SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

The emergence of probability theory

av

Jimmy Calén

2015 - No 5

The emergence of probability theory

Jimmy Calén

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Paul Vaderlind

2015

Abstract

This bachelor thesis will examine how, and for what reason, some of the fundamental probabilistic concepts emerged. The main focus will be on the transition from empiricism to science during the 17^{th} and 18^{th} century.

Acknowledgements

I would first like to express my gratitude to my supervisor Paul Vaderlind for suggesting a topic in consummate accordance with my main mathematical interest, and also for his guidance during the writing process.

Special thanks go to my loved ones, for their support and company during my studies.

Contents

1 Introduction

*If you are in a room with 22 others, what is the probability that at least 2 of you have the same birthday?

*If an ape hits a keyboard 5 times on random – what is the chance that it succeeds to write a real word found in the English dictionary?

*A group of n mathematicians enter a restaurant and check their hats. The hat-checker is muddle-headed, and upon leaving, she redistributes the hats back to the mathematicians at random. What is the probability that no man gets his correct hat?

The three straight-on questions above can all be usefully answered with knowledge of probability theory.

TV-shows where numbered balls are picked on random and a hopeful attender prays for her exact sequence to show up. Gambling in casinos where people desperately try to come up with an unbeatable strategy. Stock brokers on the Wall Street who try risk-minimizing their portfolio. Public fears that are endlessly debated in terms of probability – meltdowns, earthquakes, muggings.

Probability theory answers daily questions as well as it with huge relevance contributes to complex mathematical models in everything from financial economics, risk analysis, betting, and abstract models in physics, to in tradition non-mathematical fields like psychology and sociology.

2 Philosophical exposition of the probability concept

Like all mathematical concepts, the probabilistic one has occurred gradually. The relation between probability and randomness (lack of pattern or predictability in events) was in many cases not known. As late as in the 19th century, both the conceptual content and the severity of the mathematical method changed significantly. As our modern perception of mathematics and its relationship to reality has evolved – a considerable rigor has emerged. Due to this, every branch of mathematics must be developed with axioms as a basis. The theory should therefore be disengaged from its applications, and only consist of logical theories free from contradictions. Since probability theory, often has had its base in what we today consider the applications – adversities of becoming a tantamount part of the mathematical field occurred. It was rather considered to be in the suspect borderland of mathematics and philosophy – a role that it held until year 1933 when the Soviet mathematician Andrey Kolmogorov laid the modern axiomatic foundations of probability theory and therefore paved the way for an uncompromised acceptance of the field. These clarifications of the probabilistic fundaments have led to a leap-like progress with an extensive number of applications. [9]

Even though the axiomatic revolution of the theory opened up for an acceptance, many would still not want to accept that their specific field could not be perfectly examined using 100% risk free, deterministic methods. In the field of quantum mechanics, Albert Einstein – by many considered one of the best physicists of all time – refused to accept the probabilistic impact. For example, finding an electron in a particular region around the nucleus at a particular time can impossibly be done with 100% surety. You can only find the probability of that occurrence.

In 1943, Einstein said the following in a conversation with William Hermanns:

"As I have said so many times, God doesn't play dice with the world." [10]

3 Games of chance in early times

Even though there was no probabilistic theory to lean back on, games of chance have in thousands of years fascinated people. Groups of people from diverse areas of the world have since thousands of years ago, independent and without knowledge of each other invented games of chance. Several Babylonian gaming boards dating back to 2700 BC have been found complete with playing pieces and signs of that the game must have been driven with some kind of chance mechanism. There are also evidences of games of chance being played by ancient Egyptians and Chinese, dating back to 2100 BC. [11]

Excavations in Egypt show that six-sided dice have been used at least since 1320 BC, but other chance-devices have turned up in even earlier Egyptian sites. Similar to excavations in Egypt, archeologists in the ancient Greek and Roman world consistently digs up a disproportionate big amount of “astragali”, the knucklebones of sheep and other vertebrates. The astragalus, colloquially speaking called “talus”, has six asymmetrical sides, but when thrown to land on a level surface, it can only come to rest in four ways. The ones found are often well polished and engraved, which undoubtedly contributes to the now plausible hypothesis that they were part of games. [12]

Greek and Roman games of chance used four astragali in a simple “rolling of the bones”. With background in a study of the writings of the classical time the, in generally agreed, scoring was as follows.

4 points: The upper side of the bone, broad and slightly convex.
3 points: The opposite side, broad and slightly concave.
1 point: The lateral side, flat and narrow.
6 points: The opposite narrow lateral side, slightly hollow.

[22]

The empirical probabilities based on the tossing of a modern astragalus of a sheep are approximately 1/10 each of throwing a 1 or 6, and about 4/10 each of throwing a 3 or 4. With these approximated probabilities given, the probabilities associated with the four thrown astragali would have been as shown in the table below. These probabilities are obviously calculated using theory not yet invented when the game was played thousands of years ago.

				10000*probability
(1^4)	(6^4)			1
(3^4)	(4^4)			256
($1^3 3$)	($1^3 4$)	($6^3 3$)	($6^3 4$)	16
($3^3 1$)	($3^3 6$)	($4^3 1$)	($4^3 6$)	256
($1^3 6$)	($6^3 1$)			4
($3^3 4$)	($4^3 3$)			1024
($1^2 3^2$)	($1^2 4^2$)	($6^2 3^2$)	($6^2 4^2$)	96
($3^2 4^2$)				1536
($1^2 6^2$)				6
($1^2 34$)	($6^2 34$)			192
($3^2 16$)	($4^2 16$)			192
($3^2 14$)	($3^2 64$)	($4^2 13$)	($4^2 63$)	768
($1^2 63$)	($1^2 64$)	($6^2 13$)	($6^2 14$)	48
(1634)				384

Above, (1^4) means four astragali numbered one, ($4^2 13$) means two astragali numbered 4 and one astragalus each numbered 1 and 3. The best throw was (1634) – called the venus. Other outcomes of the games have a smaller probability, which indicates that the Greeks and Romans playing the game had not taken notice of the magnitudes of the corresponding relative frequencies. [13]

The lack of mathematical knowledge and drawing conclusions from statistics is of course the reason to why the probabilities above could not be stated. Researchers' today claim that there might have been religious reasons to the prevention of scientific studies of games of chance. Namely, the belief of that every event was predetermined by God – and therefore – there was no such thing as chance. [2]

4 The faltering steps before the 17^{th} centuries breakthrough

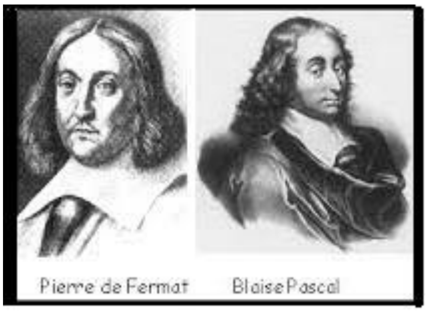
In spite of fact that the probability theory known today shows itself in so many different shapes, the development was initially slow. In the 17^{th} century, respectable discoveries in algebra, geometry and trigonometry had been general mathematical knowledge for centuries. For example, the well-known book *Algebra* written by the Persian mathematician Muhammad ibn Musa al-Khwarizmi was written 830 A.D, more than 600 years before the theory of probability was created.

The main breakthrough of the underlying theory - and therefore the actual start of what we today entitle *probability theory* – was in the 17^{th} century. Before then, the field of probability was with few exceptions completely nonmathematical. The Italian mathematician Girolamo Cardano wrote a book on games of chance around 1565, which was not published before 1663. He managed to compute probabilities of simple events, like a number of dice showing a given number. On the other hand, the book contains deficient solutions derived through trial and error together with many unclear, confusing statements regarding various games. Cardano also takes part in one of the first attempts of solving the "problem of points" (which will be examined in chapter 5). There are in his reasoning traces of probabilistic arguments which should not be regarded, but he never arrives to a correct conclusion. He can due to these reasons not be consider the founder of the theory – but instead one of the first who tried getting an understanding. [2]

"Even if gambling were altogether an evil, still,
on account of the very large number of people who play,
it would seem to be a natural evil" [14]

– Girolamo Cardano

5 Blaise Pascal & Pierre de Fermat – The 1654 years foundation of probability theory



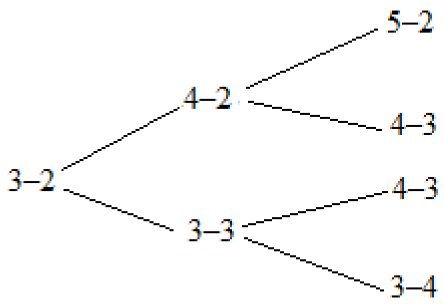
[23, 24]

In year 1654, Blaise Pascal (19 June 1623 in Clermont-Ferrand, France – 19 August 1662 in Paris, France) was asked some questions on games of chance. He communicated his solutions to Pierre de Fermat (17 August 1601 or 1607 in Beaumont-de-Lomagne, France – 12 January 1665 Castres, France) for approval. A correspondence between the two of them ensued, in which the foundations of the theory of probability were laid. Taking the addition and multiplication rules for granted, they introduced what we today entitle the expected value by means of the “problem of points”, also called the “problem of division of the stakes”. Pascal also introduced recursion as a method for solving probability problems, and they together discussed the problem of Gambler´s ruin. The famous correspondence consists of seven letters between July and October 1654. During the same period of time Pascal wrote the important treatise on what is today known as the arithmetical triangle or Pascals triangle. [2]

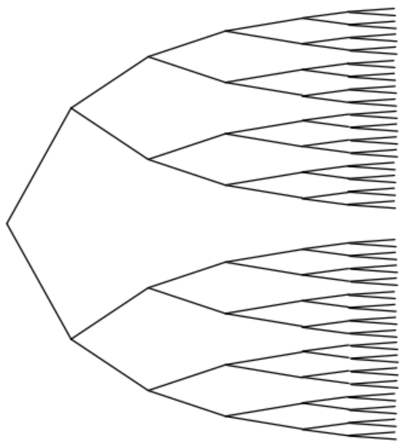
5.1 The problem of points

The problem of points concerns a game of chance with two players with equal chances of winning each round, let us assume tossing coin. The players contribute the same amount of money to a prize pot, and agree that the first player to win a certain number of rounds will collect the entire prize. Now, if the game is interrupted by external circumstances and unlikely resumed anytime soon, before either player has achieved victory – how will the pot be divided fairly among the two? The starting insight for Pascal and Fermat was realizing that the division should primarily depend on the possible ways the game might have continued if not interrupted. Therefore, it is clear that a player with a 8–2 lead in a game to 10 has the same chance of winning as a player with a 98–92 lead in a game to 100, even though the intuition says that the later game should be of more even character. We now look at an example for the case of a 3–2 lead, in a game with 7 rounds. In case the game is played out before the 7th round – namely if one player reaching 4 wins before – the players will finish the 7 rounds anyway.

Fermat computed the odds for each player to win by writing down all possible outcomes. [2]



The illustration above shows the different number of outcomes in terms of “number of wins”. Out of a “3–2”–situation, the one having the lead will achieve victory in 3 out of 4 cases. Therefore, he should be divided 3/4 of the total pot. If extending the problem, and instead looking at the “1–0”–situation, it immediately gets a lot more time consuming to use the tree-method as in the “3–2”–situation. We would then have the following illustration:



[15]

Since every round can have two different outcomes, player A winning or player B winning, we have 2^6 possible combinations of the 6 remaining rounds. We now introduce the following notation, focusing on how many more wins needed to achieve victory, instead of the number of historic wins: If the game is interrupted and A lacks *a*, and B lacks *b* games in winning, we let e(*a, b)* denote A’s share of the total pot. By the same reasoning, in the general case these remaining $a + b - 1$ rounds have 2^{a+b-1} possible outcomes. The cases favorable to A in relation to 2^{a+b-1} will then give A’s fraction of the total stake.

In a letter from Pascal to Fermat dated August 24, year 1654 – Pascal writes:

"Your method is very safe and is the one which first
came to my mind in this research: but because the trouble
of the combinations is excessive I have found an abridgement
and indeed another method that is much shorter and more neat,
which I would like to tell you here in a few words…" [1]

In the further correspondence Fermat receives the main ideas and a complete table for a stake of 512, and the cases $b = 6$, $a = 1,2,..., 6$ and a stake of 512. [2]

Pascal's procedure may be presented in the following way: We assume a (a, b)-situation. It can be followed by either $(a - 1, b)$ or $(a, b - 1)$, both equally likely.

Hence:

$$e(0,n)=1 \qquad \text{and} \qquad e(n,n)=\frac{1}{2}, \qquad n=1,2,...,$$

$$e(a,b)=\frac{1}{2}\left[e(a-1,b)+e(a,b-1)\right], \qquad (a,b)=1,2,....$$

We now examine a few base cases.

$$e(2,1)=\frac{1}{2}\left[e(2-1,1)+e(2,1-1)\right]=\frac{1}{2}\left[e(1,1)+e(2,0)\right]=\frac{1}{4}.$$
$$e(3,1)=\frac{1}{2}\left[e(3-1,1)+e(3,1-1)\right]=\frac{1}{2}\left[e(2,1)+e(3,0)\right]=\frac{1}{8}.$$
$$e(3,2)=\frac{1}{2}\left[e(3-1,2)+e(3,2-1)\right]=\frac{1}{2}\left[e(2,2)+e(3,1)\right]=\frac{5}{16}.$$
$$e(4,1)=\frac{1}{2}\left[e(4-1,1)+e(4,1-1)\right]=\frac{1}{2}\left[e(3,1)+e(4,0)\right]=\frac{1}{16}.$$
$$e(4,2)=\frac{1}{2}\left[e(4-1,2)+e(4,2-1)\right]=\frac{1}{2}\left[e(3,2)+e(4,1)\right]=\frac{6}{32}.$$

Pascal compared the terms yielded from the difference equation with the ones in his arithmetical triangle (which will be examined in chapter 9 as well), concluding that they differ only by a factor of 1/2 and by the boundary conditions. [2]

First, consider the triangle for the first 6 rows:

$$
\begin{array}{ccccccccccc}
 & & & & & \binom{0}{0} & & & & & \\
 & & & & \binom{1}{0} & & \binom{1}{1} & & & & \\
 & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & \\
 & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & \\
 & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & \\
\binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5}
\end{array}
\qquad
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

One may if being observant see that all the denominators in the presented base cases correspond to the sum of all elements in a specific row of the triangle. The numerators all seem to be the sum of the b first elements in that row. Further, the following pattern seems to emerge: Considering e(a, b). If $a + b = k$, we look at the k^{th} row in the above presented triangle. For e(4, 2) we get:

$$\frac{1+5}{1+5+10+10+5+1} = \frac{6}{32},$$

which is obviously correct.

In general, this may be stated by the following expression:

$$e(a,b) = \sum_{i=0}^{b-1} \binom{k-1}{i} \left(\frac{1}{2}\right)^{k-1} = \sum_{i=0}^{b-1} \binom{a+b-1}{i} \left(\frac{1}{2}\right)^{a+b-1}.$$

Pascal uses the method of induction to prove this formula. Since it holds for $a + b = 2$, he assumes that it holds for $a + b = k$ and then proves that it as well holds for $a + b = k + 1$. Both Pascal and Fermat solved the problem of points using combinatorial methods. [2] Pascal may especially be credited introducing what is today known as the "expected value" when stating $e(a,b) = \frac{1}{2}\left[e(a-1,b) + e(a,b-1)\right]$.

This would today be written as

$$E\left[X\right] = \frac{E\left[X \mid A\right] + E\left[X \mid B\right]}{2},$$

X: A's share of the total stake.

A: A wins next round.

B: B wins next round.

6 The giant leap forward

After the considered 1654 years start of probability theory – with Blaise Pascal and Pierre de Fermat as prominent figures – the development of the probabilistic field stagnated for nearly half a century. In 1708, a benchmark of a period of hectic activity with plenty of weighty publications was set. The development took off after the French mathematician Pierre Rémond de Montmort (27 October 1678 in Paris, France – 7 October 1719 in Paris, France) published his work. The list below presents the works of several, for the period and field, prominent mathematicians.

- 1708. Pierre Rémond de Montmort, *Essay d'Analyse sur les Jeux de Hazard*. Paris
- 1709. Nicolaus I Bernoulli, *De Usu Artis Conjectandi in Jure*. Basel.
- 1712. John Arbuthnott, *An Argument for Divine Province, taken from the constant Regularity observed in the Births of both Sexes. Phil. Trans.* London.
- 1712. Abraham de Moivre, *De Mensura Sortis. Phil. Trans.* London.
- 1713. Jacob (also known as James or Jacques) Bernoulli, *Ars Conjectandi*. Basel. (Published 8 years after his death).
- 1713. Pierre Rémond de Montmort, *Essay*. 2nd Edition. Paris.
- 1716. Nicolaas Struyck, *Uytreekening der Kanssen in het spelen, door de Arithmetica en Algebra, beneevens eene Verhandeling van Looterijen en Interest*. Amsterdam.
- 1717. Nicolaus I Bernoulli, *Solutio Generalis Problematis XV propositi à D. de Moivre, in tractatu de Mensura Sortis. Phil. Trans.* London.
- 1717. Abraham de Moivre, *Solutio generalis altera præcedentis Problematis. Phil. Trans.* London.
- 1718. Abraham de Moivre, *The Doctrine of Chances*. 1st edition. London.

The works of especially the today well-known mathematicians de Montmort, de Moivre, & the Bernoulli family contained many excellent ideas, methods, and problem. Their works had such an impact on the field of probability theory, so that it took nearly a century to digest – and thereafter develop the presented material. A few decades after the presentations of the works in the list above – year 1738, de Moivre's 2nd version of "*The Doctrine of Chances"* was published. The doctrine was by many accredited the role of something that can be likened to a Bible of probability theory, a role that it held until 1812 when it was superseded by the work of Pierre Simon de Laplace (23 Mars 1749 in Calvados, France – 5 Mars 1827 in Paris, France) – *Théorie Analytique des Probabilités.* [2]

7 Pierre Remond de Montmort & his Essay d'Analyse sur les Jeux.

"In 1708 he [de Montmort] *published his work*
on chances, where with the courage of Columbus
he revealed a new world to mathematicians." [16]

– Isaac Todhunter

Many claim that the two editions of de Montmorts *Essay* never came to enjoy the popularity that a respected probabilistic work would have deserved. Unlike the works of the Bernoulli family and de Moivre, *Essay* had deficient structure. The second edition – an expanded version of the first, with additions and generalizations – mixes the structure of a textbook and a scientific paper. Problems were stated in the text, and the associated, expanded theory often spread out in many places in the book. The Bernoulli family and de Moivre were – as if the pedagogical problems were not enough – well known mathematicians while de Montmort was considered an amateur by many. On the other hand, de Montmort inspired and stimulated curiosity in both the Bernoulli family and de Moivre. De Montmort developed the work of Fermat, solving problems of chance with means of the mathematical field combinatorics. At a time when card games gained popularity, a mathematical perspective of the theory of the games was needed. In the 1708 years edition of de Montmorts *Essay,* the first discussions of probabilistic art mentioning the problems of coincidences and derangements were presented. [2, 13]


[28]

A symbolic insight into life around the
gambling table, from Montmorts *Essay*.

De Montmort exemplified his theory in connection with the card game *le Jeu du Treize*. [3]
In the 1713 years edition of *Essay*, he presented the rules as follows (translated from French):

The players draw to see who will be the dealer. Let's call the dealer 'Pierre', and let's suppose that there are as many other players as you like. Pierre takes a full deck of 52 cards, shuffles them, and deals them out one after the other, calling out 'one' as he turns over the first card, 'two' as he turns over the second, 'three' as he turns over the third, and so on up to the thirteenth, which is a king. ['one', 'two', 'three', 'four', 'five', 'six', 'seven', 'eight', 'nein', 'ten', 'eleven','twelve', 'thirteen'.] Now if, in this whole series of cards, he never once turns over the card he is naming, he pays out what each other player has put up for the game, and the deal passes to the player sitting to his right. But if in this sequence of thirteen cards he happens to turn over the card he is naming, for example, if he turns over an ace as he calls out 'one', or a two as he calls out 'two', or a three as he calls out 'three', etc., he collects all the money that is in play, and begins over as before, calling out 'one', and then 'two', etc. [5]

Supposing that the dealer has *n* different cards sorted randomly. What is the probability of at least one coincidence as the dealer turns over the cards?

De Montmort says that due to the difficulty of finding the dealer´s advantage, he simplifies the problem to include only one suit of 13 cards. He then examines the problem for $n = 2,\ldots,5$, in each case following the same principle. [2]

He arguments in a way that is best explained using an example.

7.1 Practical approach to derangements

First, consider the following notation for the probability of at least one coincidence: P_n.

The base cases for $n = 0,1$ yield $P_0 = 0,$ and $P_1 = 1.$ For $n = 2$ there can only be two outcomes: {1} is in the first place and therefore, {2} is in the second place. Or the opposite, they are both in the wrong place. $\rightarrow P_2 = 1/2.$

The section below examines the somewhat more complicated cases for $n = 3, 4, 5$.

Starting with the case for $n = 3$ which corresponds to using cards labeled 1 to 3.

{1}, {2}, {3}.

The three of these can be rearranged, permuted in $3! = 6$ ways, namely:

{1}, {2}, {3}　　{1}, {3}, {2}　　{2}, {1}, {3}

{2}, {3}, {1}　　{3}, {1}, {2}　　{3}, {2}, {1}

Among these, there are:

2 ways when {1} is in the first place.
1 way when {2} is in the second place without {1} being in the first.
1 way when {3} is in the third place without {1} being first or {2} being second.

The probability of at least one coincidence will therefore be $P_3 = \frac{2+1+1}{6} = \frac{2}{3} = 0.66666...$

Continuing in the same manner for $n = 4$ the $4! = 24$ possible permutations will be:

{1}, {2}, {3}, {4}　　{1}, {2}, {4}, {3}　　{1}, {3}, {2}, {4}

{1}, {3}, {4}, {2}　　{1}, {4}, {2}, {3}　　{1}, {4}, {3}, {2}

{2}, {1}, {3}, {4}　　{2}, {1}, {4}, {3}　　{2}, {3}, {1}, {4}

{2}, {3}, {4}, {1}　　{2}, {4}, {1}, {3}　　{2}, {4}, {3}, {1}

{3}, {1}, {2}, {4}　　{3}, {1}, {4}, {2}　　{3}, {2}, {1}, {4}

{3}, {2}, {4}, {1}　　{3}, {4}, {1}, {2}　　{3}, {4}, {2}, {1}

{4}, {1}, {2}, {3}　　{4}, {1}, {3}, {2}　　{4}, {2}, {1}, {3}

{4}, {2}, {3}, {1}　　{4}, {3}, {1}, {2}　　{4}, {3}, {2}, {1}

Among these, there are:

- 6 ways when {1} is in the first place.
- 4 ways when {2} is in the second place without {1} being first.
- 3 ways when {3} is in the third place without {1} being first or {2} being second.
- 2 ways when {4} is in the fourth place with {1}, {2}, {3} being out of their places.

The probability of at least one coincidence will therefore be $P_4 = \frac{6+4+3+2}{24} = \frac{5}{8} = 0.625$

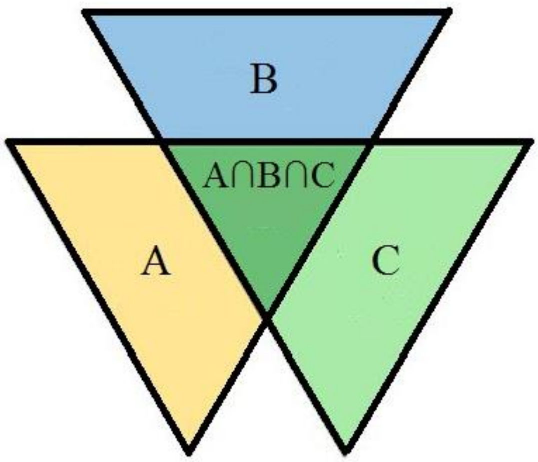
By following the same principle one may state the following for the case when n = 5.
Out of the 5! = 120 permutations, there are:

- 24 ways when {1} is in the first place.
- 18 ways when {2} is in the second place without {1} being first.
- 14 ways when {3} is in the third place without {1} being first or {2} being second.
- 11 ways when {4} is in the fourth place with {1}, {2}, {3} being out of their places.
- 9 ways when {5} is in the fifth place with {1}, {2}, {3}, {4} being out of their places.

The probability of at least one coincidence will therefore be:

$$\frac{24+18+14+11+9}{120} = \frac{19}{30} = 0.63333...$$

Using this technique to approach the problem, de Montmort avoided adding the "intersection" more than once. The picture below illustrates for the case when n = 3.



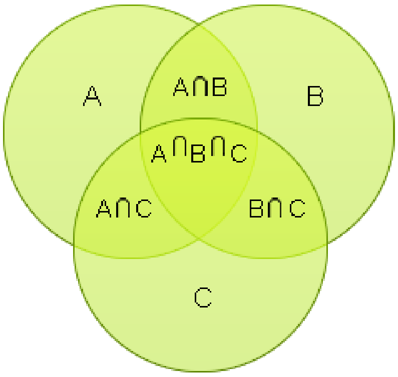
Let A denote the case when {1} is in its correct place.
Let B denote the case when {2} is in its correct place.
Let C denote the case when {3} is in its correct place.

$A \cap B \cap C$ denotes the intersection: When they are all in the correct place.

Characteristic for the field of derangements is the following for the case when n = 3.

$$A \cap B = A \cap B \cap C,$$
$$A \cap C = A \cap B \cap C,$$
$$B \cap C = A \cap B \cap C.$$

This characteristic is derived from the fact that if we know that n – 1 out of n elements are in the correct places – the n:th element will as well be placed correctly. If this would not be the case, as common in other fields of mathematics, we would instead have the following illustration:



7.2 Theorems of coincidences

De Montmort remarks that giving the general proof will take up too much space and then states the general solution in two forms: as a recursion formula, and an explicit solution as a series. [2]

De Montmorts recursion formula may be presented as:

$$P_n = \frac{(n-1)P_{n-1} + P_{n-2}}{n}, \qquad n \geq 2, \qquad P_0 = 0, \quad \text{and} \qquad P_1 = 1 \qquad (1)$$

This yields:

$$P_{13} = \frac{109339663}{172972800} = 0.632120558.$$

He, as mentioned, also states the solution in terms of an alternating series, which by means of the numbers in the arithmetic triangle yields:

$$P_n = 1 - \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{(-1)^{n-1}}{n!}, \qquad n \geq 1 \qquad (2)$$

Since

$$\lim_{n\to\infty}\sum_{k=0}^{n}\frac{x^k}{k!}=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+...=\mathrm{e}^x \qquad (3)$$

de Montmort could prove

$$\lim_{n\to 0}P_n=1-e^{-1}=0.632120558.$$

The relation between the two formulas (1) and (2) is not examined by de Montmort. Presumably though, he was probably aware that (1) may be written as

$$P_n-P_{n-1}=-\frac{P_{n-1}-P_{n-2}}{n},$$

and that (2) gives

$$P_n-P_{n-1}=\frac{(-1)^{n-1}}{n!},$$

from which he would easily have derived the relation between them. [2]

7.3 Proof:

In *Essay* de Montmort includes several of examples from which his proof could be found. [2]

Let $d_n(i)$ denote the set of permutations with the first coincidence at place *i*. Note that $d_n(1)=(n-1)!$ since {1} is fixed first. The remaining numbers may then be permuted in $(n-1)!$ ways. For $d_n(2)$ we fix {2} at the second place and the remaining numbers may again be permuted in $(n-1)!$ ways. Since {2} must be the first coincidence, we must deduct the number of permutations of both {1} and {2} being fixed, which is $(n-2)!$.

$$d_n(2)=(n-1)!-(n-2)!=d_n(1)-d_{n-1}(1).$$

In the general case, with the same method as above, we have:

$$d_n(i+1) = d_n(i) - d_{n-1}(i), \qquad n \geq 2, \qquad i = 1, 2, \ldots, n-1.$$

Using this recursion, de Montmort states the following table for the cases when n = 1, 2,…,8.

Let $d_n = d_n(1) + d_n(2) + \ldots + d_n(n)$.

	i								
n	1	2	3	4	5	6	7	8	d_n
1	1								1
2	1	0							1
3	2	1	1						4
4	6	4	3	2					15
5	24	18	14	11	9				76
6	120	96	78	64	53	44			455
7	720	600	504	426	362	309	265		3186
8	5040	4320	3720	3216	2790	2428	2119	1854	25487

Since n numbers can be arranged in $n!$ ways, and d_n represents the number of events with at least one coincidence, we get:

$$P_n = \frac{d_n}{n!}.$$

When setting $c_0(n) = n! - d_n$, de Montmort could give the following recursion formula

$$d_n = nd_{n-1} + (-1)^{n-1}, \qquad n \geq 1, \qquad d_0 = 0.$$

For given n, the probability that the first coincidence appears at the i:th place equals:

$$\frac{d_n(i)}{n!} = \frac{d_n(i)}{d_{n+1}(1)}.$$

He then stated the quite obvious:

$$d_n(n) = c_{n-1}(0) = (n-1)! - \mathrm{d}_{n-1}.$$

De Montmort finally arrives to the distribution of the number of coincidences by realizing the following: Out of a sequence of *n* numbers, the *k* coincidences may be chosen in $\binom{n}{k}$ ways, and there must be no coincidence at the remaining *n* – *k* places. [2]

This yields:

$$c_n(k) = \binom{n}{k} c_{n-k}(0).$$

In conclusion we have that

$$p_n(0) = 1 - P_n = \sum_{i=0}^{n} \frac{(-1)^i}{i!},$$

and in general

$$p_n(k) = \frac{p_{n-k}(0)}{k!} = \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}.$$

7.4 Conclusions of the theorems

Returning to the card game *Treize* that first interested de Montmort – and gave fire to his great advance in the theory of derangements/coincidences, the following should now be known:

k	0	1	2	3	4	5	…	13
$p_{13}(k)$	0.368	0.368	0.184	0.061	0.015	0.003	…	1/13!

Pierre-Rémond de Montmorts work occurs frequently in textbooks of probability and combinatorics today – and we may successfully by means of it answer the kind of questions asked in the introduction (by the author) of this composition.

- A group of 7 mathematicians enter a restaurant and check their hats. The hat-checker is muddle-headed, and upon leaving, she redistributes the hats back to the mathematicians at random. What is the probability that no man gets his correct hat?

$$P_7 = \sum_{i=0}^{7} \frac{(-1)^i}{i!} = 1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} + \frac{1}{720} - \frac{1}{5040} = 0.367857 \approx e^{-1}.$$

"I very willingly acknowledge his [de Montmorts]
Solution to be extreamly good, and own that he has in this,
as well as in a great many other things, shewn himself entirely
master of the doctrine of Combinations, which he has
employed with very great Industry and Sagacity." [8]

–Abraham de Moivre

7.5 Anecdote of de Montmorts correspondence with the Bernoulli family

While de Montmort with great success expanded the theory of combinations, he seems to have had a wide circle of friendship with other mathematicians. Apart from the – in their field – very prominent mathematicians Newton and Leibnitz, he also had contact of mathematical art with the Bernoulli family. The second edition of the *Essay* is therefore not only the work of de Montmort – even though he was the one compiling all the theory. [13]

In his 1713 years edition of *Essay*, he includes 132 pages of letter between him and Johann and Nicolaus I Bernoulli. The correspondence is both friendly and scientific, showing their creativity and how they inspired each other to formulate and solve probabilistic problems with more and more complexity. It all started when de Montmort sent a copy of his *Essay* to Jacob

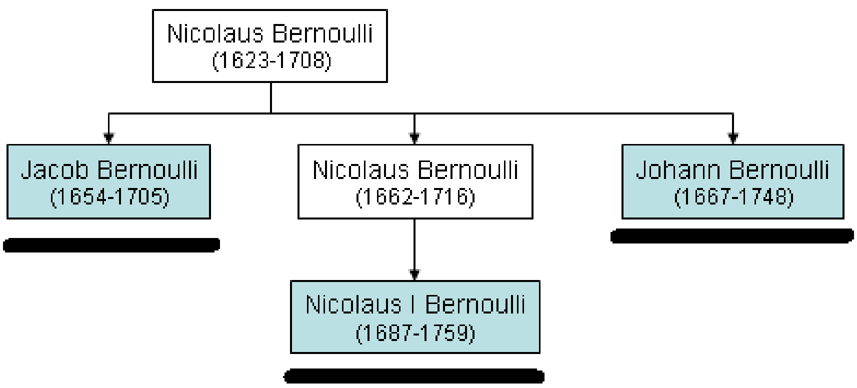
Bernoulli´s older brother, Johann (also known as Jean or John). [2]
He answered in the following way in March 17 1710 (Published in *Essay*).

"As I have received your beautiful Book only a long time after your last Letter, I have well wished to defer the response until I has received & read it, in order to be in a state to tell you of my sentiment of it. Although a flow on the eyes, of which I am often inconvenienced, prevents much work on some things which demand long calculations, especially in the time of winter, I have not left to examine in the idle hours the principal ends of your Treatise, & to do myself, as much as the weakness of my eyes has permitted me, the calculation of the greater part of the Problems…"[3,4]

De Montmort, who was rather unknown, humbly thanks Johann Bernoulli for the honor of reading his book and leaving important remarks. In his correspondence with Nicholas I Bernoulli from 1710 – 1712, the theory of coincidences examined in this chapter is frequently discusses – something that most likely contributed mighty to the theory. [2]
26 February 1711, Nicolaus I Bernoulli wrote the following in a letter to de Montmort.

"I have not yet attempted the general solution of the Problem on the Game of Treize, because it appears to me to be nearly impossible…" [5]

Two years later, the problems of the described impossible kind were obviously conquered. The illustration below shows the family relationship with the important Bernoullis marked.



8 Jacob Bernoulli & his Ars Conjectandi.

We define the art of conjecture, or stochastic art,
as the art of evaluating as exactly as possible the probabilities
of things, so that in our judgments and actions we can always base
ourselves on what has been found to be the best, the most appropriate,
the most certain, the best advised: this is the only object of the wisdom
of the philosopher and the prudence of the statesman…

…

… It seems that to make a correct conjecture about any event whatever,
it is necessary to calculate exactly the number of possible cases and
then to determine how much more likely it is that
one case will occur than another. [6]

– Jacob Bernoulli

Ars Conjectandi was nearly completed when Jacob Bernoulli passed away year 1705, but due to quarrels within the family, not published until 1713. [13]

The composition consists of four parts:

1

The treatise *De ratiociniis in Ludo Aleae* by Huygens with annotations by Jacob Bernoulli. He among other things, develops Huygens' concept of expected value.

2

The doctrine of permutations and combinations, including the twelvefold way. It also discusses the general formula for sums of integer powers, later used by de Moivre.

3

The use of preceding doctrines on various games of chance and dice games. Practical applications of the theory stated in part 2.

4

The use and application of the preceding doctrines on civil, moral, and economic affairs. This part is the shortest one, including the “golden theorem”.

The fourth part of *Ars Conjectandi* is by many, including Jacob Bernoulli himself, considered the most important one.

> *"Therefore, this is the problem which I have*
> *decided to publish here after I have pondered*
> *over it for twenty years. Both its novelty and its*
> *great utility, coupled with its just as great*
> *difficulty, exceed in weight and value all*
> *the other chapters of this doctrine…"* [2]
>
> – Jacob Bernoulli

"The golden theorem", today known as "the law of large numbers" is one of the most fundamental theorems of probability theory and statistics. It is based on the following reasoning stated by Bernoulli.

> *"The more observations that are taken, the*
> *less danger will be of deviating from the truth."* [2]

He continues adding that this is well known and that everyone knows that one or two observations will not be enough to determine the probability of an event. He describes a possible application of the theorem in the following way:

> *…To illustrate this by an example, I suppose that without your knowledge there are concealed in an urn 3000 white pebbles and 2000 black pebbles, and in trying to determine the numbers of these pebbles you take out one pebble after another (each time replacing the pebble you have drawn before choosing the next, in order not to decrease the number of pebbles in the urn), and that you observe how often a white and how often a black pebble is withdrawn. The question is, can you do this so often that it becomes ten times, one hundred times, one thousand times, etc., more probable (that is, it be morally certain) that the numbers of whites and blacks chosen are in the same 3 : 2 ratio as the pebbles in the urn, rather than in any other different ratio?"* [7]
>
> – Jacob Bernoulli

[17]

Swiss commemorative stamp of
Jacob Bernoulli displaying the
formula and a simulation
of his "golden theorem".

8.1 Approach to the golden theorem:

With modern terminology and notations, we consider:

n independent trials with probability *p* for the occurrence of a certain event.

s_n denotes the number of successes, and is binomially distributed (the distribution is further examined in its more natural context in chapter 9). We now look at the relative frequency

$$h_n = s_n / n.$$

We let E denote the event when $|h_n - p| \leq \varepsilon,$ and it may be proven that $P(E) > 1 - \delta$ for $n > n(p, \varepsilon, \delta)$ where ε and δ are any "small" positive numbers. This may be expressed in terms of "convergence in probability",

$$h_n \xrightarrow{p} p \text{ as } n \to \infty.$$

In his proof, Jacob Bernoulli considers a trial of $t = r + s$ equally likely outcomes with r favorable. So that $p = r/(r+s)$. Bernoulli first starts with the simplest case, namely $p = 1/2$ for which he first gives a numerical example followed by a general proof. He then develops the proof for the general probability *p*. [2]

We take note of his own formulation of the theorem:

"It must be shown that so many observations can be made that
it will be c times more probable than not that the ratio of the number
of favorable observations to the total number of observations will be
neither larger than (r + 1)/t nor smaller than (r – 1)/t." [2]

– Jacob Bernoulli

His stated inequality

$$\frac{r-1}{t} \le h_n \le \frac{r+1}{t},$$

May be written as

$$\left|h_n - p\right| \le 1/t,$$

from which we see that $1/t$ corresponds to ε and $1/(1+c)$ to δ.

8.2 Bernoullis theorem of large numbers:

For any positive real number *c*, we have

$$P\left(\left|h_n - p\right| \le \frac{1}{t}\right) > \frac{c}{1+c}, \text{ (1)}$$

for $n = kt$ sufficiently large, for $k \ge k(r,s,c) \vee k(s,r,c)$, where $k(r,s,c)$ denotes the smallest positive integer satisfying

$$k(r,s,\mathrm{t}) \ge \frac{m(r+s+1)-s}{r+1}, \text{ (2)}$$

where *m* denotes the smallest positive integer satisfying

$$m \ge \frac{Log_{\,e}\left[c(s-1)\right]}{Log_{\,e}\left[(r+1)/r\right]}. \text{ (3)}$$

If given p, we may choose t as large as we like so that the interval of the relative frequency $h_n = s_n / n$, becomes arbitrarily small. Jacob Bernoulli´s proof is lengthy and lacks of elegance, due to his omitted indices and, in some cases, unnecessary calculations. Bernoulli for some reason evaluated tail probabilities for both the right and the left tail, even though one follows from one another [2] – something that will therefore not be presented in the proof below.

8.2.1 Proof:

The expected number of successes $np = kr$ is the "central term" in the expansion of $(p+q)^n$. To its left, there are kr terms and to its right, ks terms. This is due to the fact that the expansion of $(p+q)^n$ contains $n + 1 = kr + ks + 1$ terms.

We have to find n such that

$$P_k = P\left(\left|s_n - kr\right| \le k\right) > \frac{c}{1+c},$$

or equivalently, finding n such that

$$\frac{P_k}{1-P_k} > c.$$

We now look at the binomially distributed s_n and remind ourselves that it denotes the number of successes.

$$(p+q)^n = \left(\frac{r}{r+s} + \frac{s}{r+s}\right)^n = \left(\frac{r+s}{t}\right)^n = (r+s)^n t^{-n}.$$

Examining $(r+s)^n$ in terms of the binomial theorem (as well thoroughly evaluated in chapter 9) yields

$$(r+s)^n = \sum_{x=0}^{n} \binom{n}{x} r^x s^{n-x} = \sum_{i=-kr}^{ks} f_i,$$

where f_i was set to represent the following:

$$f_i = \binom{kr+ks}{kr+i} r^{kr+i} s^{ks-i}, \qquad i = -kr, -kr+1, -kr+2, ..., ks.$$

Unlike Jacob Bernoulli, we will only give the proof of the right tail. This will be enough, with reference to the fact that f_{-i} for $i = 0,1,2,...,kr$ is easily obtained from f_i for $i = 0,1,2,...,ks$ by an interchange of *r* and *s*.

Proving that P_k, the central term of the series plus *k* terms to each side, is larger than $c(1-P_k)$, may be done by proving that

$$\sum_{1}^{k} f_i \geq c \sum_{k+1}^{ks} f_i, \qquad k \geq k(r,s,c). \quad (4)$$

By investigating the ratio

$$\frac{f_i}{f_{i+1}} = \frac{\binom{kr+ks}{kr+i} r^{kr+i} s^{ks-i}}{\binom{kr+ks}{kr+i+1} r^{kr+i+1} s^{ks-i-1}} = \frac{(kr+i+1)s}{(ks-i)r} = \frac{rs+(i+1)s/k}{rs-ir/k} > 1, \qquad i = 0,1,2,...,ks-1, \quad (5)$$

the following can be stated:

a) f_i is a decreasing function of *i* for $i \geq 0$.

b) $f_0 = \max(f_i)$.

c) f_i / f_{i+1} is an increasing function of *i* for $i \geq 0$.

d) $f_0 / f_k < f_i / f_{k+i}$ for $i \geq 1$.

Now, by partitioning the tail probability into *s* – 1 terms, each containing *k* terms, the upper bound may be found from property a).

$$\sum_{k+1}^{ks} f_i \leq (s-1) \sum_{k+1}^{2k} f_i,$$

which, when combining with property d), leads to the following inequality:

$$\frac{\sum_{1}^{k} f_i}{\sum_{k+1}^{ks} f_i} \geq \frac{\sum_{1}^{k} f_i}{(s-1)\sum_{k+1}^{2k} f_i} > \frac{f_0/f_k}{s-1}. \quad (6)$$

Therefore, to prove (4), it will be sufficient to prove

$$\frac{f_o}{f_k} \geq c(s-1).$$

It follows from (5) that

$$\frac{f_0}{f_k} = \frac{\binom{kr+ks}{kr} r^{kr} s^{ks}}{\binom{kr+ks}{kr+k+1} r^{kr+k+1} s^{ks-k}} = \ldots = \frac{rs+s}{rs-r+(r/k)} \frac{rs+s-(s/k)}{rs-r+(2r/k)} \cdots \frac{rs+(s/k)}{rs}.$$

To find a lower bound for the above presented ratio, Jacob Bernoulli states that the *k* factors must lie between $(rs+s)/(rs-r)$ and 1. [2] Therefore, any fixed number between these limits can be chosen. Say, $(r+1)/r,$ so that the first factors are larger and the following smaller than this number. Then, by an appropriate choice of *k*, the *m*:th factor, $1 \leq m \leq k,$ becomes equal to $(r+1)/r.$ That is

$$\frac{rs+s-(m-1)(s/k)}{rs-r+(mr/k)} = \frac{r+1}{r},$$

with the following relation between *k* and *m*:

$$k = \frac{m(r+s+1)}{r+1}.$$

Therefore, examining the ratio f_0/f_k for the above stated value of *k*, we get that *m* factors are larger, or equal to $(r+1)/r,$ and *k* – *m* factors larger than 1. We get

$$\frac{f_0}{f_k} \geq \left(\frac{r+1}{r}\right)^m.$$

In conclusion, it thus sufficient to find *m* from

$$\left(\frac{r+1}{r}\right)^m \geq c\left(s-1\right),$$

which yields

$$m \geq \frac{Log_e\left[c\left(s-1\right)\right]}{Log_e\left[\left(r+1\right)/r\right]}.$$

as stated in (3).

From (7) we get (2). Finally, *k* is found as the larger one of the two integers $k\left(r,s,c\right)$ and $k\left(\mathrm{s},r,c\right)$, and *n* is found as *kt*.

This completes “the golden theorem”, or “the law of large numbers”.

8.3 Conclusions of the law of large numbers

Returning to one of the introductory quotes in this chapter, namely about the pebbles in the urn – Jacob Bernoulli could state the following:

Set the number of white pebbles, $r = 20$ and the number of black pebbles, $s = 20$. From his earlier presented quote

> *“It must be shown that so many observations can be made that it will be c times more probable than not that the ratio of the number of favorable observations to the total number of observations will be neither larger than (r + 1)/t nor smaller than (r – 1)/t.”* [2]
>
> – Jacob Bernoulli

and the corresponding inequality

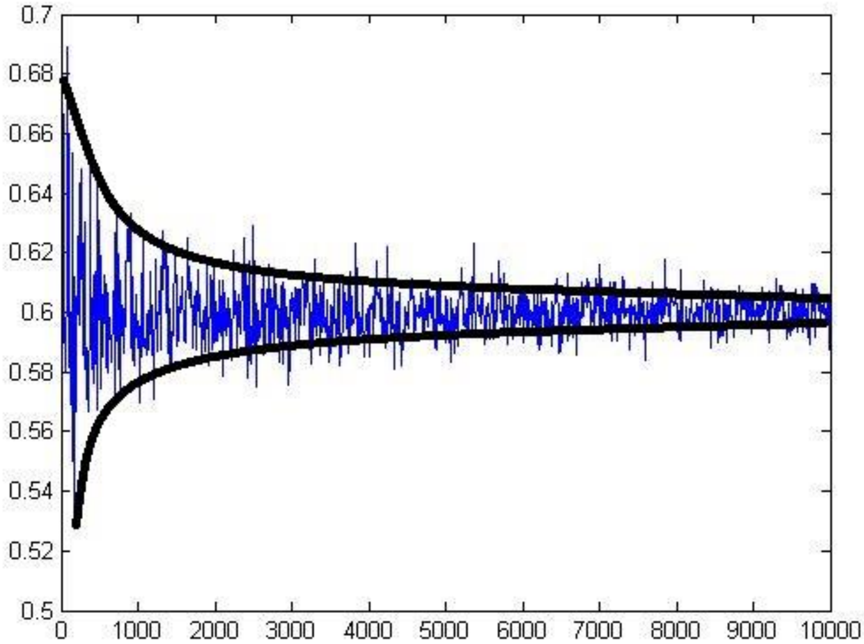
$$\frac{r-1}{t} \leq h_n \leq \frac{r+1}{t},$$

we may using our values set

$$\frac{29}{50} \leq h_n \leq \frac{31}{50}.$$

Choosing $c = 1000$, giving a moral certainty of 1000/1001 for the inequality to hold, Bernoulli finds for the right tale $m = 211$, $k = 511$, and $n = 25550$. In conclusion – for 25550 observations, it is at least 1000 times more probable that h_n will fall inside, than outside the specified interval. [2]

Below, a matlab-simulation clarifies the meaning of Bernoullis golden theorem. We consider a set of 50 elements, of which 30 denote white pebbles, and 20 black pebbles. We pick, and then put back, a random one out of these 50 elements, and repeat the process for a different number of times (varying from 10 to 10000). After each time, we look at the ratio of white pebbles to the total number of pebbles, which by means of the law of large numbers, should approach $\frac{30}{20+30} = \frac{30}{50} = 0.6$ as the number of observations increase.

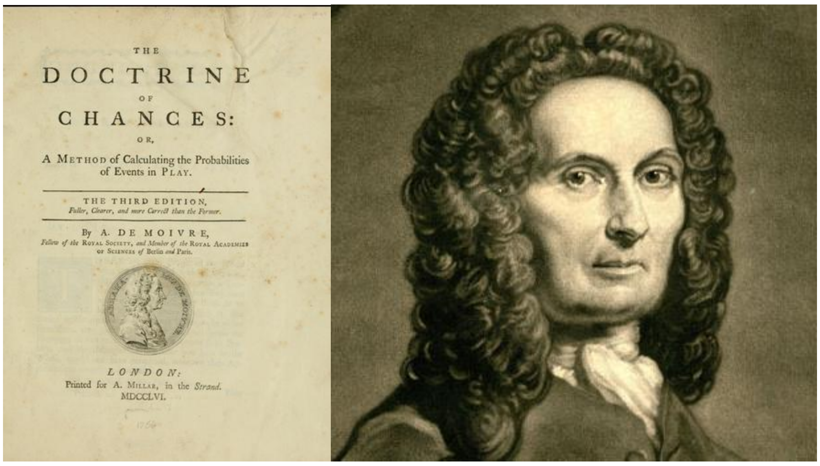


The simulation clarifies, and clearly illustrates a tapered shape with less and less deviation from 0.6 as the number of random-chosen pebbles increases.

The golden theorem is a lot easier proved using more modern methods, like the characteristic functions, which shortens and gives tons of more elegance to the proof.

9 Abraham de Moivre and his Doctrine

Since both Jacob Bernoulli and Pierre Rémond de Montmort died young, and Nicolaus I Bernoulli had become Professor of law – it became the fate of de Moivre to fulfill the work that they all together, in such a splendid way started the century with. [2]
Many would undoubtedly assert that he succeeded with verve.
The section below examines one of the most fundamental parts of de Moivres work.



[18, 19]

9.1 Normal Distribution

The normal distribution is one of the most famous and useful tools in the field of probability theory. The credit for the discovery is often attributed the mathematician Abraham de Moivre (26 May 1667 in Champagne, France – 27 November 1754 in London, England). His writings have had a tremendous impact on the theory used today, even though de Moivre himself never drew all the necessary conclusions to get all credit for inventing the normal distribution. [13]

The name of the distribution, “normal” – apart from that it has had several names during the time – has had more than one meaning. It has both meant orthogonal, and later on more carrying the meaning of common. [20]

To get a better insight in the normal distribution one should first get familiar with another, quite simpler distribution – the binomial distribution. The theory of it is based on the theorem examined in the following section.

9.2 The binomial theorem

The binomial theorem and its triangular arrangement is a discovery that Blaise Pascal is improperly credited with, since many mathematicians from around the world derived similar results hundreds of years earlier. The Indian mathematician Bhaskara seems to have known the general formulae for the number of *n* objects and the number of combinations of *r* among *n* objects about year 1150. [21] The arithmetical triangle and the construction of it were also derived in 1265, by the Arabian mathematician al-Tusi. [2]

The importance in Pascal's work lies in his systematic exposition and the related applications.

The binomial theorem states:

$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k},$$

where $\binom{n}{k}$ from the above described binomial theorem equals:

$$\frac{n!}{\mathrm{k}!(n-k)!}.$$

9.2.1 The binomial distribution

The binomial distribution is one of the discrete distributions used for answering questions of the following kind: "If a fair coin is flipped 100 times, what is the probability of getting exactly 50 heads?". Let *X*: number of heads.

$$P(X=50) = (0.5)^{50} * (1-0.5)^{50} = 0.0795892373871788.$$

Or generally:

$$P(X=k) = \binom{n}{k} p^k q^{n-k}.$$

Where X is the stochastic variable corresponding to the number of heads, $p = 0.5$ is the probability for the number of favorable outcomes, $q = 1 - p = 0.5$ is the number of non-favorable outcomes.

The case where we extend the question to: "If a fair coin is flipped 3600 times, what is the probability of 1900 or more heads?" immediately impedes the calculating process in a time where there were no computers. Examining this would yield:

$$\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k} = \sum_{k=1900}^{3600}\binom{3600}{k}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{3600-k} = \sum_{k=1900}^{3600}\binom{3600}{k}\left(\frac{1}{2}\right)^{3600}.$$

This gave fire to invention of the normal distribution, a discovery that has been part of laying the probabilistic foundation that impact the approximations and algorithms used today.

9.3 Normal approximation to binomial distribution

As the problem of adding together large amounts of $(a + b)^n$ interested de Moivre, very few mathematicians had, due to its seemingly great difficulty and absence of intuitiveness, undertaken the task.

Year 1733 a brief 7-page paper named "*Approximatio ad summam terminorum biomii $(a + b)^n$ in seriem expansi*" was communicated privately to de Moivres friends. In which de Moivre presented interesting observations concerning the coefficients of the binomial expansions of $(a + b)^n$. A few years later, 1738, his own translated version with some extra additions was included in *The Doctrine of Chances* (1738) and the first statements of what we today consider the "normal curve" were now published. [25]

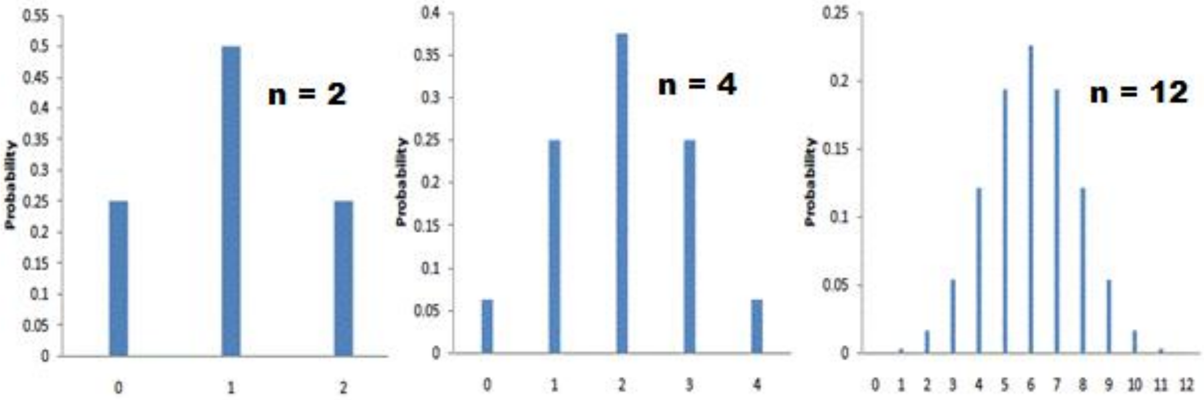
De Moivre wrote the following:

"Altho' the Solution of problems of Chance often requires
that several Terms of the Binomial $(a + b)^n$ be added together,
nevertheless in very high Powers the thing appears so laborious,
and of so great difficulty, that few people have undertaken that Task:
for besides Jacob and Nicolaus I Bernouilli, two great Mathematicians,
I know of no body that has attempted it: in which, tho' they have shown
very great skill, and have the praise that is due to their Industry,
yet some things were further required: for what they have done
is not so much an Approximation as the determining
very wide limits, within which they demonstrated that
the Sum of the Terms was contained. " [8]

In Abraham de Moivre´s year 1738 edition of “*The Doctrine of Chances*” he worked out the mathematics for the binomial expansion of $(a + b)^n$ by analyzing the tosses of a coin. Though, whether de Moivre thought of his events as tossing coins or not, remains unimportant. What´s important is the elucidation that he derived events so dependent of chance, that the probabilities of favorable/ non-favorable outcome, to be equal.

He wanted to find a function that approximated the shape that came about when the number of binomial events (coin flips) increased, with a continuous curve. He would then be able to calculate the sum corresponding to "If a fair coin is flipped 3600 times, what is the probability of 1900 or more heads?” in a much less strenuous way.

The matlab-plot below illustrates the probability for different numbers of “heads” when tossing 2, 4 or 12 coins. The smooth shape seems clearer as the number of coins increases.



Once again looking at the binomial theorem – and examining it for a few base cases:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Examining $(1 + x)^n$ using the theorem above yields the following for n = 1, 2, 3, 4, 5:

$$1 + x$$

$$1 + 2x + x^2$$

$$1 + 3x + 3x^2 + x^3$$

$$1 + 4x + 6x^2 + 4x^3 + x^4$$

$$1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5$$

Now as $\binom{n}{k}$ represents the coefficient of x^k in the expansion of $(1+x)^n$, the following can be stated:

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

The coefficients in each expansion, clearly correspond to the arithmetical triangle – with the following look for n = 0, 1 ,2 ,3 ,4 ,5.

$$\begin{array}{c}
\binom{0}{0} \\
\binom{1}{0} \quad \binom{1}{1} \\
\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2} \\
\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3} \\
\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4} \\
\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5}
\end{array}$$

Below, we will familiarize with two theorems – in this context of fundamental relevance – composed by Abraham de Moivre.

9.4 De Moivres approximation to the maximum of the symmetric binomial

Abraham de Moivre approximated the biggest term in the binomial expansion of the symmetric binomial $(p + q)^n$, $p = q = 0.5$, by the following.

$$\binom{n}{0.5n}\left(\frac{1}{2}\right)^n \sim 2.168\frac{\left(1-\frac{1}{n}\right)^n}{\sqrt{n-1}} \cong \frac{0.7976}{\sqrt{n}} \qquad (1)$$

9.4.1 Proof:

As a benchmark of the proof, observe:

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

Since the middle term – due to symmetry – is the biggest one in the expansion, the proof of the above stated approximation will be found advantageously by analyzing the following term:

$$\binom{n}{0.5n}\left(\frac{1}{2}\right)^{n}.$$

We now introduce the following notation:

$$b(k)=b(k,n,p)=\binom{n}{k}p^{k}(1-p)^{n-k},\quad k=0,1,\ldots,n,\qquad n=1,2,\ldots.$$

For n = 2m, we have

$$b(m+d)=\binom{2m}{m+d}\left(\frac{1}{2}\right)^{2m},|d|=0,1,\ldots,m,\qquad m=1,2,\ldots.$$

Since the middle term, as stated, will be the biggest one we set d = 0.

$$b(m)=\frac{2m(2m-1)\cdots(m+1)}{1\cdot 2\cdots m}2^{-2m}=2^{2m-1}\prod_{i=1}^{m-1}\frac{m+i}{m-i},$$

taking the logarithm yields

$$Log_{e}\left((b(m)\right)=Log_{e}\left(2^{-2m+1}\right)+\sum_{i=1}^{m-1}Log_{e}\left(\frac{1+i/m}{1-i/m}\right)$$

$$=\left(-2m+1\right)\cdot Log_{e}\left(2\right)+\sum_{i=1}^{m-1}Log_{e}\left(\frac{1+i/m}{1-i/m}\right).\qquad (2)$$

Now, by referring to Newtons series

$$Log_{e}\left(\frac{1+x}{1-x}\right)=2\sum_{k=1}^{\infty}\frac{x^{2k-1}}{2k-1},\qquad (3)$$

the sum stated in (2) may be written as

$$2\sum_{i=1}^{m-1}\sum_{k=1}^{\infty}\frac{1}{2k-1}\left(\frac{i}{m}\right)^{2k-1}=2\sum_{k=1}^{\infty}\frac{1}{\left(2k-1\right)m^{2k-1}}\sum_{i=1}^{m-1}i^{2k-1}.\qquad (4)$$

Now, once again referring to the work of the great mathematician presented in chapter 8 – Jacob Bernoullis formula for the summation of powers of integers:

$$\sum_{i=1}^{n}i^{m}=\frac{n^{m+1}}{m+1}+\frac{n^{m}}{2}+\frac{1}{2}\binom{m}{1}B_{2}n^{m-1}$$

$$+\frac{1}{4}\binom{m}{3}B_{4}n^{m-3}+\frac{1}{6}\binom{m}{5}B_{6}n^{m-5}+\ldots,$$

where B_i represents the Bernoulli number. (Named after and first published by Jacob

Bernoulli in his *Ars Conjectandi*). [26]

De Moivre could therefore ensure that the last sum in (4) could be written as

$$\sum_{i=1}^{m-1} i^{2k-1} = \frac{(m-1)^{2k}}{2k} + \frac{1}{2}(m-1)^{2k-1} + \frac{1}{2}(2k-1)B_2(m-1)^{2k-2} + \ldots, \qquad (5)$$

Setting $(m-1)/m = t,$ and inserting expression (5) into (4) yields

$$(2m-1)\sum_{k=1}^{\infty} \frac{t^{2k-1}}{(2k-1)2k} + \sum_{k=1}^{\infty} \frac{t^{2k-1}}{2k-1} + \frac{B_2}{m}\sum_{k=2}^{\infty} t^{2k-2} + \ldots. \qquad (6)$$

To continue on, valuating the first of these three sums, de Moivre integrated (3) from $x = 0$ to $x = t$, and dividing by t. This yields:

$$2\sum_{k=1}^{\infty} \frac{t^{2k-1}}{(2k-1)2k} = Log_e\left(\frac{1+t}{1-t}\right) + t^{-1}Log_e\left(1-t^2\right)$$

$$= Log_e(2m-1) + \frac{m}{m-1}Log_e\left(\frac{2m-1}{m^2}\right).$$

The second sum in (6) may be found directly by means of Newton´s series stated in (3).

$$\sum_{k=1}^{\infty} \frac{t^{2k-1}}{2k-1} = \left(\frac{1}{2}\right)Log_e\left(\frac{1+t}{1-t}\right) = \left(\frac{1}{2}\right)Log_e(2m-1).$$

The third sum may successfully be written as

$$\frac{B_2}{m}\sum_{k=2}^{\infty} t^{2k-2} = \frac{B_2}{m}t^2\left(1-t^2\right)^{-1} = \frac{(m-1)^2}{6m(2m-1)},$$

and

$$\lim_{m\to\infty} \frac{(m-1)^2}{6m(2m-1)} = \frac{1}{12}.$$

De Moivre also found that the 3 following terms tend to –1/360, 1/1260 and –1/1680. (7)

Inserting these results into expression (2) yields the following

$$Log_e\left(b(m)\right) \sim \left(2m-\frac{1}{2}\right)Log_e\left(2m-1\right)-2mLog_e\left(2m\right)+Log_e\left(2\right)$$

$$+\frac{1}{12}-\frac{1}{360}+\frac{1}{1260}-\frac{1}{1680}+\ldots \text{ (8)}$$

Which clearly corresponds to the following: (9)

$$(2m-\frac{1}{2})\,\mathrm{Log}_e\left(2m-1\right)-2m\,\mathrm{Log}_e\left(2m\right)+\mathrm{Log}_e\left(2\right)+\frac{B_2}{1\times2}+\frac{B_4}{3\times4}+\frac{B_6}{5\times6}+\frac{B_8}{7\times8}\cdots$$

$$=2m\,\mathrm{Log}_e\left(2m-1\right)-2m\,\mathrm{Log}_e\left(2m\right)-\frac{1}{2}Log_e\left(2m-1\right)+\mathrm{Log}_e\left(2\right)+\frac{B_2}{1\times2}+\frac{B_4}{3\times4}+\frac{B_6}{5\times6}+\frac{B_8}{7\times8}\cdots$$

Finally, we note from the above stated expression (9) that

$$2m\,\mathrm{Log}_e\left(2m-1\right)-2m\,\mathrm{Log}_e\left(2m\right)=2m\,\mathrm{Log}_e\left(1-\frac{1}{2m}\right) \text{ (10)}$$

With knowledge of Jacob Bernoullis discovery of the constant e, one sees that since:

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n=e \text{ and } \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n=e^{-1}$$

the quantity in (10)

$$2m\,\mathrm{Log}_e\left(1-\frac{1}{2m}\right)\cong-1$$

when *n*, and therefore *m*, is a big number.

From the remaining part of (9) we see that

$$-\frac{1}{2}Log_e\left(2m-1\right)+\mathrm{Log}_e\left(2\right)+\frac{B_2}{1\times2}+\frac{B_4}{3\times4}+\frac{B_6}{5\times6}+\frac{B_8}{7\times8}\ldots\approx-\frac{1}{2}Log_e\left(2m-1\right)+0.7739.$$

We then finally arrive to the part where only the following expression needs to be evaluated

$$Log_e\left(b(m)\right)\cong\frac{-1}{2}Log_e\left(2m-1\right)-1+0.7739.$$

$$\Rightarrow \left(2m-1\right)^{-1/2}\cong\left(2m\right)^{-1/2} \Rightarrow$$

$$b(m)\cong\frac{0.7976}{\sqrt{2m-1}}\cong\frac{0.7976}{\sqrt{2m}}=\frac{0.7976}{\sqrt{n}}=\frac{2}{B\sqrt{n}}.$$

where

$$B=e^{1-\frac{1}{12}+\frac{1}{360}-\frac{1}{1260}+\frac{1}{1680}\ldots}.$$

which equals the expression stated in (1).

The proof is of huge importance since it clearly demonstrates the methods found in de Moivres other proofs. [2] The only part of the proof that may not be consider satisfactory, is deriving the constant 2.168 found from the exponential of the following expression:

$$Log_e(2)+1/12-1/360+1/1260-1/1680.$$

We look at the sum

$$\sum_{k=1}^{\infty}\frac{B_{2k}}{(2k-1)2k},$$

that gives the sequence $1/12-1/360+1/1260-1/1680\ldots$

By publishing only the first four terms, de Moivre avoided the discussion of the properties of the series, namely whether it was divergent or convergent.

Deriving the first 10 terms in the series yields

$$\begin{aligned}&0.0833-0.0028+0.0008-0.0006+0.0008\\&\qquad-0.0019+0.0064-0.0296+0.1796-1.3924.\end{aligned}$$

The terms start increasing after the fourth term, which clarifies why de Moivre chose to stop there.

Here, de Moivre desisted from proceeding farther until his friend, the mathematician James Stirling proved that $B=e^{1-\frac{1}{12}+\frac{1}{360}-\frac{1}{1260}+\frac{1}{1680}\cdots}$ actually equals $\sqrt{2\pi}$. [2]

A statement that de Moivre credited Stirling with by phrasing the following:

> "*…I own with pleasure that this discovery, besides that it has saved trouble, has spread a singular Elegancy on the solution.*" [8]

The table below shows the exactness of the approximation for different *m*

	$\binom{2m}{m}\left(\frac{1}{2}\right)^{2m}$	$0.7976\big/\sqrt{2m}$
m = 3	0.3125	0.3256
m = 8	0.1964	0.1994
m = 10	0.1762	0.1783
m = 50	0.0796	0.0798
m = 400	0.0282	0.0282

9.5 De Moivres approximation to b(*m*)/b(*m + d*).

Continuing on in the same manner, de Moivre came up with an approximation to b(*m*)/b(*m* + *d*). The methods used in the proof are very similar to the approximation to b(*m*) above.

$$Log_e\left(\frac{b(m)}{b(m+d)}\right) \sim \left(m+d-\frac{1}{2}\right)Log_e\left(m+d-1\right)$$

$$+\left(m-d+\frac{1}{2}\right)Log_e\left(m-d+1\right)-2mLog_e\left(m\right)+Log_e\left(\frac{m+d}{m}\right).$$

This being admitted, de Moivre could ensure himself that if n = 2m being infinitely great, then

$$Log_e\left(\frac{b(m+d)}{b(m)}\right)=Log_e\left(\frac{\binom{2m}{m+d}\left(\frac{1}{2}\right)^{2m}}{\binom{2m}{m}\left(\frac{1}{2}\right)^{2m}}\right)\cong\frac{-2d^2}{n}.$$

9.5.1 Proof:

$$Log_e\left(\frac{b(m+d)}{b(m)}\right)=-Log_e\left[\frac{m+d}{m}\cdot\frac{1+1/m}{1-1/m}\cdot\frac{1+2/m}{1-2/m}\cdot\ldots\frac{1+\left(m-1\right)/m}{1-\left(m-1\right)/m}\right]$$

$$=\ldots$$

$$=-\left(m+d-\frac{1}{2}\right)Log_e\left(m+d-1\right)-\left(m-d+\frac{1}{2}\right)Log_e\left(m-d+1\right)-2mLog_e\left(m\right)-Log_e\left(\frac{m+d}{m}\right)$$

$$= -\left(m + d - \frac{1}{2}\right) Log_e\left(1 + \frac{d-1}{m}\right) - \left(m - d + \frac{1}{2}\right) Log_e\left(1 - \frac{d-1}{m}\right) - Log_e\left(1 + \frac{d}{m}\right)$$

$$\cong -(m+d)\, Log_e\left(1 + \frac{d}{m}\right) - (m-d)\, Log_e\left(1 - \frac{d}{m}\right).$$

By means of the following taylor expansion

$$Log_e(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots \text{ for } |x| \le 1,$$

one may continue in the following manner:

$$\cong -(m+d)\left(\frac{d}{m} - \frac{d^2}{2m^2} + \dots\right) - (m-d)\left(-\frac{d}{m} - \frac{d^2}{2m^2} - \dots\right)$$

$$= \frac{-d^2}{m} = \frac{-2d^2}{n},$$

and the approximation

$$Log_e\left(\frac{b(m+d)}{b(m)}\right) \cong \frac{-2d^2}{n} \rightarrow \frac{b(m+d)}{b(m)} \cong e^{\frac{-2d^2}{n}}$$

was proven.

In conclusion, we note that

$$\frac{b(m+d)}{b(m)} \cong e^{\frac{-2d^2}{n}} \rightarrow b(m+d) \cong b(m) \cdot e^{\frac{-2d^2}{n}} \cong \frac{0.7976}{\sqrt{n}} \cdot e^{\frac{-2d^2}{n}} \approx \frac{2}{\sqrt{2\pi n}} \cdot e^{\frac{-2d^2}{n}}.$$

9.6 Conclusions of the theorems.

As the starting point of de Moivres work was to come up with an easier way of deriving cases when several terms of the binomial $(a + b)^n$ be added together, he applied his newly found knowledge in the following manner:

$$P_d = \sum_{k=\frac{n}{2}-d}^{\frac{n}{2}+d} \frac{2}{\sqrt{2\pi n}} e^{\frac{-2k^2}{n}} \cong \frac{2}{\sqrt{2\pi n}} \int_0^d e^{\frac{-2x^2}{n}} dx \cong \frac{4}{\sqrt{2\pi}} \int_0^{\mathrm{d}/\sqrt{n}} e^{-2x^2} dx.$$

He found two inflection points (points on a curve at which the curve changes from being concave to convex or vice versa).

Inflection point 1: $\frac{n}{2} + \frac{1}{2}\sqrt{n}.$

Inflection point 2: $\frac{n}{2} - \frac{1}{2}\sqrt{n}$.

Which clearly corresponds to setting $d = \frac{1}{2}\sqrt{n}$.

To calculate the integral for small values of d, de Moivre expanded the exponential function and integrated the resulting series. [2]

For the sum of seven terms, this yields:

$P_{\sqrt{n}/2} \cong 0.682688.$

He derives the accuracy of the approximation for various values of *n* and concludes with the following quote:

"Still, it is not to be imagined that there is any
necessity that the number n should be immensely great:
for supposing it not to reach beyond the 900th power, nay not even
beyond the 100th, the rule here given will be tolerably accurate,
which I have confirmed by Trials." [8]

He also calculated the integral for the values $d = \sqrt{n}$ and $d = \frac{3}{2}\sqrt{n}$ by means of numerical integration, namely Newton's three-eight's rule. [2] (This rule will not be further examined here).

He landed in the following:

$P_{\sqrt{n}} \cong 0.95428,$

$P_{(3/2)\sqrt{n}} \cong 0.99874.$

The table presented below compares de Moivre´s results with the correct values:

	De Moivre´s approximations	Correct value
$P_{\sqrt{n}/2}$	0.68268<u>8</u>	0.682689
$P_{\sqrt{n}}$	0.954<u>28</u>	0.95450
$P_{(3/2)\sqrt{n}}$	0.99<u>874</u>	0.99730

With the notation of today, the above stated correspond to deviations of one, two, and three times the standard deviation.

By applying the newly examined approximations on a more applied problem, de Moivre could now calculate the probabilities of coin-tossing in a much less strenuous way. If repeating the experiment of tossing 3600 coins unpleasantly many times, one should now be very aware of the following:

Since

$$n = 3600 \to \frac{1}{2}n = 1800 \to \frac{1}{2}\sqrt{n} = 30,$$

"...then the probability of the Event's neither appearing oftner than 1830 times, nor more rarely than 1770, will be 0.682688." [8]

and the kind of problems that gave fire to the, of huge importance, approximations were now with enormously greater elegance examined.

De Moivre completes his analysis of the binomial by extending the theorems to the general case for p and q – the skew binomial. He writes:

"If the Probabilities of happening and failing be in any given Ratio of inequality, the Problems relating to the sum of (a + b)^n will be solved with the same facility as those in which the Probabilities of happening and failing are in Ratio of equality". [8]

He established

$$b(np+d,n,p) \cong \frac{1}{\sqrt{2\pi pq}} e^{-\frac{d^2}{2npq}}.$$

From which it is easily derived that

$$P_d \cong \frac{4}{\sqrt{2\pi}} \int_0^{\frac{d}{2\sqrt{npq}}} e^{-2x^2} dx.$$

The proof is omitted, but with no difficulty derived using only the methods and results from the previous proofs presented in this chapter.

9.7 Further developments

Some would claim that a chapter describing the emergence of the normal distribution should instead be dedicated to “the prince of mathematicians”, Carl Friedrich Gauss (30 April 1777 – 23 February 1855). Abraham de Moivre never used the concept of “probability density function” and nor did he ever interpret his theorems as anything more than an approximation for the binomial coefficients. [7]

Gauss invented the two-parameter distribution, depending on both the expected value and the standard deviation. He among other things used it for estimating errors in astronomical observations. [27]

The details of Gauss’ work will not be presented further here.

A normal distribution density function has the following look with modern notation.

$$f\left(x,\mu,\sigma\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

10 References

[1] http://www.york.ac.uk/depts/maths/histstat/pascal.pdf (2015-04-14)

[2] A history of probability and statistics and their applications before 1750, Anders Hald, 1990.

[3] Essay d'Analyse sur les Jeux de Hazard, Seconde edition, 1713.

[4] http://www.cs.xu.edu/math/Sources/JakobBernoulli/montmort.pdf (2015-04-14)

[5] https://math.dartmouth.edu/~doyle/docs/treize/treize.pdf (2015-04-14)

[6] http://www-history.mcs.st-and.ac.uk/Quotations/Bernoulli_Jacob.html (2015-04-14)

[7] The History of Statistics: The Measurement of Uncertainty Before 1900, Stephen M. Stigle, 1990.

[8] The Doctrine of Chances, 3rd edition, Abraham de Moivre, 1756.

[9] Sannolikhet - en brevväxling. Pascals brev till Fermat, Alfred Rényi, 1972.

[10] Einstein and the Poet: In Search of the Cosmic Man Paperback, William Hermanns, 1983.

[11] Randomness, Deborah J. Bennett, 1999.

[12] The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference 2nd edition, Ian Hacking, 2006.

[13] Games, Gods & Gambling: A History of Probability and Statistical Ideas, F. N David, 1998.

[14] The Biology of Gambling, Mikael Aasved, 2003.

[15] http://www.math.nyu.edu/~hand/teaching/2006-2007/fall/ma224/Sec1-Problem-of-the-Points.pdf (2015-04-14)

[16] A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace, Isaac Todhunter, 1865.

[17] http://www.mathematics-in-europe.eu/en/home/76-enjoy-maths/strick/756-jacob-bernoulli-january-6-1655-august-16-1705-by-heinz-klaus-strick-germany (2015-04-14)

[18] https://ia600807.us.archive.org/zipview.php?zip=/0/items/olcovers580/olcovers580-L.zip&file=5806510-L.jpg (2015-04-14)

[19] http://www.famous-mathematicians.com/abraham-de-moivre/ (2015-04-14)

[20] http://www-biba.inrialpes.fr/Jaynes/cc07s.pdf (2015-04-14)

[21] Handbook of combinatorics 2nd edition, Ronald L. Graham, Martin Grötschel, 1995.

[22] http://johnquincy.blogspot.se/2012_09_01_archive.html (2015-04-14)

[23] http://www.educ.fc.ul.pt/icm/icm2002/icm101/pagina9.html (2015-04-14)

[24] http://www.elperiodico.com/es/noticias/agenda/experimento-pionero-502657

[25] http://www.york.ac.uk/depts/maths/histstat/demoivre.pdf (2015-04-14)

[26] http://global.britannica.com/EBchecked/topic/36202/Ars-Conjectandi (2015-04-14)

[27] A History of Mathematical Statistics From 1750 to 1930, Anders Hald, 1998.

[28] https://specialsamlingar.files.wordpress.com/2014/03/spelbord1.jpg (2015-04-14)

For understanding mathematical proofs, ideas of proofs, and formulations of theorems, the author has used primarily the following books:

- A history of probability and statistics and their applications before 1750, Anders hald, 1990.
- An intermediate course in probability, Allan Gut, 2009.
- Discrete and Combinatorial Mathematics: An Applied Introduction, 5th edition, Ralph P. Grimaldi, 2003.