Introduction

The Origin of Probability

Probability owes its existence to gambling. It was game-theoretical problems, which was the main motivation for the pioneers of probability, and the notions of odds and numerical probabilities arose from the study of chance games. But let me begin the story at its beginning. The beginning seems to be a small bone, called *astragalus*, lying above the heel bone, called *talus*.[1] The astragalus from sheep has form and size, which makes it suitable for dice games. The astragali is reported to be much more numerous (10–50 times) than other kinds of bones in prehistory archeological excavations and, moreover, many of the prehistoric astragali show clear signs of being used for some purpose. Since there is clear evidence that the astragalus was commonly used as a die from 3500 B.C. until 1600 A.D.,[2] it is a fair guess that primitive man used the astragalus for some kind of dice game. The earliest evidence of gambling with astragali is an Egyptian tomb painting from 3500 B.C. showing a nobleman and his wife using an astragalus in a board game, and in the tomb of Reny-Soube at Thebe in Upper Egypt archeologists have found a complete board game[3] containing three astragali. From the antique Greek and Roman authors, we know that the astragali was diligently used in many different kinds of games.

The astragalus is asymmetric and contrary to a die it only has four sides on which it can rest. According to Greek and Roman authors the four possible sides of the astragalus was numbered: 1, 3, 4 and 6. Two of the sides (1 and 6) are narrower than the two others and have probability $\approx$ 10% of showing, and the two remaining sides (3 and 4) have probability $\approx$ 40% of showing. Four astragali was often used in a simple game called *rolling of the bones*. The worst throw (called *the dog*) in this game was four 1's and the best (called *Venus*) was 1, 3, 4, 6. In Table 1 you will find the probabilities of the different throws in the game "rolling the bones. Note that "the dog" has the smallest probability, and that the probability of getting "the Venus" is 384 times more probable than getting "the dog". Thus, the assignment of the values differs highly from the probabilities.[4] The Roman author Propertius (50 B.C. – 15 A.D.) writes: *I was seeking the Venus with favorable tali,*

---

[1] The astragalus and the talus are two different bones, but in the literature they have been used indiscriminately to mean the astragalus bone.

[2] The astragalus is still in use on the Faroe Islands for gambling.

[3] Named "hounds and jackals" by the archeologists.

[4] In the distant past as now people hated to lose and that may be reason that the highly improbable "dog" was claimed to be the worst possible throw.

but the dammed dog always leapt out. Because “the dog” comes approximately 1 out of 10,000 times, the good Propertius must have been very unlucky that day.

Table 1

Probabilities in the game "rolling of the bones"				
Throw				Probability
1111	6666			1×10^{-4}
3333	4444			256×10^{-4}
1113	1114	6663	6664	16×10^{-4}
3331	3336	4441	4446	256×10^{-4}
1116	6661			4×10^{-4}
3334	4443			1024×10^{-4}
1133	1144	6633	6644	96×10^{-4}
3344				1536×10^{-4}
1166				6×10^{-4}
1134	6634	3316	4416	192×10^{-4}
3314	3364	4413	4463	768×10^{-4}
1163	1164	6613	6614	48×10^{-4}
1346				384×10^{-4}

The astragalus was popular in the ancient Greece and Rome, and then as now the authorities tried to control gambling in various ways. The Roman Emperor Augustus (63 B.C. – 14 A.D.) banned gambling except for the Saturnalia (which is the month December). However, he himself was the first to break the ban. In a letter to the later Emperor Tiberius (42 B.C. – 37 A.D.) Augustus writes: *I dined dear Tiberius with the same company, . . . We gambled like old men during the meal both yesterday and today, for when the astragali were thrown whoever turned up the dog or the six set a denarius in the pool for each of the one of the astragali, and the whole was taken by anyone who threw the Venus.*

The most popular gambling tool of all times is the die. The first die might have been made by carving astragali to a more symmetric shape. The earliest known die dates back to $\approx$ 3500 B.C. The ancient dice seems to have been numbered 1–6 in a random way; however, from the 18th Dynasty (Egypt, 1370 B.C.), die makers settled on the three 2–partitions of 7, that is, 1 opposite 6, 2 opposite 5, and 3 opposite 4. There is no evidence that the Greek or the Romans made any attempt to compute odds or had a notion of quantitative or even qualitative probability, in spite of the fact that they had the necessary knowledge of mathematics to do so. Actually most of the ancient dice found are clearly skew, and it seems as if there were no attempts to make an "honest" die. There may be many reasons for this. First, due to the skewness of the astragalus it would be difficult to find regular patterns; second due to the number system of the Greeks and Romans the elementary arithmetic (in particular, multiplying and dividing numbers) was a difficult task only known to experts; the third and in my opinion the most probable reason is that there were no notion of randomness in the sense we have today. Serious gamblers then as now don't believe in randomness, but believe the die is governed by the Gods or Destiny or Luck. Actually in many religions the deity is supposed to reveal his (or hers) will in some holy random experiment, such as drawing lots or picking a white or black ball at random from an urn.

The medieval Europeans gambled eagerly with astragali, dice, and various board games, and the Church and State tried to ban or restrict gambling with the same negative result as today. For instance, the English crusaders in 1190 with rank below Knight were banned from gambling and knights were allowed to lose at most 20 shilling per day in gambling. When Bishop Wibold of Cambray ($\approx$ 960 A.D.) couldn't prevent the monks from gambling, he invented a moral game of dice. Three dice were the thrown, and to each of 56 possible outcomes he assigned a virtue, and the monk was supposed to exercise the assigned virtue for a certain length of time. The game never became popular, but it is interesting because of the fact that here we find the first recorded attempt at counting the total number of outcomes in a random experiment. In a Latin poem called *De Vetula*, which usually is ascribed to the Chancellor of the Amiens Cathedral, Richard de Fournival (1200 – 1250), we find the complete count of all possible outcomes in a random game. Again three dice are thrown, and the number of points is counted, in a table to "De Vetula" we find a lists of the number of ways that the sums 3, 4, 5, . . . , 18 can occur in one throw with three dice (called Cadentia, see Table 2), together with a list of the number of ways that the numbers 3, 4, 5, . . . , 18 can be written as a sum of three numbers between 1 and 6 (called Punctatura, see Table 2). It is uncertain why the table was made and there is no evidence, that the numbers "Cadentia" have been used for computing odds or probabilities. But from a manuscript, we know how the numbers were computed: All the 56 possible outcomes of one throw with 3 dice were listed, and a simple count of the appropriate numbers was made.[5] Note

---

[5] Pascal's triangle, that is, the binomial coefficients was known in Europe from $\approx$ 1300 A.D., and long time before that by the Arabs and Chinese. However, the relation between the binomial coefficients and combinations was first observed by B. Pascal in 1654, and what we nowadays call Pascal's triangle was called the arithmetic triangle.

that the unknown author of Table 2 has observed the symmetry between 3 and 18, 4 and 17 etc., and that the sum of the punctaturas equals 56, which is the number of possible outcomes discarding permutations, and the sum of the cadentias equals 216, which is the total number of outcomes including permutations.

Table 2
From the poem: *De Vetula*

3	18	Punctatura	1	Cadentia	1
4	17	Punctatura	1	Cadentia	3
5	16	Punctatura	2	Cadentia	6
6	15	Punctatura	3	Cadentia	10
7	14	Punctatura	4	Cadentia	15
8	13	Punctatura	5	Cadentia	21
9	12	Punctatura	6	Cadentia	25
10	11	Punctatura	6	Cadentia	27

Punctatura: The number of ways that a number can be written as a sum of 3 integers between 1 and 6.
Cadentia: The number of ways that a number can be obtained as the sum of points in one throw with 3 dice.

At the end of the 15th century some sort of notion of qualitative probability seemed to emerge. In a commentary to De Vetula, the unknown author claims that the sums 10 and 11 are more likely to show than the sums 3 and 18, but there is still no evidence of a quantitative notion of probabilities or odds. The first recorded probabilistic problem is found in the book *Summa de Arithmetica, Geometria, Proportioni e Proportionalitá* by Luca Pacioli (1445–1509). Luca was born in Toscana near the village Borgo. As young man he studied mathematics in Pérouse, Rome, Naples, Pisa and Venice, and later he became professor in mathematics in Milan and Firenze. He didn't contributed in any significant way to mathematics, but in his book *Summa de Arithmetica, Geometria, Proportioni e Proportionalitá* (1494) he summarized virtually all mathematics known at that time, and the book became a standard textbook of mathematics for a long period of time. The book contains a wealth of exercises, and many exercises of modern elementary mathematics textbooks are descendants of the exercises in *Summa de Arithmetica,* *Geometria, Proportioni e Proportionalitá.* In a section entitled "Various Exercises" we find the following exercise, see also Exc.2.53:

> Two men A and B are playing "balla" (probably a medieval ball game). Both have set 10 gold coins at stake, and the one who first wins 6 games gets the 20 gold coins at stake. However, at a point where A has won 5 games and B has won 3 games, the play is interrupted. It supposed that the two men are equally skillful; how should they divide the stake?

Luca gave the answer 5:3; in 1556 N. Tartaglia (1499–1557) objected to the solution and proposed the answer 2:1; two years later G. F. Peverone claimed that the answer rightfully should be 6:1; and in 1654, B. Pascal gave the correct answer 7:1. Luca didn't considered the exercise as a problem of probability, but as a problem of the theory of proportions. It is striking that it should take 160 years for the foremost mathematicians to solve an exercise, which any student with minimal knowledge of probability nowadays can solve in a few minutes. The story shows that the notion of probability is a modern notion, which simply was not available – not even in an intuitive or informal form – in the days of Luca Pacioli.

The Emergence of Probabilities

Around the year 1500 some notion of qualitative probabilities had emerged and around 1550, Gerolamo Cardano wrote a little book *Liber de Ludo Alea* (The Book on Dice Games), in which he develops a theory of odds and probabilities. The two major results of the book are:

- The probability that one of two exclusive events occurs equals the sum of their probabilities
- The probability that two independent events occurs simultaneously equals the product of their probabilities

The book appears more as a diary than as a book. At the beginning of the book Cardano has several wrong forms of the two propositions above; for instance, he claims that since the probability of getting "1" in one throw with a die equals $\frac{1}{6}$, then the probability of getting at least one "1" in two throws with a die must be $\frac{2}{6}$. But later he realize the mistakes and gives the correct answers without removing the false statement. The book was never published, actually Cardano kept it secret,[6] and the book was first discovered more than hundred years after his death. But at that time the theory of probabilities had been rediscovered by P. Fermat, B. Pascal and C. Huyghens. Cardano had a stormy life of which you will find a brief account of in the following.

Gerolamo Cardano (1501–1576): Gerolamo Cardanó was born the 24th of September 1501 in the Italian city Pavia. He was the illegitimate son of Facio Cardano and Chiara Michena.[7] Facio was a lawyer and he practiced medicine in Milan; he was also known as good geometrist,[8] and he taught geometry at the university of Milan. Chiara is described as a short, fat, healthy woman with a lively wit and a quick temper. Facio kept her as mistress in a separate house, but in 1506 he move into the house of Chiara, and he married her shortly before his death.

We have a most detailed description of the life of Gerolamo Cardano. Partly from his strange and extremely frank (and often self-contradictory) autobiography *De Vita Propia.* He was a man of great capacities and of great contradictions; he had a stormy life, and he was a true exponent of the new Renaissance man.

Facio instructed his son in mathematics, but Gerolamo insisted on studying medicine. Facio was opposed to Gerolamo's plans, but Chiara supported her son, and as usual the woman got her way. So after a violent quarrel with his father, Gerolamo went to the University of Pavia to study medicine in 1520. They were uneasy days; in 1521 King Francis I of France and Emperor Charles V of Spain started a war in Italy.[9] Due to the unrest, Cardano first moved to Milan and later to Padua. He became more known for his diligence at the gambling table than his diligence at the studies, and we know that he succeeded in losing the inheritance from his father at the gambling table in a short time. According to his autobiography he was elected Rector at the University of Padua in 1525 while still a student.[10]

After having failed twice he became a doctorate of medicine in 1526. His exam was so poor that he could not become a member of the medical societies of Milan, Pavia or Padua.[11] In those days membership of the Medical Society was necessary for practicing medicine, so in 1527 he went to a small village Sacco to practice there. Here he fell in love with Lucia Bandarina, the daughter of the local

---

[6] Cardano was a passionate gambler, and since he was the sole person in the world having a notion of odds and the capacity of computing them, it gave him a significant advantage over his opponents, and a good reason to keep his findings secret.

[7] Cardano hotly denies his illegitimacy, even though it was quite common and socially accepted.

[8] According to some sources, Leonardi da Vinci occasionally sought the advice at Facio on geometric matters

[9] France and Spain was the two great powers at the time of Cardano and, as today, the great powers preferred to fight their wars on other countries territory.

[10] The claim seems unlikely, it was not unusual that a student was elected as Rector, but usually the Rector was a person of a high social rank, and according to the archives of the University of Padua no Rectors was elected in the period 1509–1527.

[11] Cardano ascribes this to his illegitimacy, but again the claim seems unlikely.

innkeeper, and they married in 1531. In 1532 he tried again to enter the Medical Society of Milan, but again he was rejected, and he went to another poor village Gallerat to practice. He shared his spare time between the gambling table[12] and the study of mathematics. In 1534 he was ruined, and he moved into the poorhouse of Milan together with his wife and his newborn son, Giambatista (1534–1560). At this low point of his life his luck changed, as so many other times in the years to come. Through the influence of good friends of his father he obtained a lecturing position in geometry at the University of Milan. At this point, his mother Chiara pardoned his loss of the inheritance, and she moved into the house of Gerolamo. Soon after, Cardano wrote his first book on mathematics (1539) and in his lifetime he wrote 150–200 books on all kinds of subjects, for instance, medicine, law, mathematics, philosophy, pedagogics,[13] etc. Being extremely tenacious, Cardano finally was accepted as a member of the Medical Society of Milan. And later in his life he became known all over Europe partly for his mathematical and mainly for his medical skills.[14]

The time from 1547–1557 was a happy time for Cardano with prosperity and fame, but in 1557 luck changed again. His eldest son Giambatista married a prostitute of Milan. The marriage was unhappy; in 1560 Giambatista had a fierce quarrel with his wife, at which she told him, that he was not the father to any of their three children. After the quarrel Giambatista went down to the kitchen, where he made a bread stuffed with white arsenic; he served the bread to his wife and her family. The wife died and her family was close to join her. Giambatista was soon caught, and together with a servant, who assisted with the baking, he was accused of murder. He claimed himself not guilty, and his father Gerolamo, who also was practicing law, defended his son, but in vain. On a gloomy November day in 1560 Giambatista was executed for the murder of his wife. This was a serious blow for Gerolamo, and so much more because Giambatista was his favorite son. Gerolamo had three children, two sons Giambatista and Aldo, and one daughter Chiara. His younger son Aldo (1543–??) behaved very badly from a young age and became involved in a criminal gang at the age of about 13. At the age of 18, Aldo had already payed several visits to several Italian prisons. Shortly after the execution of Giambatista, Aldo's gang visited Gerolamo, locked him up in the cellar and robbed the house. In 1562 Gerolamo could no longer bear his reputation as the father of a murderer and a notorious outlaw, and he moved to Bologna, where good friends had offered him a professorship at the University of Bologna. But the suffering and

---

[12] According to his autobiography he earned an essential part of his income by gambling. Since he was ruined many times in his life, he couldn't possible be as lucky as he claims in his autobiography.

[13] Cardano wrote a book on how to raise you sons properly; however, then as now theoretical pedagogues seems to have difficulties with the practical raising of their own children. Of the two sons of Cardano one was executed for murder, and the other became a notorious outlaw and gang leader.

[14] In 1552 he was called to Scotland to treat the Archbishop, John Hamilton, for astomia. Cardano cured the Archbishop telling him to avoid feathers of all kinds and to sleep with furs instead of feather linen. The cure worked, and the Archbishop became a loyal friend of Cardano for the rest of his life.

the depressions over his two sons had brought him to madness, and in his last years he was insanely paranoid, spending all his time fighting real but mostly imaginary enemies. His lectures at the university was avoided by the students because of his strange behavior. He was expelled from his position, accused of unnatural sexual behavior with minor boys. This might have been an excuse from the university to get rid of an old, paranoid, strange man, but it seems that his daughter Chiara believed the accusations telling him in a letter, that she was ashamed of his conduct. This brought Gerolamo still further into madness.

But fate had reserved a last blow for Gerolamo. On October 6, 1570 he was arrested by the Inquisition,[15] accused of heresy. The cause for the accusation has never been revealed[16], but probably it was an act attributed to the counterreformation movement led by the severe and orthodox Pope Pius V. In Gerolamo's books there were certainly many things which could displease a strict Inquisitor. Gerolamo Cardano was not a religious man, and he had engaged himself in may different kinds of divinations, for instance, astrology, chiromancy,[17] etc. After a few months in prison, Gerolamo was released, but he was banned from publishing and writing. His autobiography, which was concluded in secret after his release, contains no hints about the trial and the exact charges. He spend his last years in the house of his former student Rodolfo Silvestri. Finally, death released him from his sufferings in 1576.

In his lifetime he published 131 books, and after his death 111 more manuscripts were published, and it was said that just before his death he burned another 170 manuscripts. He engaged himself in all kind of sciences, and there is no doubt that he was a genius, but he had a bad reputation of plagiarizing others peoples' work or inventions without due reference. He probably did not invent the cardan shaft, but he had a horse carriage with a cardan shaft made as a present for the Spanish military governor Alfonso d'Avalos Machese; nor did he invent the suspension of the ship compass, which today bears his name, but he did invent a new way of raising a sunken vessel, a new kind of chimney improving the draft, a new method of distilling wine to make better brandy, and many other ingenious inventions. He was always filled with new ideas, but often the ideas were more theoretical than practical.

The Rise of Probability Theory

As mentioned in the previous section Cardano kept his findings about probabilities secret, and for a period of almost 100 years after the death of Cardano

---

[15] The Inquisition was a religious court for the Catholic Church. It was responsible for the execution (usually by burning) of numerous persons who were found guilty in heresy.

[16] The court trials at the Inquisition was usually kept secret.

[17] Chiromancy is the art of telling the future of a person by reading the lines of the hand.

there were no attempt to make a general theory. In this period we only find few scattered investigations of particular probabilistic problems. The real breakthrough came in the summer of 1654, where Pierre de Fermat and Blaise Pascal laid down the foundation of elementary probability theory in a series of letters. All the letters except the first (from Pascal to Fermat) have been saved. The second letter is undated and starts as follows:

> *Monsieur*
> *Si j'entrepends de faire un point avec un seul dé en huit coups . . .*

(In English: *Sir, If I try to make a certain score with a single die in eight throws . . .*). The letter is an answer to a lost letter from Pascal to Fermat in the letter Fermat pointed out a mistake of Pascal, and in the response (dated July 29) Pascal acknowledged the mistake and reported a series of solutions to a variety of probabilistic problems. The correspondence went on all summer and fall until October 27, and in the letters Fermat laid down the theoretical foundation for elementary probability theory, which is the lottery model described in (2.15), and Pascal applied the theory of Fermat to solve a large variety of problems. For instance, in a letter dated August 24 Pascal solves the problem of Pacioli mentioned above, and he generalizes the result to an arbitrary number of plays interrupted at an arbitrary position. In November 1654 Pascal had a bad horse carriage accident, and the correspondence ended at this time. On July 25, 1660 Fermat wrote to Pascal and suggests a meeting, but Pascal responded that he is too ill to travel. It is uncertain what triggered Pascal's investigations into probability theory, he wasn't a gambler, but he knew a notorious gambler, Chevalier de Meré [see Exc.2.54], who may have fed Pascal with problems of chance games.

Pierre de Fermat (1601–1665): Fermat was born in the village Beaumont-de-Lomagne in Gascogne near Toulouse. He studied law at the University of Toulouse, and after his graduation in 1631 he was appointed as "Conseilleur de la Chambre des Requêtes du Parlement de Toulouse", that is, as counsellor at a special court taking care of civil lawsuits. He advanced slowly but steadily, and he finally, became the head of the court. In his spare time he studied mathematics and languages with great success. Considering that mathematics was only one of the hobbies of Fermat, he reached astonishing results, and he was one of the leading mathematicians of his time. Unlike mathematicians before and after, public recognition meant little or nothing to Fermat; most of his findings are contained in his huge correspondence with scientists in France, England, Italy, Germany and The Netherlands. He had only little personal contact with other scientists. He seems to be one of the rare persons who bloom in isolation, and his modesty made him believe that a rough sketch of a proof or the mere formulation of a theorem were sufficient for everybody to fill in the details.[18] The most famous and difficult problem of mathematics goes under the name *Fermat's last problem.* The problem is the existence of solutions to *the Diophantic equation*: $x^n + y^n = z^n$, where $n \geq 2$ is a given and known integer, and x , y and z are unknown. A Diophantic equation is an equation where only integers are allowed as solutions; it is named after the Egyptian mathematician Diophantus ($\approx$ 250 A.D.) who wrote a book on the subject and found a formula for the integer solutions to $x^2 + y^2 = z^2$ [there are infinitely many solutions, the most well-known is $(3,4,5)$]. In a copy of Diophantus' book belonging to Fermat we find the following note in the margin:

> *I have found a wonderful proof that a cube cannot be divided into two cubes, and a biquadrate cannot be divided into two, etc. . . .*

Thus, Fermat claims that there are no integer solutions to the equation $x^n + y^n = z^n$, when $n = 3$ or 4 and possibly also for $n \geq 5$, but due to difficulties in reading the handwriting of Fermat that is, somewhat uncertain. For three centuries the foremost mathematicians have tried to prove the claim of Fermat. In 1770 the claim was proved for $n = 3,4$, in 1808 the claim was proved for $n = 3,4,5,6$, in 1910 the claim was proved for $3 \leq n \leq 31$, in 1954 the claim was proved for $3 \leq n \leq 619$, in 1982 the claim was proved for $3 \leq n \leq 253{,}747{,}889$. Thus, we are progressing, but there is a long way to infinity. The latest news (January 1994) is, that the theorem is claimed to be proved by A. Wiles, but there seems to be several gaps in the proof and at this time the proof is incomplete.

Blaise Pascal (1623–1662): Pascal was a wonder kid with a fragile health. His father, Étienne Pascal, was judge in Clermont-Ferrand, and he is described as a competent mathematician. In 1631 the family moved to Paris to stimulate the education of Blaise; in 1639, only 16 years old Blaise wrote a remarkable paper on conical sections (*Essai pour le coniques*). Two years later he invented the first mechanical computer, which could add, subtract and multiply numbers.[19] In 1646 Blaise and his family converted to a Christian sect called Jansenism.[20] The Jansenists tried to reform the Catholic Church and were strong opponents of the Jesuits, but the sect was finally (1656) declared heretic by Pope Innocent X, and the members were persecuted. Jansen emphasized the "internal life" and he claimed natural sciences as a perversion against the will of God. Blaise was at first not a serious believer of Jansenism, and he continued his scientific work. But in the fall of 1654 he had a serious horse carriage accident, which he took as a sign

---

[18] In his last letter of the 1654 correspondence he formulated a series of number theoretical theorems, and writes that he was sure that Pascal easily could fill in the proofs. Pascal couldn't prove any of theorems!

[19] There still exists a few samples of Pascal's computer.

[20] Named after the Bishop of Ypres, Cornelius Jansen (1585–1638).

of God, and went to the Jansenistic monastery in Port Royal outside Paris, where his sister was a nun. After that Pascal gave up his work with natural sciences.[21]

In 1655 a young Dutchman named Christiaan Huyghens (1629–1695) made a stop in Paris on his way to complete his studies in mathematics and physics at the University of Anger. At that time, the Parisian mathematicians met regularly in the house of Carcavi, and our young man carried a letter of introduction from his mathematics teacher Francis van Schooten to the Carcavi circle. He never met Fermat, who lived in Toulouse, or Pascal, who had withdrawn to the monastery of Port Royal, but he learned about their findings. However, he was only told about the results – not the proofs and arguments. He became intrigued by the problems of chance games, and he wrote a small book on probability theory. The original version was in Dutch, but it was soon translated into Latin, and was published in 1656 bearing the title *De Ratiociniis in Alea Ludo* (How to Reason in Dice Games). The book was warmly received by his contemporary fellow mathematicians, and for half a century it was the unique introductory text to probability theory, and it is fair to say that the book represents the real beginning of probability theory as a mathematical subject. *De Ratiociniis in Alea Ludo* consists of 14 problems with complete solutions. In the first three problems, Huyghens defines *the expectation*[22] of a random experiment [see (4.1)] and he uses the notion in a very effective and ingenious way to solve a series of nontrivial probabilistic problems. For instance, the 14th and last problem of *De Ratiociniis in Alea Ludo* goes as follows:

> If another and I throw in turn and turn about with 2 dice on the condition that I will have won when I have thrown 7 points, and he will win, when he has thrown 6; if I let him throw first find the ratio of my chance to his.

Thus, we have two persons, say A and B, throwing two dice in turn with the rule, that A begins and wins if he get 6 points before B obtains 7 points. Note that the probability of getting 6 and 7 points, respectively, in one throw with two dice equals $\frac{5}{36}$ and $\frac{1}{6}$, respectively. Thus, the probabilities are slightly disadvantageous to A, but A has the advantage of beginning. In the solution, Huyghens applies expectations in a very clever way, and his argument (in a slightly modernized language) goes as follows: Let a be the stake, and let x be expected winnings of B, then $a - x$ is the expected winnings of A. Moreover, let y be the expected winnings of B on the condition that B begins. If A is to throw, he has a probability of $\frac{5}{36}$ of winning immediately, and so $x = \frac{31}{36} y$, and if B is to throw, he has a probability of $\frac{1}{6}$ of winning immediately, and so

---

[21] In 1658 he had a slight fall back. According to his sister he had toothache, and in order to forget the pain he wrote a short paper on the cycloid (that is, the curve described by a fixed point on a rolling wheel). The paper was published under a pseudonym.

[22] In Huyghens notation, "the value of my chance".

$y = \frac{1}{6}\,a + \frac{5}{6}\,x$. Solving the two equations gives $x = \frac{31}{61}\,a$ and $a - x = \frac{30}{61}\,a$. Thus, if p and q are the probabilities that A and B respectively wins, then we find that $p = \frac{30}{61}$ and $q = \frac{31}{61}$, which shows that the advantage of beginning increases A's probability of winning.

Later Huyghens went into physics, where he laid his major scientific work, and he only returned to probability theory a few times in his life. He is best known for his theory of light, which became the theoretical background for the construction of microscopes and telescopes,[23] he also invented the pendulum clock and showed, that the time of one oscillation of a pendulum rolling on a cycloid doesn't depend on its position.[24] In 1669 Huyghens brother Ludovick wrote a letter to Christiaan telling him that he had made a table of life expectations to be used for computing the premiums of a certain kind of life insurance called "tontine". In his reply Christiaan pointed out the difference between the expected and probable lifetime, and that it is the latter which is needed for computing the premium of life insurance.[25]

The Origin of Statistics

The original meaning of the word "statistics" is "science of states", and in its early existence it was also called "political arithmetic". Like probability theory, statistics owe its existence to a small object – much smaller than the astragalus – but in spite of it smallness it is the most vicious creature who had ever existed on this earth. It is called *Yersina Pestis*, and it is the bacteria which causes the most infectious and lethal decease of all times – *the plague*. In 1346 A.D. the first plague epidemic hit Europe. It arrived at Sicily from the Middle East and it moved on to all of Europe (including remote parts like Iceland and Greenland), and when it finished in 1352 it had killed approximately one-third of the population of Europe (the estimated death toll is 50,000,000 persons). After that time the plague occurred regularly until about 1712. In Table 3 you find the years of the plague in Copenhagen and London. The last outburst of plague in Copenhagen occurred in 1711; in Table 7.1 you will find a table of the weekly mortality in Copenhagen. In all 22,535 persons out of a total population of approximately 60,000 died in Copenhagen that year, mostly due to the plague. Usually the plague would kill about one-third of the population. The bacteria *Yersina Pestis* has a complicated life cycle, which includes three hosts, rats, rat lice and humans. It is lethal to all three hosts and death is caused by prussic acid produced by the bacteria. Prussic

---

[23] The microscope and telescope was invented before that time, but was improved largely by the methods of Huyghens.

[24] The first pendulum clocks were modified for the "circular inaccuracy," but it was soon found that the correction was marginal to the other sources of inaccuracy.

[25] This is still a common mistake, in particular, in the computations of risk probabilities for dangerous industrial plants such as nuclear power plants and chemical plants.

Table 3

Plague epidemics in Copenhagen and London	
Copenhagen	1349 ? 1405 ? 1536 1546 1583 1601 1625 1630 1637 1655 1665 ? 1711
London	1348 ? 1406 ? 1537 ? 1563 ? 1601 1625 ? 1665 ? ?

The question marks "?" indicate periods where I was unable to find data, but where I am certain that one or more plague epidemics has occurred.

acid is a highly poisonous chemical and causes a blue or dark coloring of the skin; for this reason the plague is also named *the black death*.[26] Around 1712 people recognized the connection between the plague and rats, and from that time we have tried by all means to exterminate the rats,[27] but in vain. However, we have pushed the rats down in the sewers, and thereby stopped the transition of the *Yersina Pestis* from rats to humans.

As mentioned earlier the original meaning of "statistics" is "the science of states". The science of states dates back to Aristotle (384 B.C. – 322 B.C.), whose book *Politeia* contains a description of 158 states. With the emergence of the new ideas of the Renaissance, scientists of states began to collect data such as population number, yearly production of various goods, etc. However, due to lack of real data the numbers was often purely hypothetical or quantitative. The most famous population count of all times took place on Christmas day in the year 1 B.C., when the Roman Emperor Augustus ordered all the inhabitants of the Roman Empire to go to their town of birth to be counted. Population counts even today are a difficult and expensive matter, and it requires a high administrative level, which was present in the Roman Empire, but was lost with its fall, and from 400–1800 there were no real population counts in Europe. Population counts are important for two reasons: They are the basic data needed for estimating the total tax revenue of the state and the size of army that the state can assemble at wartime. In the Middle Ages and the Renaissance, the states of Europe didn't have the administrative capacity to make population counts, but the Church was capable of counting the number of deaths (or rather burials) and the number of newborn babies (or rather baptized babies). The first regular collection of birth and death data goes back to 1538, when the British king Henry VIII introduced the so-called yearly bills of mortality. The year before, The British Isles was hit by the

---

[26] Today the plague is no longer a serious threat to humans; most of the known antibiotics will kill the Yersina Pestis and cure the person.

[27] It is ironic that the human race today are exterminating approximately one species per day without really trying, but the rat which we have tried so hard to exterminate is surviving without any sign of being an endangered species.

plague, and Thomas Cromwell, the real ruler of Britain at that time, ordered the Church to make a count of burials and baptized babies, probably with the purpose of estimating the future military and tax revenue potential. At the beginning these counts were irregular and disorganized, but from 1603 and onward we have regular and reasonably precise counts from the city of London. In Denmark such counts were performed from $\approx$ 1600, but the first reliable counts covering the whole country dates back to $\approx$ 1700 [see Table 7.1].

Political arithmetic: Statistics (or political arithmetic as it was named in its first years of existence) is the art of deducing estimates and properties of quantities which can not be observed directly. The inventor seems to be an English tradesman and haberdasher of small wares, John Graunt (1620–1674). In 1661 he published a book bearing the title *Natural and Political Observations upon the Bills of Mortality*, where he reported about his studies of the bills of mortality. He applied the bills to estimate the birth mortality, the number of inhabitants in London, the number of years to recover the former population level after a plague epidemic, life-tables (that is, a table of the probabilities that an x-year-old person will survive his or hers next birthday), etc. His methods are bold but dubious; however, many of his findings are in accordance with later and more reliable observations. For instance, he found that 36% of all children died before their sixth year birthday – a result which were confirmed by later and more precise computations. In 6 different ways he found that the population of London were $\approx$ 380,000. To give you an idea of his methods, I describe two of John Graunt's techniques to obtain the population number of London:

From the Bill of Mortality for London in the year 1660 John Graunt knew three facts. (1): 13,000 persons died in London in 1660; (2): 12,000 babies were born in London in 1660; (3): Of the 13,000 deaths in London that year, 3,200 occurred inside the walls, that is, in central London (at that time the Bills of Mortality also contained information about the number of burials in the individual parishes of London). Studying a map of central London gave him an estimate of 11,800 houses inside the walls. In his first method, John Graunt assumes that to each house there corresponds a family of 8 persons (man, wife, 3 children and 3 servants), and he thereby obtained the number 380,160 in the following way:

$$\begin{gathered}\text{Number of families inside the walls } = 11,800 \\ \text{Total number of deaths } = 13.000 \\ \text{Number of deaths inside the walls } = 3,200 \approx \tfrac{1}{4}\cdot 13,000 \\ \text{One family } = 8 \text{ persons (man, wife, 3 children, 3 servants)} \\ \Downarrow \\ \text{Total population number } = 8 \times 4 \times 11,800 = 380,160\end{gathered}$$

In his second method, John Graunt assumed that a woman in the age between 16 and 40 years old is pregnant approximately every second year, and that the number of women between 16 and 76 years old is twice the number of the women 16 and 40. Moreover, assuming that to each woman between 16 and 76 there corresponds a family of 8 persons (man, wife, 3 children and 3 servants) he obtained the number 384,000 in the following way:

$$\begin{gathered}
\text{Total number of newborns } = 12,000 \\
\text{One family } = 8 \text{ persons (man, wife, 3 children, 3 servants)} \\
\Downarrow \\
\text{Number of women aged 16-40 } = 2 \times 12,000 = 24,000 \\
\Downarrow \\
\text{Number of women aged 16-76 } = 2 \times 24,000 = 48,000 \\
\Downarrow \\
\text{Total population number } = 8 \times 48.000 = 384,000
\end{gathered}$$

No matter how shaky and dubious the methods of Graunt may seem retrospectively, his investigations were a breakthrough in statistics and the art of estimating quantities, which cannot be observed directly, and his work received due recognition. Soon the probabilists of the time (for instance, Abraham de Moivre, Daniel Bernoulli and Edmund Halley) took up the challenge to invent rigorous methods, particularly to compute reliable life-tables, which were an important tool for the emerging life insurance companies.

Insurance: In 1492 Christopher Columbus discovered the continent of America[28] and, soon after, the sail routes south of Africa and America to India, China and Japan were discovered. In the Middle Ages trade with India, China and Japan followed long and dangerous caravan routes through Asia to the ports of the Middle East, where the goods were shipped to the Italian city-states (Venice, Firenze, Genova etc.). The new sail routes cut the cost of transport immensely, and they were very profitable but also very risky. The high risk called for insurance, and the first marine insurance was established in the beginning of 16th century in Northern Italy. Originally, insurance was an ordinary bet; for instance, a merchant sending a ship on an intercontinental voyage could make a bet with his fellow merchants that his ship would not return in safety. If the ship was lost he would thereby get some compensation, and if the ship returned in safety, he would have to share some of his profit with the bet holders. But soon insurance was organized by companies. In 1601, the British Queen Elisabeth I passed the first law of insurance, and in 1688 Edward Lloyd opened a coffee house on Tower street, London. There as in other coffee houses, merchants, seafarers and insurance brokers assembled

---

[28] Or rather the sail route from Europe to America, the continent was, of course, discovered centuries before by the Indians.

to drink coffee and to make business. From its informal beginning, Lloyd's coffee house grew in repute and influence, and from 1700 on he published *Lloyd's News*, giving information of ship movements and other information of interest to marine insurance brokers. As time went by, the publication grew into what nowadays is entitled *Lloyd's Register of Shipping*. Today, *Lloyd's* is a insurance market with about 5,000 *underwriters of Lloyd* (that is, insurance companies associated to Lloyd's) and about 2,000 *insurance brokers* (that is, companies which on the behalf of the policy holders seeks the best possible insurance contract). Life insurance came into existence at the same time as marine insurance. At the beginning, life insurance like marine insurance was ordinary bets, but soon it was organized in life-companies. In the 17th and 18th centuries *the tontine*[29] was a popular form of life insurance. The tontine policy provided a yearly refund at the end of a definite period such as 15 or 20 years until the death of the policy holder, and usually the premium was paid in total at the signing of the policy contract. It was frequently used by cities to raise money for some definite projects. In the middle of the 18th century fire insurance was added to the options of insurance.

It is of vital interest for an insurance company to be able to compute reliable risk probabilities, such as "the probability that a cargo ship bound for China will return in safety" or "the probability that a man aged 35 years will survive his 70th birthday". Inspired by the works of John Graunt, scientists of the time tried to construct more rigorous methods of computing reliable life-tables. Since 1584, regular and precise recordings of funerals (including the age of the deceased) and births was maintained by the Silesian city Breslau. The clergyman and scientist Kaspar Neuman of Breslau was attracted by the materials and communicated his findings to the Royal Society in London. The society forwarded the material to the English astronomer Edmund Halley (1656–1742) who published a remarkable paper in 1693 with the title *An Estimate of the Degrees of Mortality of Mankind Drawn from the Curious Tables of Births and Funerals of the City of Breslaw.* In the paper Halley invented a rigorous and precise method of computing life-tables. The method of Halley goes as follows: Let a_x denote the (unknown) number of living aged x years, let $a(x,y)$ denote the (unknown) number of living of age $\geq x$ and $< y$, and let d_x denote the (known) number of deceased in the previous year aged x years. Assuming that $a_{101} = 0$ and that the survival probabilities are constant from year to year we get $a_{x+1} = a_x + d_x$ from which we obtain the following formulas for the unknown quantities a_x and $a(x,y)$:

$$a_x - a_y = \sum_{j=x}^{y-1} d_j \quad , \quad a_x = \sum_{j=x}^{100} d_j$$

$$a(x,y) = (y-x)\sum_{j=y}^{100} d_j + \sum_{j=x}^{y-1} (j-x+1)\, d_j$$

---

[29] Named after Lorenzo Tonti.

Table 4

The probabilites of human Life, according to Dr. Halley
From Abraham de Moivre: "Doctrine of Chances"

Age	Living	Age	Living	Age	Living	Age	Living	Age	Living	Age	Living
1	1000	16	622	31	523	46	387	61	232	76	78
2	855	17	616	32	515	47	377	62	222	77	68
3	798	18	610	33	507	48	367	63	212	78	58
4	760	19	604	34	499	49	357	64	202	79	49
5	732	20	598	35	490	50	346	65	192	80	41
6	710	21	592	36	481	51	335	66	182	81	34
7	692	22	586	37	472	52	324	67	172	82	28
8	680	23	580	38	463	53	313	68	162	83	23
9	670	24	574	39	454	54	302	69	152	84	19
10	661	25	567	40	445	55	292	70	142	*	*
11	653	26	560	41	436	56	282	71	131		
12	646	27	553	42	427	57	272	72	120		
13	640	28	546	43	417	58	262	73	109		
14	634	29	539	44	407	59	252	74	98		
15	628	30	531	45	397	60	242	75	88		

Having computed the numbers a_x it is an easy matter to compute the probability that a person aged x years will survive his next birthday. In de Moivre's book *Doctrine of Chances* we find Halley's life-table [see Table 4] which shows the expected evolution of a group of 1000, 1-year-old babies. I will leave the history of statistics at this point and turn to the progress of probability; in the text you find more historical notes, for instance, in (4.38) you will find a brief account of the history of taking averages, and in (4.39) you will find Daniel Bernoulli's solution to the so-called insurance paradox.

The Progress of Probability theory

In the beginning of the 17th century Nicholas Bernoulli moved from the city of Antwerp in The Netherlands due to the severe government of the Duke of Alba.

After a while he settled down with his family in the Swiss city of Basel, and he soon established himself as a merchant banker. On Christmas Day 1654, his oldest son James Bernoulli was born. James Bernoulli (1654–1705) became the first of a series of eminent mathematicians of the Bernoulli family. From 1676 to 1682 he travelled between the European universities learning science and mathematics. On his journeys he learned about Huyghens book *De Rationiciis in Alea Ludo*, and he became attracted by the subject. From 1690 until his death in 1705 he worked on a book on probability theory entitled *Ars Conjectandi* (The Art of Conjecturing). The book was left unfinished at his death, and the publication was delayed for eight years by his brother John Bernoulli under the pretence that his nephew Nicholas Bernoulli was to complete the book. John became an eminent mathematician like his elder brother James, but he never ceased to envy him.[30] The mathematical community awaited the publication of *Ars Conjectandi* with excitement and finally, John had to publish the unfinished manuscript in 1713. The book contained a wealth of new results, in particular, the theorem which we now call "the law of large numbers" and which James named "the golden theorem" [see (4.8)–(4.12)]. The law of large numbers is, together with the central limit theorem which was discovered by de Moivre in 1733, unquestionably the most important results of probability theory. In the last and unfinished part of the book, James tried to apply the law of large numbers to make a solid foundation of daily-life probabilities, such as "the probability that a man aged 35 years will survive his 70th birthday," which indeed not were covered by the simple lottery model of Fermat and Pascal; but as so many of his successors without success – probability theory had to wait until 1933 before a firm foundation was found.

Several years later Abraham de Moivre (1667–1754) took up the tread in a most ingenious way. De Moivre was born in the French village Vitry; he studied mathematics at Sedan and later at the Sorbonne University in Paris. Together with his family he belonged to the Protestant sect called the Huguenots. On October 18, 1685 the Edict of Nantes[31] was evoked, and the 18-year-old Abraham was imprisoned together with many other Huguenots. Three years later he escaped the prison and fled to England. There he arrived 21 years of age, with no money and no friends, but with a solid knowledge of mathematics. He set up a small school of mathematics on Fleet Street in London, and all his life he had to walk from house to house giving mathematics lessons to children of wealthy people. Besides his work as schoolmaster he also earned money advising gentlemen about the proper odds in betting, but he doesn't seem to have been engaged in gambling otherwise.[32] In 1718, he wrote a marvellous book *Doctrine of Chances* (revised in

---

[30] James was indeed a better mathematician than John.

[31] The Edict of Nantes was a cease-fire treaty between the French King and the Huguenots allowing them to exercise their religion in certain cities of France.

[32] Cardano is the only probabilist of the past, who was seriously engaged in gambling, and for a good reason: In the dedication to "Doctrine of Chances" de Moivre writes: *There are many people in*

1738 and 1756). In 1733 de Moivre discovered the second pearl of probability[33] – the central limit theorem – and the result was included (without proof) in the second and third edition of *Doctrine of Chances*; in (4.22), (5.1) and (5.16) you will learn more about de Moivre's work on the central limit theorem. The book also contains a wealth of solutions to nontrivial problems (even measured with today standards) among these the so-called "ruin problem", which you can find in Sections (11.12+13) of Volume II.

De Moivre never found the normal density function, but he found a series expansion of the normal distribution function. Probability had to wait ≈ 60 years before de Moivre's theorem was conceived as we do today. In 1808 the American surveyor Robert Adrain (1775–1843) published a remarkable paper, in which he suggested the normal density as the proper distribution of errors of measurements. One year later, Karl Friederich Gauss (1777–1855) published a paper with the same conclusion. At that time United States was far from the leading mathematical circles, and the paper of Adrain passed unnoticed. Because Gauss was, without competition, the leading mathematician of his time, the normal distribution is also called the Gauss distribution [see (4.22)]. In the period 1770–1820, Pierre-Simon de Laplace concluded the works of J. Bernoulli, A. de Moivre and K. F. Gauss, and the result is included in his famous book *Théorie Analytique des Probabilités* (1812, 1814, 1820). In the book, Laplace extends the central limit theorem of de Moivre and demonstrates the connection to the newly found normal distribution. Moreover, he essentially exhausted the possibilities of the classical Fermat-Pascal model of probability theory. For the rest of the century, probability theory was on the move to find a new foundation which would include the real-life probabilities of the statisticians and insurance companies.

The Search for a New Foundation

After the publication of Laplace' book *Théorie Analytique des Probabilités*, probabilists set out to extend probability beyond the classical Fermat-Pascal model, but with the problem of lacking a firm mathematical foundation. The trail was found by Lvovich Pafnufty Chebyshev (1821–1884), who introduced for the first time the random variables on the stage of probability. He defined a random variable as "a real variable which can assume different values with different probabilities", and he proved his famous inequality [see (4.4.4)], and the first version of the central limit theorem for general random variables. The original proof of his

---

the world who are prepossessed with an opinion that the doctrine of chances has a tendency to promote play; but they will soon be undecieved . . . it is rather a guard against it, by setting in a clear light, the advantages and disadvantages of those games wherein chance is concerned.

[33] The three pearls of probability are the law of large numbers, the central limit theorem and the law of the iterated logarithm.

inequality was extremely long and complicated but was simplified to the three-line proof we know today by his student Andrei Andreiwich Markov (1856–1922), who together with Alexander Mikhailovich Lyapounov (1857–1918) (likewise a student of Chebyshev) corrected and extended the central limit theorem of Chebyshev [see (5.16)]. Around the turn of this century the status of probability theory was such that it was considered as a part of physics instead of mathematics. In the year 1900 a large conference on mathematics was held in Paris, and the leading mathematician of that time, David Hilbert (1862–1943), was invited to give his opinion of the agenda of mathematics in the 20th century. Hilbert did that with an astonishing prevision in the form of 23 problems,[34] which he predicted would be the central mathematical themes of this century. In his sixth problem Hilbert ask for axiomatizations (with as few and simple axioms as possible) of those parts of physics in which mathematics is used, and he explicitly mentions mechanics and probability theory as two physical subjects to be axiomatized. After the lecture Émile Borel suggested that one could use the unit interval with the Lebesgue measure, which Borel had invented the year before, as a model for "continuous" probabilities. A few years later Rademacher constructed a sequence of independent functions on the unit interval taking the values ± 1 with probability $\frac{1}{2}$, and Steinhaus succeeded in constructing a sequence of independent normally distributed functions on the unit interval. However, the unit interval with the Lebesgue measure imposes too few degrees of freedom,[35] in particular, when it comes to complicated objects like Brownian motions.

Brownian motions: In 1827 the English botanist Robert Brown studied the irregular movements, which later was named Brownian motions, of small particles in water. These motions were discovered long time before Brown was born, and the general accepted theory (the so-called vitalism) at that time was that the particles were alive. Brown rejected the theory and showed that small particles of stone or glass performed the same kind of movements; however, he was not able to find a satisfactory theory, and the problem of Brownian motions was discussed among physicists for the rest of the 19th century until Albert Einstein in 1905 suggested the probabilistic model, which we use today [see (9.8) and (9.10)]. Actually, Einstein was surpassed by the economist Bachelier, who in 1900 suggested exactly the same model as Einstein, but as a model of the fluctuations on the French stock market.[36] Einstein's model was later confirmed in a series of experiments performed by Jean Pierre Perrin, who received the Nobel Prize in 1926 for his work with Brownian motions. It was soon found that a stochastic process satisfying the requirements

---

[34] Most of the mathematics of this century is somehow related to one or more of the 23 problems of Hilbert.

[35] Actually, it can be shown that essentially all stochastic processes can be defined on the unit interval, but the proof requires modern measure theory.

[36] Today the Brownian motion is widely used in economics, biology, ecology etc., as a model for practical everything except original Brownian motions.

of Einstein would have continuous sample paths, which was to be expected, but it was also shown that the sample paths would have to be nowhere differentiable, which was shocking as it meant that a Brownian particle travels with infinite speed. This fact was so discomforting that probabilists began to doubt that there existed a consistent theory covering the model of Einstein. However, the eminent mathematician Norbert Wiener set out to prove the existence of Einstein's model for Brownian motions, and after a "tour de force" he finally, succeeded to do so at the end of the 1920's.

The method of Norbert Wiener was tailor-made for Brownian motions and could hardly be used for proving the existence of other complicated processes such as stable processes or the Ohrenstein-Uhlenbeck process. But in 1933 the ingenious mathematician Andrei Nikolaevich Kolmogorov (1903–1989) published *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Foundations of Probability Theory), which was to set probability on its modern course. In *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Kolmogorov applied abstract measure theory to provide a new foundations for probability theory: A *probability space* is defined to be a triple $(\Omega, \mathcal{F}, P)$ satisfying:

(Pr.1) Ω is a non-empty set, $\mathcal{F}$ is a non-empty set of subsets of Ω, and P is a function from $\mathcal{F}$ into the non-negative real numbers $\mathbf{R}_+$

(Pr.2) $$F, F_1, F_2, \ldots \in \mathcal{F} \quad \Rightarrow \quad F^c \in \mathcal{F} \quad \text{and} \quad \bigcup_{n=1}^{\infty} F_n \in \mathcal{F}$$

(Pr.3) $$P(\emptyset) = 0 \quad \text{and} \quad P(\Omega) = 1$$

(Pr.4) If $F_1, F_2, \ldots \in \mathcal{F}$ are disjoint, then $P\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} P(F_n)$

A non-empty collection $\mathcal{F}$ of sets satisfying (Pr.2) is called a *σ-algebra*, and a non-negative function P satisfying (Pr.3) and (Pr.4) is called a *probability measure* [see (1.1)]. With this notion at hand we define a *random variable* to be a function from Ω into the real line $\mathbf{R}$ such that the set $\{\omega \mid X(\omega) \leq a\}$ belong to $\mathcal{F}$ for every $a \in \mathbf{R}$ [see (2.7)]. In *Grundbegriffe der Wahrscheinlichkeitsrechnung* Kolmogorov proves his famous *consistency theorem* which, loosely speaking, states that every random phenomenon can be modeled within the axioms (Pr.1)–(Pr.4) unless the specifications contains evident inconsistencies [see Section (9.4) of Volume II, where you will find the precise statement and proof]. The consistency theorem solves in one stroke all the problems of existence of rigorous models for random systems such as the Brownian motion, the Poisson process, the stable processes, the Ohrenstein-Uhlenbeck process, etc., and the book was received by the contemporary probabilists with almost euphoria. Suddenly, all the mess of the house of probability was cleaned up and probability could enter mathematics as a rigorous and firmly founded discipline. Since 1933 the vast majority of probabilists have been working within the framework of Kolmogorov. However, a few years before (1931), von Mises found another rigorous and firm foundation of probability theory, taking frequencies and the law of large numbers as the basic object, but the work of von Mises was completely overshadowed by Kolmogorov's model, and only a few – very few[37] – probabilists have worked within the framework of von Mises.

Kolmogorov also consider the weaker notion of a *finitely additive probability space*, which is defined to be a triple $(\Omega, \mathcal{F}, P)$ satisfying:

(Fpr.1) Ω is a non-empty set, $\mathcal{F}$ is a non-empty set of subsets of Ω, and P is a function from $\mathcal{F}$ into the non-negative real numbers $\mathbf{R}_+$

(Fpr.2) $F, G \in \mathcal{F} \quad \Rightarrow \quad F^c \in \mathcal{F}$ and $F \cup G \in \mathcal{F}$

(Fpr.3) $P(\emptyset) = 0$ and $P(\Omega) = 1$

(Fpr.4) If $F, G \in \mathcal{F}$ are disjoint, then $P(F \cup G) = P(F) + P(G)$

A non-empty collection $\mathcal{F}$ of sets satisfying (Fpr.2) is called an *algebra*, and a non-negative function P satisfying (Fpr.3) and (Fpr.4) is called a *probability content* [see (1.1)]. The notion has had little impact on the course of probability, and today, the study of *pure probability contents* [that is, a probability contents, which doesn't satisfy (Pr.4)] is closer to logic and the foundation of set theory than to probability theory. For instance, if $\mathcal{F}$ is a σ-algebra, then it can be shown that it is not possible to ***construct*** a pure probability content, and that the existence of a pure probability content is equivalent to a certain form of the axiom of choice.

What Is a Probability?

After this brief account of the history of probabilities it is appropriate to ask: "What is a probability really? And what does history teaches us about probabilities?" The history demonstrates that the notions of randomness and probabilities are difficult – very difficult – to apprehend, and that the notions are fairly new and alien to humans.[38] Second, the history and a wealth of examples, many of which can be found in this text, demonstrate that our intuition about probability is poor and often takes a wrong track. Thus, when you evaluate probabilities trust your computations and doubt your intuition. For this reason the interpretation of probabilities presents

---

[37] For instance, Kolmogorov in the 1970's applied the ideas of von Mises to define independence of the digits in a given sequence of 0's and 1's.

[38] I don't think that mankind ever was meant to apprehend randomness and probabilities.

a difficult and fundamental problem, and the past – from the emergence of probabilities in 1550 until today – contains numerous examples of misinterpretations and miscalculations. There are three (at least) ways of interpreting probabilities:

A priori probabilities: If someone shows a die and claims that the six possible outcomes have probability $\frac{1}{6}$, you would certainly agree without hesitation, unless you know the particular die and have observed some kind of skewness. Probabilities assigned in this way are called *a priori probabilities*, that is, probabilities that all sensible persons can agree on. A priori probabilities usually stem from an assumption of uniformity, that is, the assumption that certain outcomes are equally probable, and they are often based on the uniform distributions [see (2.15) and (2.24)]. For instance, the classical model of Fermat and Pascal is a typical example of an a priori assignment of probabilities. In statistics the idea is closely related to the so-called *Bayesian school of statisticians*. In Chapters 1 and 2, you will see how to model a priori probabilities within the axioms of Kolmogorov. The main disadvantage of the notion are the many cases where there is no consensus about the probability assignment, and the less numerous cases where the ostensible obvious probability assignment disagrees with observations; see for instance, the anomal distribution described in Section (4.40).

Frequential probabilities: The most common interpretation of a probability is that if you repeat the event independently a certain number of times, then the probability approximately equals the frequency of the occurrence of the event. Such an interpretation is called a *frequential probability*. For instance, claiming that the probability of getting one head in one toss with a coin is $\frac{1}{2}$ is usually interpreted as follows: if you throw the coin 100 times, then you will get approximately 50 heads and 50 tails. According the recordings of one of largest Danish fire insurance companies, the probability that my house will burn down completely within the next year is $\frac{1}{20{,}000}$. In this case, the next year can, of course, not be repeated, but in the files of the insurance company we find records of fires in houses of the same kind and age as my house, which can be used to find a frequency and, thus, a frequential probability. It is important to observe that a frequential probability is a collective notion, which says little or nothing about the individual. For instance, if a doctor tells a cancer patient that he has a 60% chance of surviving the next year, then the statement has little meaning to the patient – either he survives or he don't – but to the doctor the statement has perfectly valid meaning: Of a group of say 100 patients with the same kind cancer the doctor expects 60 to survive and 40 to die within the next year. The model of Kolmogorov gives the basic rules of the calculus of probability, but it isn't a model of frequential probabilities and it doesn't say anything about the probability of a given event; nevertheless, the frequential interpretation of probabilities is the main link between reality and probability theory. Thus, within the framework of Kolmogorov we have to ***prove*** the validity of the frequential interpretation, that is, to prove the law of large numbers.[39] This will be done in Sections (4.8)–(4.12) and (4.32). The main advantage of a frequential probability is its robustness; that is, new information is unlikely to alter the probability in any significant way (provided the probability is based on many reliable observations). The main disadvantage is that doesn't applies to individual events and requires that the event can be repeated independently many times.

Subjective probabilities: At present the Danish State is building the longest bridge in the world over the highly trafficked waters between the islands of Funen and Zealand. In the summer of 1991, I was asked by a member of the Danish Parliament to give my opinion of the evaluation of the probability that the bridge would by destroyed by a collision with a ship within the next 100 years, a projection made by a consulting firm. The firm had computed the probability to be 5%.[40] In this case the probability is neither an a priori probability – I doubt that anyone would be able to assign a probability without thorough investigations – nor is it a frequential probability – the bridge is not yet finished, and is only built to last 100 years; moreover, there are no comparable bridges in the world.[41] It is a so-called *subjective probability*. A subjective probability is interpreted as the degree of belief in the person stating the probability. Many daily-life probabilities are subjective; for instance, the bookmakers odds are of this type, and often the exchange rates on the currency market are of this type.[42] Kolmogorov's model is in reality a model of subjective probabilities: The axioms don't tell what the probabilities are, but when you have fixed your subjective probabilities they tell you how to calculate other probabilities. The main advantage of a subjective probability is its broadness of application, and its major disadvantages is its subjectivity – you may disagree with practically everybody about the probability assignment – and its sensibility to new information. Before the Chernobyl accident the probability of a serious accident on a nuclear power plant was (subjectively) considered as almost negligible, but after the accident the probability was drastically increased.

References to the History of Probability

The history of probability and statistics serves as entrance for a better understanding of the subjects. There is a number of books covering the history of probability up to about 1820 (for instance, [2], [3], [4] and [5]) but only few

---

[39] In many ways it would be more natural to follow the ideas of von Mises, and take the law of large numbers as an axiom and not as a theorem to be proved.

[40] This is a very high risk probability; for instance, the probability that my house will burn down completely within the next 100 years is 0.5%.

[41] The consulting firm derived the probability by simulations on a computer.

[42] Often the exchange rates on the currency market doesn't reflect reality but the speculators' expectation of the future market.

books covering the history of statistics (for instance, [3] and [6]) or the history of probability after 1820 (for instance, [1]).

[1] Adams, J. (1974): *The Life and Times of the Central Limit Theorem*, Kaedman Publishing Co., New York

[2] David, F. N. (1962): *Games, Gods and Gambling*, Charles Griffin & Co., London

[3] Hald, A. (1990): *A History of Probability and Statistics and Their Applications before 1750*, John Wiley & Sons, New York

[4] Maistrov, E. (1974): *Probability – a Historical Sketch*, Academic Press, New York and London

[5] Ore, O. (1953): *Cardano – the Gambling Scholar*, Princeton University Press, Princeton

[6] Westergaard, H. (1932): *Contribution to the History of Statistics*, P. S. King & Son Ltd., London