprint of Huygen's book with a commentary by Bernoulli; the second part was a complete and valuable exposition of that part of the theory of combinatorics then in existence along with many new combinatorial properties that he derived; the third part had detailed solutions to 24 problems on games of chance; and the fourth part (left incomplete by the author) was on the applications of probability to civil, moral, and economic problems.

It is in the fourth part of Bernoulli's book that another new era in the history of probability begins. There appears here a limit theorem that now bears Bernoulli's name and is also commonly referred to as the **weak law of large numbers.** This theorem articulates the relationship between the mathematical probability associated with trials of an experiment and the empirical (experimental) probabilities $s/t$ ($s$ successes in $t$ trials) for larger and larger values of $t$. In simple terms, the theorem states that as the number of trials of an experiment becomes larger and larger (i, e., at $t$ increases), the probability approaches 1 that the difference of the empirical probabilities $s/t$ from the constant probability associated with each experiment will be as close to 0 as we desire. More formally, if $p$ is the constant probability associated with each trial, then for any positive value $\varepsilon$, we can assert with probability as close to 1 as we please that $|s/t-p|$ is less than $\varepsilon$ for sufficiently large numbers of trials.

With Bernoulli's theorem, the number of practical applications of probability theory becomes virtually unlimited. For this reason and the fact that his theorem opened the door for further limit theorems, he made an immense contribution to probability theory.

A French nobleman, Pierre-Remond de Montmort (1678–1719) worked on probability in the early part of the 18th century. Like many other early mathematicians, he was also a student of philosophy and religion. Although very wealthy, he was a hard-working mathematician, and enjoyed a good reputation in the scientific community. For instance, Leibniz obviously valued him highly as he asked him to be his representative on a commission that was to rule on the controversy between Leibniz and Newton as to who first discovered the calculus.

Montmort is best known for two editions of his book entitled **Essai d'Analyse surles Jeux de Hazard (Essay on the Analysis of Games of Chance)**, published in 1708 and 1713.

The book had four sections, the first dealing with a theory of combinatorics, the second with games of chance with cards (unlike all the previous works that strictly pertained to games of chance with dice), the third with games of chance with dice, and the last section contained solutions to various problems, including those proposed by Huygens as well as letters between Montmort and N. Bernoulli, a nephew of James.

Montmort, unlike James Bernoulli, did not include applications of probability theory to civil, moral, and economic problems. He writes:

*If I were going to follow M. Bernoulli's project I should have added a fourth part where I applied the methods contained in the first three parts to political, economic, and moral problems. What has prevented me is that I do not know where to find the theories based on factual information which would allow me to pursue my researches.* (12)

His use of algebra in his work on probability theory is noteworthy and comparable to that of James Bernoulli. He, like other probabilists before him and others yet to come, made tireless efforts to develop a field that was still in its infancy.

Working in probability near the time of J. Bernoulli and R. Montmort was the Frenchman Abraham de Moivre (1667-1754). De Moivre made many contributions to mathematics; he is well-known for his work in infinite series and for determining powers and roots of complex numbers. He was imprisoned in France when he was 18 and when freed several years later he went to England, never to return to France. He was not of noble birth and the unkind rumor was that he added the noble prefix "de" upon leaving France. He was poverty-stricken thorughout his life and was dissatisfied with his inability to save money from tutoring students.

*If there were some position where I could live tranquilly and where I could save something, I should accept it with all my heart...*

He hoped that Leibniz could assist him in securing a university post:

*...he (Leibniz) is so universally respected, and one has so much confidence in him, that a word from his mouth in my favor will be extremely advantageous...*

De Moivre's most well-known work on probability theory was **The Doctrine of Chances** which appeared in 1718 and was dedicated to his good friend Newton. It went through three enditions with the third appearing in 1756. This last edition summarized the past 50 years of probability theory and can be viewed as the first modern textbook on probability. De Moivre and Montmort had unkind words for each other concerning their respective works. De Moivre implied that Montmort had only improved on Huygens and Montmort implied that de Moivre had taken his ideas from Montmort's book **Essai d' Analyse**. Both were correct to some extent. De Moivre was greatly influenced by Montmort and J. Bernoulli, and this coupled with his mathematical powers that were superior to either of these two men, allowed him to accomplish much more.

In his book **The Doctrine of Chances** he solves a **duration of play** problem in which he first introduced the idea of a **generating function** for probability. The problem was first proposed by Huygens:

*Given two players A and B, the probabilities of their winning a round are $p$ and $q=1-p$, respectively. At the beginning of the game A has a coins and B b coins. The loser of each round is to give a coin to the winner. What is the probability that A will lose all of his coins before winning all of B's coins?*

Also in this book, he develops a result which, when later extended by Laplace, is the **second basic limit theorem** in probability theory (the first limit theorem being J. Bernoulli's). What de Moivre did was to determine, for $p=1/2$, the number of observations (i.e., a value of t) that are necessary for $s/t-p$ to have a given likelihood of falling within certain limits. This complemented the result of J. Bernoulli, who had shown that it is **possible** to find a value of $t$ which would cause $|s/t-p|$ to fall within certain limits.

Some probabilists credit de Moivre with being the first to derive the **normal law** (which we will discuss presently). He died at the age of 87, and is considered by some to the first of the great analytic probabilists.

The advancement of probability in England in the early 18th century was largely the work of Thomas Simpson (1710-1761) and Thomas Bayes (1702-1761), each of whom made notable contributions to probability theory. Simpson was the first to make use of a **continuous distribution** in his work on error theory. A diagram of this distribution appears in a 1757 collection of his papers titled "Miscellaneous Tracts on Some Curious... Subjects in Mechanics, Physical Astronomy and Speculative Mathematics." Bayes' work in probability theory was contained in an essay found among his papers after he died. The essay was published in 1763 and titled "An Essay towards Solving a Problem in the Doctrine of Chances by the late Rev. Mr. Bayes, F.R.S., communicated by Mr. Price in a Latter to John Canton, A.M., F.R.S." This essay shows that Bayes was also a pioneer in using continuous distributions. He was the first to derive the curve for the **binomial distribution** and obtain all of its properties. He also established the rule for determining the probability that the probability $p$ of an event lies in a interval $(a, b)$, knowing that the event occured $m$ times in $n$ independent trials, and assuming that $p$ has a uniform probability distribution. Bayes is often given credit by modern writers for discovering the familiar probability theorem which bears his name. However, this theorem is nothing more than an alternate form of the multiplication rule for probabilities which was known prior to Bayes' time and appears nowhere in his writings. It was Laplace who assigned Bayes' name to the formula.

It appears that Leonhard Euler's (1707-1783) interest in probability began when he was asked by King Frederick of Prussia to help with setting up a lottery to replenish the government treasury. In subsequent years, Euler was consulted on other lotteries that the king was contemplating. Among other problems associated with setting up a lottery, Euler calculated the probabilities of drawing certain sequences of numbers, say 3, 4, and 5, from $n$ numbers. On the basis of such calculations, he then calculated the price at which a lottery ticket should be sold. Euler solved problems on other games of chance, including card games. He also tackled problems in demography and insurance, with his most notable contributions being made in the foundations of modern demography. His technique for solving problems is worthy of emulation-he solves them for a simple case, proceeds to more complicated cases, and finally solves the general case. His major thrust in the development of probability was in the area of applications. Although this work must be considered minor when compared to Euler's accomplishments in other fields, it illustrates the forces which helped the field of probability in its development. Eminent mathematicians, working in the positions from which they earned a living, found themselves facing practical problems probabilistic in nature. It is not surprising that such mathematicians would not only solve those problems ingeniously, but also would see beyond to a general class of theoretical problems and to development of theoretical tools for their solution.

Daniel Bernoulli's (1700-1782, son of James Bernoulli) work in probability theory included the study of probability curves, construction of the first table of the normal curve, and applying for the first time the differential calculus in solving probability problems. His main work in this area is the memoir "De usu algorithmi infinitesimalis in arte conjectandi specimen." In this memoir, he remarks:

*Each time the situation changes as a result of a continuous change process, as for example, cards with different numbers are drawn from an urn in succession and when the laws determining various changes resulting from this operation are investigated it seems useful to apply infinitesimal calculus, provided each variation can be considered infinitesmally small. This is possible as long as the number of cards remaining in the urn is very large, since in this case unity can be taken as infinitesimally small. The same hypothesis was the basis of the arithmetic of infinitesimal quantities utilized by mathematicians before differential and integral calculi were discovered. However, I realize that this separately posed question requires additional elaboration and I therefore proceed to illustrate this matter by examples in which I first use the visual analysis and then carry on to an application of the algorithm of infinitesimal quantities.* (1, *Paragraph* 1)

The Frenchman, Pierre-Simon Laplace (1749-1827), made monumental contributions to the field of probability, as well as to other fields including celestial machanics. His reputation in these fields was primarily established through his works **Theorie Analytique des Probabilities** (**Analytic Theory of Probability**) which was first published in 1812 and which sustained three editions, and **Traite de Mecanique Celeste** (Treatise on Celestial Mechanics) consisting of five volumes published in the period 1799-1825. Laplace was a staunch supporter of Napoleon Bonaparte, and dedicated his third volume of "Celestial Mechanics" to him. In this volume he writes "Bonaparte, the Pacificator of Europe to whom France owes her prosperity, her greatness, and the most brillant epoch of her glory." However, his loyalty to Napoleon ceased when Napolean's position as dictator began to weaken.

So great were the contributions of Laplace to the field that he is often called the father of modern probability. The most important of Laplace's many contributions to probability theory was the limit theorem now named for him. We referred earlier to the special case ($p=1/2$) of this theorem proved by de Moivre. The theorem deals with the distribution of deviations of the frequency of occurrence of an event in a sequence of independent trials from its probability. More precisely, he showed that this **distribution** (which resulted from a suitable normalization of a binomial distribution) **approaches the normal probability distribution.**

The major thrust of his book **Analytic Theory of Probability** was the proof of this theorem. Maistrov has this to say about the book:

*He systematized the previous, aften uncoordinated results, improved the methods of proof, laid the foundations for study of various statistical regularities, successfully applied probability theory to estimation of errors in observation, and so on.* (9, *p.* 142)

The book includes a new proof of James Bernoulli's theorem and the solutions to a variety of problems, including the division of stakes problem, the Buffon Needle Problem, and the St. Petersburg problem. It is particularly interesting to note also that it contains the **classical definition of probability**: Assuming that the outcomes of an experiment are equi-probable, the probability of an event occuring on a single trial of an experiment is equal to the number of possible outcomes that are favorable to the event divided by the total number of outcomes. Laplace was not the originator of this definition, but his inclusion of it marked an important step in its general adoption. James Bernoulli's **Ars Conjectandi** was the first well-circulated work to make use of the concept of equi-possibility, while Leibniz used this definition of probability in 1678. Laplace's book also includes some of his work on applying the theory of probability to demography. This application of probability was of special interest to Laplace, including its particular application to taking census by sampling. His accomplishments in this area contributed greatly to the development of statistics.

It was due to Laplace's book (i. e., to the limit theorem appearing there) that the application of probability theory could now be viewed as a scientifically justified method. Unfortunately, Laplace did not restrict the universe to which his results were applicable. In fact, he seemed to believe that the theory was applicable to solving the most important problems in any field, including various social problems. Hence he was vulnerable to incorrect applications, such as his claim, based on faulty use of the law of large numbers, that a judgment handed down by a court is more likely to be correct if more judges are involved in the decision. Commenting in this connection, Maistrov writes:

*It is Laplace's view that all the regularities of any field of mass phenomena may possibly be reducible to the unique normal law, as the celestial phenomena are reducd to the unique law of universal gravitation. Based on this point of view he attempts to apply probability theory to court procedures, decisions at gatherings, and so on. Such an unfounded and erroneous extension of the applications of probability theory had a negative effect on the* *development of this science.* (9, *p.* 148)

As we shall see later, misguided efforts in this direction by later practitioners of probability (who had James Bernoulli and Laplace as their forernners) contributed substantially to the movement of the theory of probability in a more formal direction, namely the use of axiomatics.

An American mathematician, Robert Adrain (1775-1843), and the immortal German mathematician Carl Friedrich Gauss (1777-1855) independently and almost simultaneously derived the basic result, the **normal law for the distribution of errors**. Adrain's derivations (he gave two) were published in 1808 in **The Analyst**, or **Mathematical Museum 1** and Gauss' in 1809 in his famous work, "Theoria Motus Corporum Coelestium". It was Gauss' derivation that formed the basis for further development of the theory of errors. In his wide-range of scientific activity, he was not one to divorce applications from theory or vice versa and this was particularly true in his probabilistic works. Maistrov writes:

*Although Gauss's contributions to probability theory are connected with applications, they were not devoted solely to applied problems. His works have had significant impact on the development of various branches of probability theory. For example his theory of errors prompted an investigation of conditions under which the normal distribution law is applicable. Gauss's contributions raised the problem of estimating the parameters of the normal distribution. He also substantiated the least-squares method using probability theory, taking as an axiom the pinciple of the arithmetic mean.* (9, *p.* 151)

The normal law of the distribution of errors that was discovered by Gauss can give rise to errors of any magnitude, as he pointed out. Unfortunately, this law was considered a universal law, hence it was assumed that all the observations should be retained. It was several decades later that probabilistic criteria for rejecting observations began to appear. Much of Gauss' work on the theory of errors has stood the test of time. Statistics textbooks of today show little or no modification of his work in this area.

Simeon Denis Poisson (1781-1840), a very prolific mathematician with a wide range of interests, is perhaps best known in the field of probability theory for his limit theorem which he designated "the law of large numbers", for a probability distribution he derived (which bears his name), and for an approximation theorem he developed. His "law of large numbers" is a generalization of Bernoulli's theorem (referred to as the weak law of large numbers) and is stated thusly: if the probability of an event E occurring in $n$ independent trials is not the same in each of the trials, then as $n$ increases without bound, the probability approaches 1 that the frequency $m/n$ of the occurrence of event $E$ will be arbitratily close to the arithmetic mean $p$ of the probabilities of the occurences of event $E$ in the individual trials. Using modern notation, we have

$$\lim_{n\to\infty} p(|m/n-\tilde{p}|<\varepsilon)=1.$$

Bernoulli's theorem is a special case of Poisson's theorem for the situation where the probability of event $E$ occuring is constant from trial to trial (i.e., $\tilde{p}=p$). Approximation theory gives an estimate (via the Poisson Distribution) to the probability of $m$ occurrences of an event $E$ in $n$ independent trials-the content of Laplace's theorem for **small** $p$, where $p$ is the probability of the occurence of event $E$ in a single trial.

These two theorems of Poisson appeared in his major work on probability "Recherches sur la probabilite des jugements en matiere criminelle et en matiere civile," published in 1837. Included in this book is a discussion by Poisson of the applicability of probability theory to evaluating the correctness of court decisions. He claims that his law of large numbers, and not Bernoulli's weak law of large numbers, is what makes it applicable. It was his belief that probability theory could be applied to all moral and physical events, which stimulated him to investigate limit theorems. His views on this wide applicability of probability theory were supported by some mathematicians and attacked by others who felt that Poisson (and also Laplace) were compromising mathematical science. Gnedenko comments on the status of probability theory at that time:

*In spite of the fact that Laplace and Poisson concluded an important and fruitful initial period in the development of probability theory, a period of philosophical cementation of the basis of this science, this period resulted in an indifferent attitude toward probability theory in the West and in a definite rejection of the possibilities of utilizing its methods in studying natural phenomena. This led to the beginning of a long period of stagnation in probability theory in the West.* (5, *p.* 394)

**Probability Theory in Russia** (in the last half of the 19th century)

Up to the middle of the 19th century, the major achievements in probability theory were the law of large numbers and Laplace's Theorem. The misapplications of these theorems by Laplace and Poisson (and others) illustrated the necessity of determining the scope of their applicability. The basic postulates of probability theory needed to be further substantiated and more accurately specified. For probability theory to overcome the impasse which it had reached in the mid-19th century, fresh ideas were needed as well as an applications oriented approach to the basic problems. A savior came along in the person of the Russian mathematician Pafnutii Lvoich Chebyshev (1821-1894).

Chebyshev was a highly regarded lecturer and instructor at St. Petersburg University. He founded a mathematical school in Russia called the St. Petersburg School, which was very instrumental in the development of mathematics in Russia. He made large contributions in many branches of mathematics with his contributions in probability theory among the most notable. Lyapunov characterizes Chebyshev's approach to solving problems:

*Chebyshev and his disciples always remained on solid ground, guided by the opinion that only those investigations initiated by applications (scientific or practical) are of value and only those theories which arise from the consideration of particular cases are actually useful.* (2, *p.* 20)

Chebyshev's status today arises primarily from his work on limit theorems which included the **construction of bounds on the errors involved when applying the theorems.**

He viewed an approximate solution as accurate if and only if bounds for the errors could be found. Even though he wrote only four papers in probability theory, their impact on this science was immence. In his Master's thesis (1846) "An essay on elementary analysis of probility theory", he gave a proof for Bernoulli's theorem as well as deriving bounds on the errors involved when applying the theorem, and he proved Poisson's theorem for a finite number of different probabilities (shortly thereafter he proved the theorem in general, giving bounds on the errors).

Chebyshev's investigations in probability theory led him to two thecrems which paved the way for the future development of this theory. In a paper "On Mean Values", which he delivered in 1866 to the Academy of Sciences, he proved what is now known as the **Chebyshev inequality**. This inequality was used to help him prove the first of his two major theorems, called Chebyshev's theorem, or **Chebyshev's form of the law of large numbers**. It states that for sufficiently large $n$, the probability will be as close to one as we please that the arithmetic mean of a number of random variables will be as close as we please to the mean of their mathematical expectations. Using modern notation, given

$$\varepsilon>0, \qquad \lim_{n\to\infty} p\left(\left|\frac{x_1+x_2+\cdots+x_n}{n}-\frac{u_1+u_2+\cdots+u_n}{n}\right|<\varepsilon\right)=1$$

Chebyshev's theorem is an extension of Bernoulli's and Poisson's theorems for the case of independent random variables with uniformly bounded variances. Indeed, to show the latter, let an event $A$ occur in $n$ trials with probabilities $p_1, p_2, \cdots, p_n$, respectively. Define the random variable $x_i$ by letting $x_i$ have the value 1 if event $A$ occurs and 0 if event $A$ does not occur in the $i$th trial. Then.

$$u_i=1\cdot p_i+0\cdot q_i; \qquad \bar{x}=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{m}{n}$$

By Chebyshev's theorem,

$$\lim_{n\to\infty} p\left(\left|\frac{x_1+x_2+\cdots+x_n}{n}-\frac{p_1+p_2+\cdots p_n}{n}\right|\leq\varepsilon\right)=1$$

or

$$\lim_{n\to\infty} p\left(\left|\left(\frac{m}{n}\right)-\tilde{p}\right|\leq\varepsilon\right)=1,$$

where $\tilde{p}$ is the arithmetic mean of the probabilities of occurrences of the event in individual trials $x$ (this is the statement of Poisson's theorem). If all the probabilities $p_i=p$, then $\tilde{p}=p$, and we obtain Bernoulli's theorem:

$$\lim_{n\to\infty} p\left(\left|\left(\frac{m}{n}\right)-p\right|\leq\varepsilon\right)=1$$

Chebyshev should not be given complete credit for the inequality which he used to prove this theorem. I. J. Bienayme (1796-1878) was involved in its development as Markov mentions:

*We associate this remarkable and simple inequality with the two names Bienayme and Chebyshev, because Chebyshev was the first to clearly express and prove it, while the basic idea of the proof was pointed out much earlier by Bienayme in a memoir containing the* *inequality itself, in a not particularly obvious form* (10, *p.* 92).

The second major theorem proved by Chebyshev, although there were gaps in his proof, was the **central limit theorem** (with different restrictions than in the modern version of that theorem) which was published in the 1887 **Proceeding of the Academy of Science**, some 20 years after he proved his law of large numbers. The paper was titled "On two theorems concerning probabilities." Chebyshev extensively used the **theory of moments** in this paper and is given credit for constructing the method of moments in probability theory. The central limit theorem establishes that under certain conditions the distribution law of probabilities of the sum of a large number of independent random variables approaches the normal distribution in the limit as the number of summands increases. Maistrov indicates why this theorem is of such importance to science:

*There are in nature a vast number of phenomena subject to the action of a large number of causes where each individual cause acts independently and exerts only a very small influence on the course of the phenomenon.* (9, *p.* 207)

By its very nature, Chebyshev's work found immediate and widespread application. His works helped place Russia at the forefront of probability theory mainly because of his allegiance to and results in obtaining exact estimates on deviations from the limiting laws, and by his establishment of the restrictions on applicability of these laws.

Andre Andreevich Markov (1856-1922) was a most important figure in probability theory, a protege of Chebyshev, and an able spokesman for Chebyshev's ideas. Markov took over the teaching of Chebychev's probability course at St. Petersburg University upon Chebychev's retirement in 1883. He was an excellent lecturer and used his book "The Calculus of Probabilities" (published in 1913) in his probability course. The book was clearly written, thorough, and rigorous. It contained many new contributions to probability theory and was to be used by the beginner as well as by the scholar. The introduction to the fourth edition contained this statement:

*The most valuable feature of this book is that it does not present the standard, dry, overworked smooth scientific material, but is teeming with original research contributions. These features render the book as a classical work with no equal in the theory of probability.* (11, *p. XIV*)

One of Markov's significant contributions to probability theory was his simpler proof, using the method of moments, of Chebyshev's central limit theorem in 1898. Markov's proof was rigorous, filling in some gaps in Chebyshev's original proof. It might be noted that A. N. Lyapunov (1857-1918) in 1900 and 1901 wrote two papers in which he proved this theorem using considerably weaker restrictions, and by utilizing the more general methods of characteristic functions. It was 20 years later that the restrictions were further weakened with a new sufficient restriction by Y. W. Lindeberg in 1922 and W. Feller's proof of the necessity of this restriction in 1935.

Markov also invented a important new branch of probability theory, namely, the study of **dependent random variables**. In early 1900 he showed the applicability of the law of large numbers and of the central limit theorem to sums of dependent random variables, in particular those connected in a chain. Chains of dependent random variables are nowadays called **Markov chains** and their study, as well as the study of **Markov processes** (which are so applicable to the study of the natural sciences and engineering) is now a large part of modern day probability theory.

There was a complete shift in the area of applications of probability from the first half to the second half of the 19th century. In the first part of this century, probability theory was being applied primarily to the processing of observations and to demographic statistics. The emphasis in the last half of the 19th century was in the area of physics, mainly due to the works of the Austrian physicist Ludwig Boltzmann (1844-1906) and the American mathematician and physicist Josiah Willard Gibbs (1839-1903). Gibbs' contributions were in the area of **statistical mechanics.** Maistrov summarizes Boltzmann's contributions:

> *Boltzmann was one of the most prominent theoretical physicists of the second half of the nineteenth century and one of the founders of modern physics. His name is primarily connected with the initiation and development of statistical physics. His main contribution was the molecular-kinetic interpretation of the second law of thermodynamics and the derivation of the statistical interpretation of entropy.* (9, *p.* 225)
>
> *Boltzmann's ideas greatly influenced the development of physics. In particular, statistical notions developed by him paved the way for quantum theory since statistical notions are among the initial concepts of this theory.* (9, *p.* 229)

**The Axiomatization of Probability Theory** (beginning in 1917 and having as its focal point the appearance of the Kolmogorov axioms in 1933. This approach continues to underlie the development of probability theory in the present day).

At the outset of the 20th century all probability textbooks used the classical foundations of probability theory which stemmed from Laplace's classical definition of probability. This definition was circular, defining "equi-probable" as "equi-possible", but this didn't cause any significant problems. More importantly, the definition was not applicable to a majority of significant probabilistic problems in the sciences that did not give equi-probable cases. The sciences were expanding rapidly and the use of probabilistic notions in these sciences required further clarification and justification of them. The foundations of probability theory were ambiguous, still causing it to be applied to areas that had no relevance to the subject. Emil Borel (1871-1956) was an excellent mathematician, yet he, along with Laplace, Poisson and others before him, incorrectly applied probability theory to social and moral problems.

A careful examination of the foundations of probability theory would require the identification of basic concepts and the establishment of rules of inference for working with these concepts. That is, any new logical foundation of probability theory had to be based on the **axiomatic method.** Then, any set of objects satisfying the axioms would be suitable for study by the methods of the theory of probability. N. I. Lobachevskii paved the way for the axiomatic method by constructing a geometry based on axioms different from Euclid's. David Hilbert and Guiseppe Peano axiomatized geometry and arithmetic at the beginning of the 20th century. The time was definitely right for the axiomatization of probability theory.

The first works on the axiomatization of probability theory are due to S. N. Bernstein (1880-1968). He also contributed to the further development of limit theorems for dependent random variables, to the derivations of new results in Markov chains, and to the applications of probability theory in the natural sciences, including biology. His first work on axiomatization was published in 1917 in the **Proceedings of the Kharkov Mathematical Association** (Vol. 15) in a paper titled "An essay on the axiomatic foundations of probability theory." A detailed axiomatization of probability theory, along with applications of probability theory in the natural sciences, appeared in Bernstein's excellent textbook "Probability Theory." This textbook was published in 1927 and lasted through four editions, with the final edition appearing in 1946. The book contains his three axioms of probability theory: the axiom of comparability of probabilities, the axiom of incompatible (disjoint) events, and the axiom of combination of events. From these three axioms he was able to construct the general structure of probability theory.

Bernsteins views on mathematical probability and its applications were expressed at the First All-Russian Conference of Mathematicians in 1927 in Moscow.

*Purely mathematical probability theory cannot be concerned whether the coefficient called mathematical probability has any practical value, either subjective or objective. The only condition to be fulfilled is the absence of contradictions, namely: various methods of calculating this coefficient under given conditions and provided the axioms are not violated, should lead to the same value of this coefficient.*

*Moreover, if we want the conclusions of probability theory not to degenerate into a simple mental exercise, but allow for empirical verification, it is necessary to consider only those propositions or assertations which can actually be established as false or true. The cognitive process, which is irreversible in its nature, actually means that certain propositions became veritable, i.e. are realized and simultaneously in this case their negation becomes false or impossible.* (15)

A positive step in the further advancement of probability theory came with A. N. Kolmogorov's axiomatization and its recognition. His work on the logical foundations of probability theory from the middle 20's on resulted in the 1933 publication of his book "Grundbegriffe der Wahrscheinlich-keitsrechnung" which contained his axioms. These axioms reflected the rapid trends that were developing in the sciences, and made use of basic concepts of set theory and other relatively new areas of mathematics. Gnedenko comments on the relationship between the basic notions of probability theory and the basic concepts of other mathematical sciences:

*These analogies between such seemingly different branches of science made it possible to* *present the logical foundations of probability theory in a different light and enrich its contents with new problems and new methods of investigations as uell as to complete the solution of certain classical problems.* (4, *p.* 363)

As a result of Kolmogorov's axiomatization, probability theory became a full-fledged modern mathematical discipline that was applicable to a variety of disciplines ranging from economics to cybernetics. The change from a very restrictive and vague notion of probability to a very broad and precise notion took over 400 years. This change is perhaps best illustrated by Cardano's solution to a specific problem in 1526 and Kolmogorov's axioms that appeared in 1933. If you recall, Cardano used the term "circuit" to devote the 36 ways in which two dice can land. He talks about the number of ways in which the sum of two dice can be 10:

*The point* 10 *consists of* (5, 5) *and of* (6, 4) *but the latter can occur in two ways, so the whole number of ways of obtaining* 10 *will be* 1/12 *of the circuit*... (13, *p.* 192)

Kolmogorov's axioms follow (8):

Let $E$ be a collection of elements $\xi$, $\eta$, $\zeta$, ..., which we shall call **elementary events,** and $F$ a set of subsets of $E$; the elements of the set $F$ will be called **random events.**

Ⅰ. $F$ is a field of sets.

Ⅱ. $F$ contains the set $E$.

Ⅲ. To each set $A$ in $F$ is assigned a non-negative real number $P(A)$. This number $P(A)$ is called the probability of the event $A$.

Ⅳ. $P(E)$ equals 1.

Ⅴ. If $A$ and $B$ have no element in common, then $P(A+B)=P(A)+P(B)$.

In case of an infinite field $F$ an additional axiom is supplemented, which for the case of finite fields follows as a corollary from the above stated five axioms.

Ⅵ. For a decreasing sequence of events,

$$A_1 \supset A_2 \supset \cdots\cdots \supset A_n \supset \cdots$$

of $F$, the product (intersection) of which is empty, the following equation holds:

$$\lim_{n\to\infty} p(A_n)=0$$

## References


1. Bernoulli, D., De usu algorithmi infinitesimalis in arte conjectandi specimen, Novi Commentarii Academiae Scientiae Imp. Petropolitanae XII (1768).
2. Chebyshev, P. L., "*Selected Mathematical Works.*" Moscow-Leningrad, 1946.
3. David, F. N., "*Games, Gods and Gambling,*" Griffin, London, 1962.
4. Gnedenko, B. V., "*A Course in Probability Theory.*" lst ed. Moscow-Leningrad, 1950.
5. Gnedenko, B. V. (Proc. Inst. History Natural Sciences) 2 (1948).
6. Huygens, C. I. "*Oeuvres Completes,*" Vol. 14. 1920. La Haye, Martinus Nijhoff (for the Societe Hollandais des Sciences).
7. Kendall, M. G., Studies in the history of probability and statistics Ⅱ, The beginnings

of a probability calculus, *Biometrika* 43, 1-14 (1956).

8. Kolmogorov A. N., "*Foundations of the Theory of Probability.*" (Second English Edition), Chelsea Pub. Co., New York, 1956, p. 2.
9. Maistrov, L. E., "*Probability Theory-A Historical Sketch.*" Acad. Press, New York and London, 1974.
10. Markov, A. A., "*The Calculus of Probabilities*", 4th ed. Moscow, 1924.
11. Markov, A. A., "*Selected Works.*" Acad. Sci. USSR, Moscow, 1951.
12. Montmort. P., "*Essai d' Analyse sur les Jeux de Hasard,*" 2nd. ed. Quilau. Paris, 1713/14.
13. Ore, O., "*Cardano: The Gambling Scholar.*" Princeton Univ. Press. Princeton, New Jersey, 1953.
14. One, O., "Pascal and the invention of probability theory, "*Amer. Math. Monthly* 67, 409-419 (1960).
15. Proceedings of the First All-Russian Conference of Mathematicians in Moscow, 1927.
16. Todhunter, I., "*History of the Mathematical Theory of Probability*". Cambridge, University Press, London and New York, 1865. (Reprinted by Chelsea, New York, 1949, 1961.)