Review[1] of Liber *De Ludo Aleae* (Book on Games of Chance) by Gerolamo Cardano

1. Biographical Notes Concerning Cardano

Gerolamo Cardano (also referred to in the literature as Jerome Cardan), was born in Pavia, in present day Italy, in 1501 and died at Rome in 1576. Educated at the universities of Pavia and Padua, Cardano practised as a medical doctor from 1524 to 1550 in the village of Sacco and in Milan. During this period he appears to have studied mathematics and other sciences. He published several works on medicine and in 1545 published a text on algebra, the Ars Magna. Among his books is the Liber de Ludo Aleae (Book on Games of Chance), written sometime in the mid 1500s, although unpublished until 1663.

2. Review of De Ludo Aleae

Cardano's text was originally published in Latin, in 1663. An English translation by Sydney Henry Gould is provided in Professor Oystein Ore's book *Cardano, The Gambling Scholar* (Princeton University Press, 1953). Professor Ore's book provides both biographical information relating to Cardano, as well as commentary on Cardano's presentation of a probability theory relating to dice and card games.

In De Ludo Aleae Cardano provides both advice and a theoretical consideration of outcomes in dice and card games. The text (as published) is composed of 32 short chapters. The present review is concerned principally with chapters 9 to 15, illustrating aspects of the theory concerning dice.

The first eight chapters provide a brief commentary on games and gambling, offering advice to players, and suggesting both the dangers and benefits in playing. It may be of interest to quote some of Cardano's comments regarding the playing of games of chance:

"...in times of great anxiety and grief, it is considered to be not only allowable, but even beneficial."

"..in times of great fear or sorrow, when even the greatest minds are much disturbed, gambling is far more efficacious in counteracting anxiety than a game like chess, since there is the continual expectation of what fortune will bring."

"In my own case, when it seemed to me after a long illness that death was close at hand, I found no little solace in playing constantly at dice."

"However, there must be moderation in the amount of money involved; otherwise, it is certain that no one should ever play."

"..the losses incurred include lessening of reputation, especially if one has formerly enjoyed any considerable prestige; to this is added loss of time...neglect of one's own business, the danger it may become a settled habit, the time spent in planning after the game how one may recuperate, and in remembering how badly one has played."

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.

Cardano especially warns that lawyers, doctors and those in like professions avoid gambling, which could be injurious to their reputations and business. Interestingly, he adds:

"Men of these professions incur the same judgement if they wish to practice music."

In Chapter 6 Cardano presents what he refers to as the Fundamental Principle of Gambling:

"The most fundamental principle of all in gambling is simply equal conditions...of money, of situation...and of the dice itself. To the extent to which you depart from that equality, if it is in your opponent's favour, you are a fool, and if in your own, you are unjust."

What is most important for our purposes, is to recognise that Cardano's fundamental principle states that games of chance can only be fairly played when there are equiprobable outcomes. This principle is the basis for his theory relating to outcomes in games of dice.

Cardano begins to present (the results) of his theory on dice in Chapter 9: *On the Cast of One Die*. Given that a die has six points, he states:

"...in six casts each point should turn up once; but since some will be repeated, it follows that others will not turn up."

We see here that his principle is at work (the symmetry of the die allows equiprobable outcomes), and also that he recognises (confirmed by experience no doubt), that the principle is an ideal, and that in practice we will not have each point turn up once in every six casts. There would appear to be an implicit understanding of a "long range relative frequency" interpretation of *"in six casts each point should turn up once".* Or, in the contemporary language of probability theory, we would say that we *expect* in six casts each point should turn up once.

In this chapter, the concepts referred to as "circuit" and "equality" are introduced:

"One-half of the total number of faces always represents equality; thus the chances are equal that a given point will turn up in three throws, for the total circuit is completed in six, or again that one of three given points will turn up in one throw. For example, I can as easily throw one, three, or five as two, four, or six."

The "circuit" refers to the number of possible (elementary) outcomes, what in contemporary probability theory may be referred to as "the size of the sample space". "Equality" appears to be a concept related to expectation. Since a given point on a die is expected to turn up once in six throws (the circuit), it could equally turn up in the first or second three casts. Cardano also provides a variation on this interpretation, indicating that in one throw, three given points (1,3,5) could turn up as easily as the three other points (2,4,6). Equality then can be understood as defined, that is, one-half of the circuit, or as (in contemporary terms) an event, which is as likely as its complementary event (that is, an event with probability one-half).

Professor Ore suggests that the concept of equality is a consequence of Cardano having "the practical game in mind":

"...he seems to assume that usually there are only two [players]...each will stake the same amount A so that the whole pot is $P = 2A$. When a player considers how much he has won or lost it is natural to relate it not to the whole pot 2A but to his own stake A. In terms of such a measure his expectation becomes

$$E = pP = 2pA = p_eA$$

[where p refers to the proportion of favourable outcomes for one player and p_e is called the *equality proportion* by Professor Ore]

so that the equality proportion or the double probability becomes the natural factor measuring loss or gain. ...In a fair game…the number of favourable and unfavourable cases must be the same and each player has the same probability [1/2]. …This means that each player has equality in his favourable cases, so that the corresponding equality proportions are [1]. And Cardano expresses this simply by saying that "there is equality".

In Chapter 11 Cardano discusses the case of two dice. He enumerates the various possible throws:

"…there are six throws with like faces, and fifteen combinations with unlike faces, which when doubled gives thirty, so that there are thirty-six throws in all, and half of these possible results is eighteen."

It is somewhat interesting that Cardano does not at this stage provide an illustration, or table to aid his explanation. If the outcomes of the cast of two dice are represented by ordered pairs, the throws with like faces are (1,1), (2,2)…(6,6), six in all, and those with unlike faces are: (1,2), (1,3), (1,4), (1,5), (1,6), (2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6), fifteen in number, and finally: (2,1)…(6,5), another fifteen. In total there are, as Cardano states, 36 possible outcomes. His use of the text only description surely would have made the subject more difficult for the reader to appreciate his reasoning – unless the reader was well versed in the subject matter. A lay reader for example, would likely ask why the unlike face combinations would have to be doubled. As it turns out, this manner of explanation is typical in Cardano's text, although he does provide some illustrations. For this reason it would seem a reasonable conjecture that the work is intended for those persons familiar with gambling.

In this chapter, a result is given, comparing how likely, relative to equality, it is to get at least one die with one point (an ace) in each of two casts of two dice:

"The number of throws containing at least one ace is eleven out of the circuit of thirty-six; or somewhat more than half of equality; and in two casts of two dice the number of ways of getting at least one ace twice is more than 1/6 but less than 1/4 of equality."

Cardano does not describe how he derived this result, however the following reasoning seems quite possible. The number of ways of getting at least one ace is 11 – (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (3,1), (4,1), (5,1) and (6,1). As 11 is less than 12, there are fewer than 12 times 12 or 144 ways of getting at least an ace in both casts of the dice. In two casts of two dice, there are 36 times 36 or 1,296 possible outcomes (the circuit in this case). Equality is defined as half of 1,296, which is 648. 144 divided by 648 is less than 1/4. Also, since 10 is less than 11, there are more than 10 times 10 or 100 ways of getting at least an ace in both casts. As 100 divided by 648 is more than 1/6, Cardano's result is confirmed. Note that the statement "more than 1/6 but less than 1/4 of equality" in terms of modern probability would be "more than 1/6 times 1/2 or 1/12 but less than 1/4 times 1/2 or 1/8", that is the probability of the event is between 1/8 and 1/12. Why wasn't Cardano more exact? If the above reasoning was followed, he would have known that there were 121 possible ways in which the aces could turn up. Possibly, for his purposes, the precise fraction is unnecessary. It may have been sufficient to explain that a game in which a player wagers on the occurrence of at least one ace in two casts of two dice, was not a fair game. The player could expect only between 1/4 and 1/6 of their own stake. Also, interestingly, the use of upper and lower bounds suggests some variation in practice – although we can not say for sure that this is what Cardano wished to express. It may have been seen as a more aesthetic way of describing the proportion. Clearly however, concepts and problems familiar to modern students of probability are being considered.

In Chapter 12, the casting of three dice is considered. Again the possible outcomes are enumerated, the total (circuit) being 216. While the wording is somewhat vague, Cardano appears to make an error in reporting the number of outcomes with at least one unspecified point (such as an ace):

"...out of the 216 possible results, each single face will be found in 108 and will not be found in as many."

According to Professor Ore his reasoning seems to be that for one cast of a die, 1/6 of a given point can be expected to turn up. In 3 throws, 3 times 1/6 or 1/2 of the point will occur. Of 216 possible outcomes, 108 would be favourable. However, if Cardano realised (as appears to be the case) that there were 11 possible outcomes for a single point in two throws of the dice, would he make such an error? Was it an approximation only? Cardano does indicate in subsequent chapters that there are 91 possible outcomes, which is correct.

Chapters 13 and 14 concern outcomes of the sum of points. The theory is an extension of the earlier results, and Cardano observes:

"In the case of two dice, the points 12 and 11 can be obtained respectively as (6,6) and (6,5). The point 10 consists of (5,5) and of (6,4) but the latter can occur in two ways, so that the whole number of ways of obtaining 10 will be 1/12 of the circuit and 1/6 of equality."

Note that Cardano uses ordered pairs to illustrate the outcomes (presuming that the translated text does not use this for simplification). Also note that the expression 1/12 of the circuit is what we would refer to as a probability of 1/12.

In Chapter 14 the use of the term "odds" is found, as we would apply it today:

"If therefore, someone should say, 'I want an ace, a deuce, or a trey, you know that there are 27 favourable throws, and since the circuit is 36, the rest of the throws in which these points will not turn up will be 9; the odds will therefore be 3 to 1.'"

An incorrect computation of odds is found in the same chapter:

"If it is necessary for someone that he should throw at least twice, then you know that the throws favourable for it are 91 in number, and the remainder is 125; so we multiplying each of these numbers by itself and get to 8,281 and 15,625, and the odds are about 2 to 1."

Cardano realises that an error has been made, and discusses this in chapter 15:

"This reasoning seems to be false... for example, the chance of getting one of any three chosen faces in one cast of one dice is equal to the chance of getting one of the other three, but according to this reasoning there would be an even chance of getting a chosen face each time in two casts, and thus in three, and four, which is most absurd. For if a player with two dice can with equal chances throw an even and an odd number, it does not follow that he can with equal fortune throw an even number in each of three successive casts."

It is interesting that after having made this observation, the earlier text is not corrected. Professor Ore notes that this is typical of Cardano's presentations. The passage is also interesting for its use of the words "chance" and "fortune" relating to the possible outcomes of throws. In the following paragraph the word "probability" is used:

"In comparison where the probability is one half, as of even faces with odd, we shall multiply the number of casts by itself and subtract one from the product, and the proportion which the remainder bears to unity will be the proportion of the wagers to be staked. Thus, in 2 successive casts we shall multiply 2 by itself, which will be 4; we shall subtract 1; the remainder is 3; therefore the player will rightly wager 3 against 1..."

Cardano continues to discuss the computation of odds in the chapter, however what is of principal interest is the use of the word "probability". The above passage is quite possibly the first application of the word in written form, with the meaning comparable to its use in the modern theory (based on symmetry or a long-range relative frequency definition).

It is well known that the theory of probability has its origins in questions on gambling. Why is this the case? Although people were aware of the variable and unpredictable character of every day phenomena (such as the weather, commodity prices, etc.), games of chance lend themselves to a mathematical discussion because the universe of possibilities is (relatively) easily known and computed, at least for simple games.

Cardano's text would appear to be the first known mathematical work on the theory of probability, although published after the more famous correspondence between Pascal and Fermat.

Review[1] of *Sopra Le Scoperte dei Dadi* (Concerning an Investigation on Dice)

1. Biographical Notes

Galileo Galilei was born in Pisa in 1564. His early education was at the Jesuit monastery of Vallombrosa, and attended the University of Pisa (with the original intention of studying medicine). He became a professor of mathematics at Pisa in 1589, and at Padua in 1592. He is famous for his interest in astronomy and physics, including hydrostatics, and dynamics (through his study of properties relating to gravitation). Works published in 1632 include support for the Copernican system. Although the publication was approved by the papal censor, it did (to some degree) contradict an edict in 1616, declaring the proposition that the sun was the centre of the solar system to be false. After an inquiry, Galileo was placed under house arrest, and died near Florence in 1642.

2. Review of Sopra Le Scoperte dei Dadi

Galileo's brief research summary on dice is believed to have been written between 1613 and 1623[2]. It is a response to a request for an explanation about an observation concerning the playing of three dice. While the possible combinations of dice sides totalling 9, 10, 11, and 12 are the same, in Galileo's words:

"…it is known that long observation has made dice-players consider 10 and 11 to be more advantageous than 9 and 12."

Galileo notes in the opening paragraph of his article:

"The fact that in a dice-game certain numbers are more advantageous than others has a very obvious reason, i.e. that some are more easily and more frequently made than others…"

Galileo explains the phenomenon by enumerating the possible combinations of the three numbers composing the sum, and presents a tabular summary. The principles allowing the enumeration are explained:

"...we have so far declared these three fundamental points; first, that the triples, that is the sum of three-dice throws, which are made up of three equal numbers, can only be produced in one way; second, that the triples which are made up of two equal numbers and the third different, are produced in three ways; third, that those triples which are made up of three different numbers are produced in six ways. From these fundamental points we can easily deduce in how many ways, or rather in how many different throws, all the numbers of the three dice may be formed, which will easily be understood from the following table:"

10		9		8		7		6		5		4		3	
631	**6**	621	**6**	611	**3**	511	**3**	411	**3**	311	**3**	211	**3**	111	**1**
622	**3**	531	**6**	521	**6**	421	**6**	321	**6**	221	**3**				
541	**6**	522	**3**	431	**6**	331	**3**	222	**1**						
532	**6**	441	**3**	422	**3**	322	**3**								
442	**3**	432	**6**	332	**3**										
433	**3**	333	**1**												
	27		**25**		**21**		**15**		**10**		**6**		**3**		**1**

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.

2 Refer to F.N. David's *Games, Gods and Gambling – A History of Probability and Statistical Ideas*, Dover

The top row of the table presents the sum of the three dice. Galileo does not provide the enumeration for sums 11 to 18, indicating earlier in his article that an investigation from 3 to 10 is sufficient because:

"…what pertains to one of these numbers, will also pertain to that which is the one immediately greater."

While the wording is awkward, he is referring to the symmetrical nature of the problem, however does not provide any more explanation.

The possible triples are shown under the sums, and to the right of each is the number of combinations for the triple. The last row sums those combinations.

From Galileo's table, it can be seen that 10 will show up in 27 ways out of all possible throws (which Galileo does indicate as 216). Since 9 can be found in 25 ways, this explains why it is at a "disadvantage" to 10 (even though each sum can be made from 6 different triples).

The article is of interest for its antiquity in the development of ideas relating to the science of probability. Although words like "chance" and "probability" are not directly used, the idea is conveyed by the application of terms such as "advantage" or "disadvantage". Combinatorial mathematics, and an appreciation for the equipossibility of individual events (gained either by a recognition of the symmetry of the die, or observation of results) form the building material for early probability science.

Publications 1962, page 62.

Review[1] of *Correspondence* between Pierre de Fermat and Blaise Pascal

1. Biographical Notes

Pierre de Fermat was born in 1601 at Beaumont-de-Lomagne, studied law at the University of Toulouse, and served there as a judge. He appears to have corresponded a great deal with scientists in Paris, as well as with others, including Pascal, about mathematical ideas. His interests included the theory of numbers, and is well known for the proposition that the equaton $x^n + y^n = z^n$ has no solutions in the positive integers (>2). He died at Castres in 1665.

Blaise Pascal was born at Clermont in 1623, and died in Paris in 1662. In addition to his contribution, along with Fermat, to the science of probability, he is well known for his work in geometry and hydrostatics. Pascal wrote the *Essai pour les Coniques*, and invented (and sold) a mechanical calculating machine. He may be most famous for his philosophical and religious writings, and is the author of the *Pensees*.

2. Review of the correspondence

This summary is primarily concerned with ideas presented in the first two letters of a collection of correspondence written over a period from 1654 to 1660. The first letter in this series is from Fermat to Pascal, and is undated, although it was likely written in June or July of 1654 (based on the dates of subsequent correspondence from Pascal). It would seem that Pascal had earlier written to Fermat, discussing the problem relating to the division of stakes in a wager on a game of dice, when the game is suspended before completion. The question appears to have been: If a player needs to get 1 point (a specific side of the die) in eight throws of the die, and after the first three throws has not obtained the required point, how much of the wager should be distributed to each player if they agree to discontinue play?

Fermat's letter suggests that Pascal reasoned 125/1,296 of the wager should be given to the player. Fermat disagrees with this, proposing that the player should receive 1/6 of the wager. Fermat's argument is surely based on the equal possibility of outcomes for points 1 to 6, due to the symmetry of the die.

Fermat distinguishes between an assessed value for a throw not taken, with subsequent continuation of the game, and the agreed completion of play before eight throws. His reasoning is as follows:

"If I try to make a certain score with a single die in eight throws…[and] we agree that I will not make the first throw; then, according to my theory, I must take in compensation 1/6th of the total sum…Whilst if we agree further that I will not make the second throw, I must, for compensation, get a sixth of the remainder which comes to 5/36th of the total sum…If, after this, we agree that I will not make the third throw, I must have…a sixth of the remaining sum which is 25/216th of the total…And if after that we agree that I will not make the fourth throw…I must again have a sixth of what is left, which is 125/1,296th of the total, and I agree with you that this is the value of the fourth throw, assuming that one has already settled for the previous throws."

The argument can be summarised in the following table:

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.

Throw	Proportion of Wager distributed	Remainder of original Wager
1	1/6	5/6
2	1/6 (5/6)	25/36
3	1/6 (25/36)	125/216
4	1/6 (125/216)	625/1,296
5	1/6 (625/1296)	3125/7,776
6	1/6 (3125/7776)	15,625/46,656
7	1/6 (15625/46656)	78,125/279,936
8	1/6 (78125/279936)	390,625/1,679,616
Accumulated Totals	1,288,991/1,679,616 (0.77)	390,625/1,679,616 (0.23)

In Fermat's theory, the player always has a chance of obtaining the whole wager (a chance at least proportional to the ratio of the one number needed and the six sides of the die – that is, a one in six chance). The table above shows the sequence of probabilities associated with the occurrence of the needed point on the last throw (1/6, 5/6·1/6, 5/6·5/6·1/6 etc.). These are the probabilities associated with the negative binomial distribution, having the required point occur on the last of a sequence of 1,2,…,8 throws.

In his response letter, dated July 29, 1654, Pascal agrees with Fermat's reasoning, and presents solutions to two specific cases of the problem of points:

1) The case involving a player needing one more point.
2) The case in which a player has acquired the first point.

For the first case, Pascal uses a recursive process to illustrate the solution. He provides the example of two players wagering 32 pistoles (gold coins of various denominations) each, and begins by considering a dice game in which three points are needed. The players' numbers have equal chances of turning up. The following table illustrates Pascal's argument: (the ordered pair notation (*a*,*b*) refers to the "state" of the game at some stage, with player A having thrown *a* points, and player B, *b* points; the pair {c,d} refers to the division of the wager).

State of Game	Division if player A's number turns up next	Division if player B's number turns up next	Division if players agree to suspend the game
(2,1)	{64,0}	{32,32}	{48,16}
(2,0)	{64,0}	{48,16}* *this corresponds to the state (2,1) shown above	{56,8}
(1,0)	{56,8}** **this corresponds to the state (2,0) shown above	{32,32}	{44,20}

The values distributed upon suspension of the game, conform to the expected values.

The argument is an (early) example of the application of a "minimax" principle. Both players wish to maximize the amount they would receive, and minimize their losses. The "motivation" is illustrated by the following "payoff matrix", with the expected proceeds for player A indicated under the relevant circumstances.

		Player B		
		Agreement to settle the wager	No agreement to settle the wager	Row Minimum
Player A	Rolls a favourable number	48	64	48
	Does not roll a favourable number	48	32	32
	Column Maximum	48	64	

Player A would like to maximize the row minimums, while player B wishes to minimize the column maximums. Both would settle on 48 (for player A).

After discussing his theory relating to the equitable distribution of the wager amount, Pascal presents the following rule :

"...the value (by which I mean only the value of the opponent's money) of the last game of two is double that of the last game of three and four times the last game of four and eight times the last game of five, etc."

Using the recursive procedure applied by Pascal in the case of a game of *three*, with a game of *four*, we would proceed as follows:

State of Game	Division if player A's number turns up next	Division if player B's number turns up next	Division if players agree to suspend the game
(3,2)	{64,0}	{32,32}	{48,16}
(3,1)	{64,0}	{48,16}* *this corresponds to the state (3,2) shown above	{56,8}
(3,0)	{64,0}	{56,8}** **this corresponds to the state (3,1) shown above	{60,4}

Using Pascal's terminology, the value of the last game of *four* is four.

The rule for distributing a wager of 2W (each player providing W), when one of the players requires one more point, is then $2W-W/2^n$, where n represents the number of points needed for the game (before play commences).

Pascal suggests that the solution to the second class of problems is more complicated:

"…the proportion for the first game is not so easy to find…[it] can be shown, but with a great deal of trouble, by combinatorial methods…and I have not been able to demonstrate it by this other method which I have just explained to you but only by combinations."

A rule is presented, without detailing a proof:

"*Let the given number of games be, for example, 8. Take the first eight even numbers and the first eight odd numbers thus:*

2, 4, 6, 8, 10, 12, 14, 16

and 1, 3, 5, 7, 9, 11, 13, 15.

Multiply the even numbers in the following way: the first by the second, the product by the third, the product by the fourth etc.; multiply the odd numbers in the same way…
The last product of the even numbers is the denominator and the last product of the odd numbers is the numerator of the fraction which expresses the value of the first one of eight games…"

If each player wagers W, then the distribution of the wager after the first throw would be

$$W + W(1/2 \cdot 3/4 \cdot 5/6 \cdot \ldots \cdot (2n-1)/2n)$$

Where n is the number of points required (after getting the first point).

How can this formula be shown to be reasonable, using elementary principles of probability? One approach is to examine the possible ways for completing various games after one player acquires the first point. The tree diagrams in Exhibit 1 (last page) illustrate three cases.

Player A has one point and needs in case a) one more point; b) two more points and c) three more points. Each connecting line between possible states of the game represents an event with probability 1/2, the probability of going from one state to the next. Using the diagrams, the probabilities for player A obtaining the required points may be assessed. For example, in a) there is a 1/2 probability of going from (1,0) to (2,0), and a $1/2 \cdot 1/2 = 1/4$ probability of going from (1,0) to (1,1) to (2,1). The probability for player A getting two points (given one point) is then 1/2+1/4 = 3/4. Similarly, for case b) the probability is 11/16 and for case c) 21/32.

If these probabilities are used to obtain the expected values for player A, we have:

In case a) $3/4 \cdot 2W = (2/4 + 1/4)2W = W + W(1/2)$

In case b) $11/16 \cdot 2W = (8/16 + 3/16)2W = W + 3/8 \cdot W = W + W(1/2 \cdot 3/4)$

In case c) $21/32 \cdot 2W = (16/32 + 5/32)2W = W + 5/16 \cdot W = W + W(1/2 \cdot 3/4 \cdot 5/6)$

The above sequence of expected values conforms to Pascal's rule.

To illustrate, Pascal considers a game in which a player has obtained 1 point and needs 4 more. He notes that at most 8 plays would be required to complete the game (either player A throws 4 more points, or player B will throw the required 5). He observes that 1/2 of the number of combinations of 4 from 8, divided by a sum consisting of this same value, plus the combinations of 5,6,7 and 8 from 8, gives the same proportion as $1/2 \cdot 3/4 \cdot 5/6 \cdot 7/8 = 35/128$.

This is the case, since in general:

$1/2 \cdot 3/4 \cdot 5/6 \cdot \ldots \cdot (2n-1)/(2n) = (2n-1)!/n!(n-1)! \cdot 1/2^{2n-1}$, with:

$$(2n-1)!/n!(n-1)! = 1/2 \cdot (2n!/n!n!), \text{ and}$$

$$2^{2n-1} = 1/2 \cdot (1+1)^{2n}, \text{ and}$$

$$1/2 \cdot (1+1)^{2n} = 1/2 \cdot \sum_{i=0}^{2n} \binom{2n}{i}$$

In the July 29th letter, Pascal also provides two tables indicating a division of wagers for games of dice suspended at different stages. The tables are not accompanied with detailed explanations. Pascal also relates the observations, and questions of Monsieur de Mere, relating to a game of dice requiring (at least) one six to turn up in 4 throws. The odds given in favour of this event are 671 to 625. Again, the computations are not provided, however, they correspond to the probability given by:

$$\sum_{i=1}^{4} \binom{4}{i} (1/6)^{i} (5/6)^{4-i}$$

Also, it is noted that there is a "disadvantage" in throwing two sixes in 24 such plays. Using the above formula, summing the combinations of 1,2,…24 out of 24, with associated probabilities 1/36 and 35/36 (to the appropriate exponents) the probability for the event can be shown to be about 0.49. That Monsieur de Mere noticed in practice this "disadvantage" is remarkable (he must have observed, and / or played, many such games).

The remaining correspondence includes an interesting dispute over the interpretation of combinations of events, used as a means for computing equitable settlements for wagers (establishing the proportion of funds to be distributed).

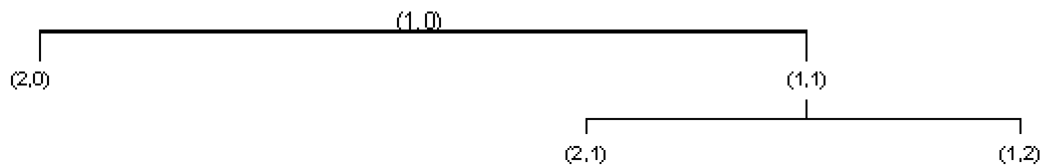
However, for our purposes, at this stage, it is sufficient to appreciate that combinatorial methods, and the identification of equipossible events, are the cornerstones for the emerging theory of probability, with early applications for binomial expansions. Instead of using the terms "chance" or "probability", our correspondents used words such as "favour" or "advantage" and "disadvantage", which convey the same meaning, in the context of a gambling environment. We also noted the early decision theory motivation, and its influence on what we will later call expected value.

A survey of the literature does give the impression that Fermat and Pascal, got the "die rolling" for the mathematical development of the science of probability.

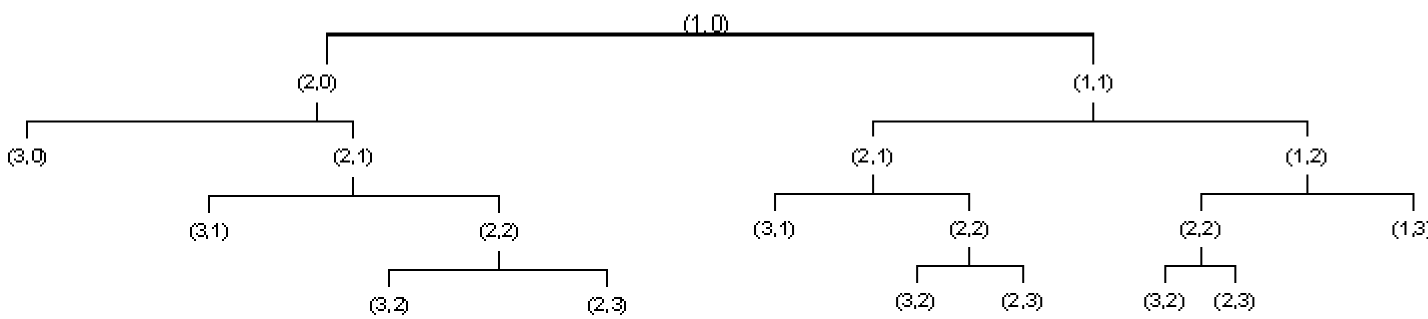
Exhibit1

Tree graphs of possible ways for completing two player games of dice after one player has acquired the first point

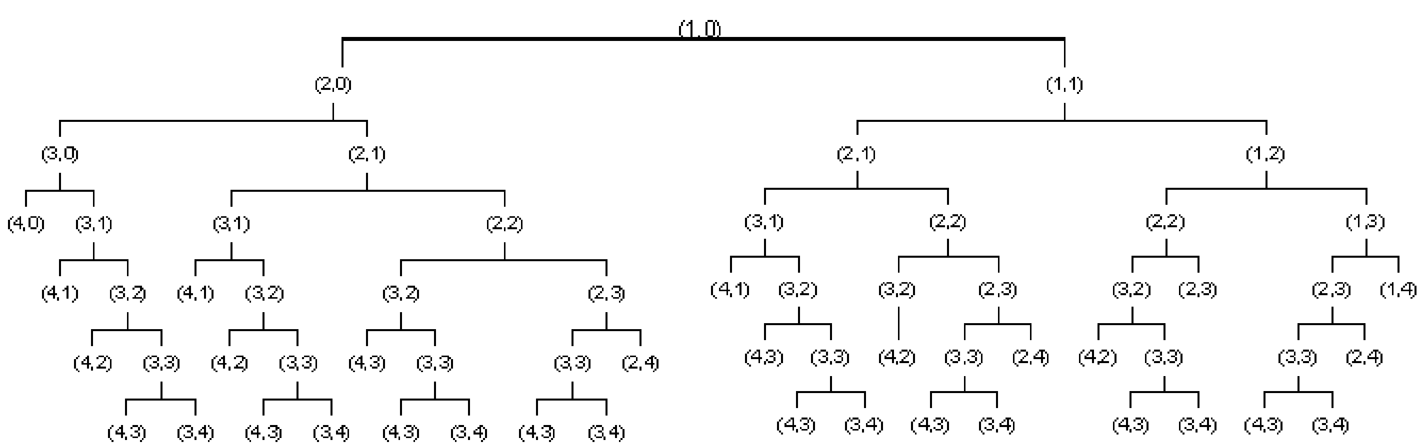
a) Game requiring two points



b) Game requiring three points



c) Game requiring four points



Review[1] of Christiaan Huygen's *De Ratiociniis in Ludo Aleae* (On Reasoning or Computing in Games of Chance)

1. Biographical Notes

Christiaan Huygens was born at The Hague, Netherlands, in 1629. He studied mathematics and law at the University of Leiden, and at the College of Orange in Breda. His father was a diplomat, and it would have been the normal practice for Huygens to follow in that vocation. However, he was more interested in the natural sciences, and with support from his father he was able to conduct studies and research in mathematics and physics. He is well known for his work relating to the manufacturing of lenses, which improved the quality of telescopes and microscopes. He discovered Titan, identified the rings of Saturn, and invented the first pendulum clock. He resided in Paris for some time, and made the acquaintance of persons familiar with Fermat and Pascal, and with their correspondence relating to "the problem of points" and similar concepts concerning games of chance. It is believed that Huygens died at The Hague, in 1695.

2. Review of the *De Ratiociniis in Ludo Aleae*

On Reasoning in Games of Chance, is cited in the literature as the first published mathematical treatise on the subject of probability[2]. The work was first printed in 1657, before the earlier correspondence between Fermat and Pascal was published, although clearly influenced by the content of those letters. The present review uses an English translation printed in 1714 by S. Keimer, at Fleetstreet, London.

The treatise is composed of a brief introduction, entitled *The Value of Chances*; the statement of a fundamental postulate; 14 propositions, a corollary and a set of five problems for the reader to consider.

The development of the theory is very systematic. Introducing the subject, Huygen's writes:

"Although in games depending entirely upon Fortune, the Success is always uncertain; yet it may be exactly determined at the same time how much more probability there is that [one] *should lose than win"*

Games of chance have outcomes that are (generally) unpredictable. At the same time, Huygens claims that it is possible to make meaningful statements, or measurements, relating to those systems. While the concept of probability, perhaps even the word itself[3], is observed within our earlier readings, Huygens' association of the phenomena (games of chance) with a relative measure of chance, is comparable to a modern treatment of the theory by first defining a random system or process, and the concept of probability. Having defined the system and the measure, Huygen's states his fundamental principle:

"As a Foundation to the following Proposition, I shall take Leave to lay down this Self-evident Truth: That any one Chance or Expectation to win any thing is worth just such a Sum, as would procure in the same Chance and Expectation at a fair Lay [or wager]*"*.

The wording is somewhat difficult follow, however he gives an example, from which it is evident that the

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.
2 See for example, Ian Hacking's *The Emergence of Probability* (Cambridge University Press, 1975), page 92 or William S. Peters' *Counting for Something – Statistical Principles and Personalities* (Springer – Verlag, 1987), page 39.
3 Refer to Gerolamo Cardano's De Ludo Aleae.

"Expectation" or value of a wager, is the mean of the possible proceeds:

"If any one should put 3 Shillings in one Hand, without telling me [which hand it is in]*, and 7 in the other, and give me Choice of either of them; I say, it is the same thing as if he should give me 5 Shillings."*

Although the word "expectation" (the Latin "expectatio" was used in the original work)[1] is used to refer to the value of a wager, its meaning in this example does correspond to its use in modern probability theory.

The first proposition states:

*"If I expect **a** or **b**, and have an equal chance of gaining either of them, my Expectation is worth (**a**+**b**)/2."*

The expectation is the fair value for a wager. How can this value be calculated in a game where the prizes are received with equal chance? Huygens reasons as follows: Suppose there is a lottery with two players, and each player buys a ticket for **x**, and that it is agreed that the proceeds are **a** and 2**x-a**, then each player could just as easily receive **a** or 2**x-a**. Setting 2**x-a** = **b**, it follows that the value of the lottery ticket is **x** =(**a**+**b**)/2.

The second proposition extends the first to the case of three prizes, **a**,**b** and **c**, such that **x** = (**a**+**b**+**c**)/3 is the value of the expectation; then in the same manner to four prizes, and so on.

Proposition III states:

*"If the number of Chances I have to gain **a**, be **p**, and the number of Chances I have to gain **b**, be **q**. Supposing the chances equal; my Expectation will then be worth **ap**+**bq** / **p**+**q**."*

This proposition generalizes the expectation to lotteries with prizes having different chances of being distributed. Huygens gives the following example:

"If I have 3 Expectations of 13 and 2 Expectations of 8, the value of my Expectation would by this rule be 11."

Propositions IV to IX illustrate solutions to "the problem of points", in a manner analogous to the reasoning of Pascal, although Huygens is looking more generally at the problem, not using any specific type of game as an example. Beginning with simple cases, Huygens solves more complicated problems, suggesting:

"The best way will be to begin with the most easy Cases of the Kind."

To appreciate the form of Huygens' arguments, consider Proposition VII:

"*Suppose I want two Games, and my Adversary four.*
*Therefore it will either fall out, that by winning the next Game, I shall want but one more, and he four, or by losing it I shall want two, and he shall want three. So that by Schol Prop.5. and Prop.6., I shall have an equal Chance for 15/16**a** or 11/16**a**, which, by Prop.1 is just worth 13/16**a**.*"

---

1 See Ian Hacking, page 95.

Propositions 5 and 6 explain the expectations when one player needs 1 game, and the other 4 games, and when one player needs 2 and the other 3 games. Then applying Proposition 1, the expectations are effectively averaged to provide the relevant value. A corollary is then given:

*"From whence it appears, that he who is to get two Games, before another shall get four, has a better Chance than he is to get one, before another gets two Games. For in this last Case, namely of 1 to 2 his Share by Prop. 4 is but 3/4**a**, which is less than 13/16**a**."*

This corollary compares the probabilities relevant to the two cases, using the expectations. This form of comparison has been made by earlier writers, using the "odds" approach[1].

In Poposition IX, Huygens provides a table showing the relative chances for three players in various game states.

Propositions IX to XIV present solutions for problems relating to games of dice. The solutions rely upon the earlier propositions, especially Proposition 3. As an example, Proposition X relates to the rolling of a single die:

"To find how many Throws one may undertake to throw the Number 6 with a single Die."

Huygens reasons that for the simplest case, one throw, there is 1 chance to get a six, receiving the wager proceeds **a**, and 5 chances to receive nothing, so that by Proposition 3, the expectation is (1·**a** + 5·**0**) / (1+5) = 1/6**a**. To compute the expectation for 1 six in two throws, it is noted that if the six turns up on the first die, the expectation will again be **a**. If not, then referring to the simplest case, there is an expectation of 1/6**a**. Using Proposition 3, there is one way to receive **a**, and 5 ways to receive the 1/6**a** (the sides 1 to 5 on the die):

$$(1 \cdot \mathbf{a} + 5 \cdot (1/6\mathbf{a})) / (1+5) = 11/36\mathbf{a}$$

This corresponds to the six appearing on the first throw with probability 1/6, or on the second throw with probability (5/6) · (1/6). In Huygens system, expectations for simpler cases are combined using Proposition 3, to solve more complex problems.

Proposition XIV uses two linear equations to derive the ratio of expected values for two players in the following case:

"If my self and another play by turns with a pair of Dice upon these Terms, That I shall win if I throw the Number 7, or he if he throw 6 soonest, and he to have the Advantage of first Throw: To find the Proportion of our Chances."

With total proceeds set at **a**, Huygens uses **a-x** to represent the expectation of the player to throw first, and **x** for the second player. When it is the first player's turn, the second player's expectation is **x**. Huygens reasons that when the second player is to throw, the expectation must be higher (it is conditioned on the first player not throwing a 6). He refers to this expectation as **y**. On the first player's turn, the second player's expectation (using proposition 3) will be (5·0 + 31**y**)/36 = 31/36**y** (as there are 5 ways for the first player to get 6). It follows that 31/36**y** = **x** (or **x**=36/31**y**). On the second player's turn the expectation is (6**a** + 30**x**) / 36 (there are 6 ways to roll 7) which equals **y**. Then:

1 Odds are given in the text by Cardano and in the correspondence between Fermat and Pascal.

$$(6\mathbf{a} + 30\mathbf{x}) / 36 = 36/31\mathbf{x} ,$$

with solution $\mathbf{x} = 31/61\mathbf{a}$. The ratio of the expected values $\mathbf{x}$ to $\mathbf{a-x}$ is then 31:30.

The text concludes with a set of five problems for the reader to solve (also the practice in most contemporary texts in mathematics).

Review[1] of Dr. John Arbuthnott's An Argument for Divine Providence, taken from the constant Regularity observed in the Births of both Sexes

1. Biographical Notes

John Arbuthnott was born in 1667 at Kincardineshire, Scotland. The following passage from *Annotated Readings in the History of Statistics*, by H.A. David and A.W.F Edwards, provides interesting biographical information:

> *John Arbuthnott, physician to Queen Anne, friend of Jonathan Swift and Isaac Newton…was no stranger to probability…In 1692 he had published (anonymously) a translation of Huygens' De ratiociniis in ludo aleae (1657) as Of the Laws of Chance, adding "I believe the Calculation of the Quantity of Probability might be improved to a very useful and pleasant Speculation, and applied to a great many events which are accidental, besides those of Games." There exists a 1694 manuscript of Arbuthnott's which foreshadows his 1710 paper [An Argument for Divine Providence].*

Dr. Arbuthnott died at London in 1735.

2. Review of *An Argument for Divine Providence*

Dr. Arbuthnott's paper is the first in our series of readings to apply the developing theory of probability to phenomena other than games of chance. Earlier, Pascal did use probabilistic reasoning in his article *"The Wager"* to advocate a life of faith, in an "age of reason". Interestingly, Arbuthnott's subject is related to Pascal's.

There are two principal arguments made in the article:

1) It is not by chance that the number of male births is about the same as the number of female births.
2) It is not by chance that there are more males born than females, and in a constant proportion.

To support the first proposition, application is made of the binomial expansion relating to a die with two sides marked M (male) and F (female). Essentially, $(M+F)^n$ is a model for the possible combinations of male and female children born. Arbuthnott observes that as n increases, the binomial coefficient associated with the term having identical numbers of M and F, becomes small compared to the sum of the other terms.

"It is visible from what has been said, that with a very great number of Dice…(supposing M to denote Male and F Female) that in the vast number of Mortals, there would be but a small part of all the possible Chances for its happening at any assignable time, an equal Number of Males and Females should be born."

Arbuthnott is aware that in reality there is variation between the number of males and females:

"It is indeed to be confessed that this Equality of Males and Females is not Mathematical but Physical,

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.

which alters much of the foregoing Calculation; for in this Case [the number of male and female terms] *…will lean to one side or the other."*

However, he writes:

"But it is very improbable (if mere Chance governed) that they would never reach as far as the Extremities…"

While it would be possible to have large differences in the numbers of males and females (with binomially distributed data having probability 1/2), the probability of this becomes very small when n is large. Contrary to Arbuthnott's argument, it could be reasoned that chance would account for the approximate equality in numbers of males and females.

The second proposition discounts chance as the cause for the larger number of male births observed annually. The form of the argument is interesting, because it is similar to a test of significance. Arbuthnott states the *Problem*:

"A lays against B, that every Year there shall be born more Males than Females: To find A's Lot, or the Value of his Expectation."

A hypothesis is being made in the form of a wager. Arbuthnott notes that the probability that there are more males than females born must be less than 1 /2 (assuming that there is an equal chance for a male or female birth). For this "test" however, he sets the chance at 1 /2 (which would result in a higher probability), and notes that for the number of males to be larger than the number of females in 82 consecutive years (for which he has data on christenings), the lot would be $1/2^{82}$. The lot would be even less if the numbers were to be in "constant proportion". Since the data do not support B (in every year from 1629 to 1710, male christenings outnumber female christenings), Arbutnott reasons:

"From whence it follows, that it is Art, not Chance, that governs."

The hypothesis of equal probability is rejected, and Arbuthnott attributes the observed proportions to Divine Providence.

The second argument has been referred to as the first published test of significance[1].

1 Ian Hacking, *The Emergence of Probability*, page 168.

Review[1] of Pierre Remond de Montmort's *On the Game of Thirteen*

1. Biographical Notes

With reference to Isaac Todhunter's text, *A History of the Mathematical Theory of Probability*2, Pierre Remond de Montmort devoted himself to religion, philosophy and mathematics. He served in the capacity of cathedral canon at Notre-Dame in Paris, from which he resigned, in order to marry. In 1708 he published his treatise on "chances", *Essai d'Analyse sur les Jeux de Hazards*. L.E Maistrov, in his text, *Probability Theory – A Historical Sketch*3, provides the following biographical information:

> *"Pierre Remond de Montmort (1678-1719) was a French mathematician as well as a student of philosophy and religion. He was in correspondence with a number of prominent mathematicians (N. Bernoulli, J. Bernoulli, Leibniz, etc.) and was a well-established and authoritative member of the scientific community. In particular, Leibniz selected him as his representative at the commission set up by the Royal Society to rule on the controversy between Newton and Leibniz concerning priority in the discovery of differential and integral calculus…His basic work on probability theory was the "Essai d'Analyse sur les Jeux de Hazard". It went through two editions, the first of which was printed in Paris in 1708…the second…appeared in 1713, although Todhunter claims* [1714]. *The first part contains the theory of combinatorics; the second discusses certain games of chance with cards; the third deals with games of chance with dice, the fourth part contains the solution of various problems including the five problems proposed by Huygens…"*

2. Review of the article

The present review concerns an article in the second part of Montmort's *Essai d'Analyse*, entitled "On the Game of Thirteen". An English translation of that article is available in *Annotated Readings in the History of Statistics* by H.A. Davids and A.W.F. Edwards.4

In the first section of the article, Montmort provides a description of the play of Thirteen:

> *"The players first draw a card to determine the banker. Let us suppose that this is Peter and that the number of players is as desired. From a complete pack of fifty-two cards, judged adequately shuffled, Peter draws one after the other, calling one as he draws the first card, two as he draws the second, three as he draws the third, and so on, until the*

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.
2 *A History of the Mathematical Theory of Probability*, page 78, by Isaac Todhunter, Chelsea Publishing Co., New York (1965 unaltered reprint of the First Edition, Cambridge 1865).
3 *Probability Theory – A Historical Sketch*, page 76, by Leonid E. Maistrov, Academic Press Inc., New York 1974.
4 *Annotated Readings in the History of Statistics* , Springer-Verlag, New York 2001.

thirteenth, calling king. Then, if in this entire sequence of cards he has not drawn any with the rank he has called, he pays what each of the players has staked and yields to the player on his right. But if in the sequence of thirteen cards, he happens to draw the card he calls, for example, drawing an ace as he calls one, or a two as he calls two, or a three as he calls three, and so on, then he takes all the stakes and begins again as before, calling one, then two, and so on…"

Provision is made in the rules for a new deck of cards should the dealer use all the cards in the first set.

The game of Thirteen provides an early example of a problem relating to coincidences or matches. Montmort describes a method for computing the chance or expectation of drawing a card matching the number called by the banker. Since the time of Cardano, questions relating to expectation have been solved by enumerating the favourable cases and the total possible events. Montmort observes:

"*Let the cards with which Peter plays be represented by a,b,c, d, etc… it must be noted that these letters do not always find their place in a manner useful to the banker. For example, a, b, c produces only one to the person with the cards although each of these three letters is in its place. Likewise, b, a, c, d produces only one win for Peter, although of the letters c and d is in its place. The difficulty of this problem is in disentangling how many times each letter is in its place useful and how many times it is useless."*

To solve the problem, Montmort first considers a game with only two cards, an ace and a two. There is only one way in which the banker can receive the proceeds of the wager, an ace has to be the first card. Montmort then computes the expectation, essentially an application of Huygens' 3rd Proposition in De Ratiociniis. If the proceeds total A, then the banker's expectation is $(1 \cdot A + 1 \cdot 0) / 2 = 1/2$ A.

The next case considered is a game with three cards, represented by the letters a,b,c. Montmort observes that of the six possible combinations for the letters (representing the possible orders for dealing the cards), four are favourable to the banker (and two are not favourable):

"...there are two with a in first place; there is one with b in second place, a not having been in first place; and one where c is in third place, a not having been in first place and b not having been in second place."

It follows that the expectation is $(4 \cdot A + 2 \cdot 0) / 6 = 2/3$ A.

Similarly four and five card games are considered, indicating the expectations as 5/8 A and 19/30 A, respectively. The expectations for games with one to five cards allows Montmort to suggest a formula for computing the banker's expectation generally, in a recursive manner:

$[g(p-1)+d] / p$

where:

p is the number of cards;
g is the espectation when there are p-1 cards, and
d is the expectation when there are p-2 cards.

A table is given showing the expectations for games up to 13 cards. The banker's expectation for the game of Thirteen is presented as 109,339,663/172,972,800 A.

Montmort observes that the expectations can be expressed as series in the form:

1 - 1/(1.2) + 1/(1.2.3) - 1/(1.2.3.4) +...,

with alternating positive and negative terms, consisting of numerators 1 and denominators 1…(p-2)(p-1)p, where p is the number of cards. The rapid convergence of the expectations to a value between 5/8 and 19/30, does not appear to have been noticed. An examination of the recursive formula may have indicated that as the number of cards (p) in the game increases, the value identified as g (expectation) stabilizes, since (p-1)/p gets closer to 1, and d/p becomes small. Montmort is principally interested in explaining *how* the expectation is computed in the game of Thirteen, and describing the mathematical expressions from which the expectations can be computed.

In addition to explaining how the expectation is derived, Montmort provides a table showing the number of possible favourable deals from specific cards in a game. For example, in a game with five cards, there are 24 ways for an ace to be dealt as the banker calls one, 18 ways for the two to be dealt as two is called and so on. The table can be used for games having up to eight cards. Montmort completes the article with some commentary relating to patterns in that table.

Montmort's article adds to the variety of problems considered by probability science since the time of Cardano. Montmort has clearly used Huygens approach, and the principle stated in Proposition IV of De Ratiociniis:

> *"…the best way will be to begin with the most easy Cases of the kind."*

Review[1] of James Bernoulli's *Theorem*…From *Artis Conjectandi*

1. Biographical Notes

According to W.W.R. Ball in *A Short Account of the History of Mathematics2:*

> *"Jacob or James Bernoulli was born at Bâle on December 27, 1654; in 1687 he was appointed chair of mathematics in the university there; and occupied it until his death on August 16, 1705…In his Artis Conjectandi, published in 1713, he established the fundamental principles of the calculus of probabilities…His higher lectures were mostly on the theory of series…"*

2. Review of the article

In Chapter IV of Part IV of his text on probability, *Artis Conjectandi* (published in 1713), James Bernoulli writes:

> *"Something further must be contemplated here which perhaps no one has thought of about till now. It certainly remains to be inquired whether after the number of observations has been increased, the probability is increased of attaining the true ratio between the numbers of cases in which some event can happen and in which it cannot happen, so that the probability finally exceeds any given degree of certainty…"3*

This review summarizes Bernoulli's proposed solution to this inquiry largely in his own words, as presented in Chapter V of Part IV of *Artis Conjectandi*.

For the purpose of this summary, an abridged English translation of a copy available in German4 has been prepared. The Latin, German and abridged English versions are attached for reference. It should be noted that in the German and English copies, some mathematical notation differs from the original, using instead a more contemporary system.

Bernoulli begins by presenting a set of increasingly complex lemmas which will be used to prove his proposition:

> <u>*Lemma 1*</u>
>
> *Given a set of natural numbers*

---

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.
2 Originally published by MacMillan & Co. Ltd., London 1912, pages 366 and 367.
3 Translation by Bing Sung (1966). *Translations from James Bernoulli*. Department of Statistics, Harvard University, Cambridge, Massachusetts.
4 Electronic Research Archive for Mathematics: (Ostwald's Klassiker d. exact. Wissensch. No. 107 u. 108.) *Published:* (1899).

0, 1, 2, ..., r-1, r, r+1, ..., r+s

...

continued such that the last member is a multiple of r+s, for example nr+ns, the new set is:

0, 1, 2, ..., nr-n, ..., nr, ..., nr+n, ..., nr+ns.

With increasing n, the terms between nr and nr+n or nr-n, similarly the terms between nr+n, nr-n or nr+ns and 0 increase. No matter how large n is, the terms greater than nr+n will not exceed s-1 times the number of terms between nr and nr+n. The number of terms below nr-n will not exceed r-1 times the terms between nr-n and nr.

Lemma 2

If r+s is raised to an exponent, then the expansion will have one more term than the exponent.

Lemma 3

In the expansion of the binomial r+s with exponent an integral multiple of r+s=t, for example n(r+s)=nt, then first, there is a term M, the largest value of the terms, if the number of terms before and after M are in the proportion s to r ... the closer terms to M on the left or right are larger than the more distant terms. Secondly, the ratio of M to a term closer to it is smaller than the ratio of that closer term to one more distant, provided the number of intermediate terms is the same.

Lemma 4

In the expansion of a binomial with exponent nt, n can be made so large that the ratio of the largest term M to other terms L_n and R_n which are n terms to the left or right from M, can be made arbitrarily large.

The proofs for Lemmas 3 and 4 are detailed in L.E. Maistrov's book Probability Theory – A Historical Sketch1. *In both cases binomial expansions of the terms are divided, and algebraic simplification presents the required limiting results.*

Lemma 5

In the expansion of a binomial with exponent nt, n may be selected so large that the ratio of the sum of all terms from the largest term M to the terms L_n and R_n to the sum of the remaining terms, may be made arbitrarily large.

1 *Probability Theory – A Historical Sketch*, pages 72 and 73, by Leonid E. Maistrov, Academic Press Inc., New York 1974.

Proof of Lemma5

According to Lemma 4, as n becomes infinitely large, M / L_n becomes infinite, then the ratios L_1 / L_{n+1}, L_2 / L_{n+2}, L_3 / L_{n+3} become all the more infinite. So it then follows that:

$$\frac{L_1 + L_2 + L_3 + ... + L_n}{L_{n+1} + L_{n+2} + L_{n+3} + ... + L_{2n}} = \infty$$

that is, the sum of the terms between M and L_n is infinitely greater than the sum of the terms left of L_n. Since by Lemma 1 the number of terms left of L_n exceeds the terms between L_n and M by only (s-1) times (that is, a finite number of times), and then from Lemma 3 the terms become smaller more distant from L_n, then (the sum of) all the terms between L_n and M (even if M is not included) will be infinitely larger than (the sum) left of L_n. In the same way ... [for the right side]

The Proposition is then stated as:

Let the number of favourable cases to the number of unfavourable cases be exactly or nearly r/s, therefore to all the cases as r/r+s = r/t - if r+s = t - this last ratio is between r+1/t and r-1/t. We can show, as many observations can be taken that it becomes more probable arbitrarily often (for example, c - times) that the ratio of favourable to all observations lies in the range with boundaries r+1/t and r-1/t.

Bernoulli observes that given a probability r/t for a favourable outcome, and a probability s/t for an unfavourable outcome, in nt trials (with t = r+s), the number of (possible) events with all favourable outcomes, all but one favourable outcomes, all but two favourable outcomes, etc. are

$$r^{nt}, \qquad \binom{nt}{1} r^{nt-1} s, \qquad \binom{nt}{2} r^{nt-2} s^2,$$

These correspond to the terms in the expansion of r+s, for which a number of useful properties were established in the lemmas. First, the number of trials with nr favourable outcomes and ns unfavourable outcomes is M. Next, the number of trials with at least nr-n and at most nr+n favourable outcomes is the sum of terms between the two limits L_n and R_n defined in Lemma 4. Bernoulli can then write:

Since the binomial exponent can be selected so large that the sum of terms which are between both bounds L_n and R_n is more than c times the sum of all the remaining terms outside of these bounds, (from Lemmas 4 and 5), it follows then that the number of observations can be taken so large that the number of trials in which the ratio of the number of favourable cases to all cases will not cross over the bounds (nr+n)/nt and (nr-n)/nt or (r+1)/t and (r-1)/t, is more than c times the remaining cases, that is, that it is more than c times probable that the ratio of the

> *number of favourable to all cases does not cross over the bounds (r+1)/t and (r*-1)/t.

James Bernoulli's solution is apparently the first proof of the Law of Large Numbers, which informally is stated as:

> *The law which states that the larger a sample, the nearer its mean is to that of the parent population from which the sample is drawn*.

Review[1] of Abraham de Moivre's *A Method of approximating the Sum of Terms of the Binomial $(a+b)^n$*…From *The Doctrine of Chances*

1. Biographical Notes

According to Isaac Todhunter in his text, *A History of the Mathematical Theory of Probability*2:

> *"Abraham de Moivre was born at Vitri, in Champagne, in 1667. On account of the revocation of the edict of Nantes3, in 1685, he took shelter in England, where he supported himself by giving instruction in mathematics and answers to questions relating to chances and annuities. He died at London in 1754…De Moivre was elected a Fellow of the Royal Society in 1697…It is recorded that Newton himself, in the later years of his life, used to reply to inquirers respecting mathematics in these words: 'Go to Mr. De Moivre, he knows these things better than I do'…"*

De Moivre is well known for the theorem:

$$[\cos(\theta) + i\sin(\theta)]^n = \cos(n\theta) + i\sin(n\theta)$$

2. Review of the article

This review relates to a supplementary article entitled *A Method of approximating the Sum of the Terms of the Binomial $(a+b)^n$ expanded into a Series, from whence are deduced some practical Rules to estimate the Degree of Assent which is to be given to Experiments* (referred to as the *Approximatio*), which appears in later editions (after 1733) of Abraham De Moivre's text on probabilities, *The Doctrine of Chances*, first published in 1718.

De Moivre's mathematical presentation in the *Approximatio* begins with a discussion relating to approximating the ratio of the middle term of the binomial (1+1) raised to very large n, to the sum of all terms (2^n). It is indicated that this approximation was developed several years earlier. As a result of contributions from James Stirling, it was found that the approximate ratio could be written as

$$2 \big/ \sqrt{nc}$$

where c is the circumference of a circle with radius equal to 1. The value of c is then 2π.

De Moivre next states:

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.
2 *A History of the Mathematical Theory of Probability*, page 78, by Isaac Todhunter, Chelsea Publishing Co., New York (1965 unaltered reprint of the First Edition, Cambridge 1865).
3 The Edict of Nantes was a proclamation by King Henry IV of France and Navarre, guaranteeing civil and religious rights to the Huguenots.

"…the Logarithm of the Ratio which the middle term of a high Power has to any Term distant from it by an Interval denoted l, would be denoted by a very near approximation, (supposing m = 1 / 2n) by the Quantities

(m+l-1/2) x log(m+l-1) + (m-l+1/2) x log(m-l+1)
- 2m x log m + log ((m+l)/m)."

De Moivre does not provide details on the derivation of the above formulae. Anders Hald describes their derivation in his *book A History of Probability and Statistics and Their Applications before 1750*[1].

He then presents Corollary 1:

"This being admitted, I conclude, that if m or 1/ 2n be a Quantity infinitely great, then the Logarithm of the Ratio, which a Term distant from the middle by the Interval l, has to the middle Term, is –2ll/n."

Again, the derivation is not shown. It follows noting that:

$(m+l-1/2)\log(m+l-1) + (m-l+1/2)\log(m-l+1) - 2m\log m + \log((m+l)/m)$

is equivalent to

$(m+l-1/2)\log(m+l-1) - (m+l-1/2)\log m + (m-l+1/2)\log(m-l+1) - (m-l+1/2)\log m + \log((m+l)/m).$

Then approximating log(m+l-1) by log(m+l) and log(m-l+1) by log(m-l), for large m, and re-writing the above terms using the properties of logarithms, we have

$(m+l-1/2)\log((m+l)/m) + (m-l+1/2)\log((m-l)/m) + \log((m+l)/m).$

Recalling that $\log(1+x) = x - x^2/2 + x^3/3 - \ldots$ when $-1 < x \leq 1$, then the above can be approximated, for *l* less than the square root of n, by

$(m+l-1/2)(l/m - l^2/2m^2) + (m-l+1/2)(-l/m - l^2/2m^2) + l/m - l^2/2m^2 = 2ll/n$
Then the approximation to the inverse ratio's logarithm is $-2ll/n$.

In Corollary 2, De Moivre notes that the number with "hyperbolic logarithm" (natural logarithm) $-2ll/n$ is:

$$1 - 2ll/n + 4l^2/2nn - 8l^6/6n^3 + \ldots$$

This is the series for $e^{-2ll/n}$, which then approximates the ratio of a term *l* terms distant from the middle term, to the middle term. If we represent the middle term by T_0, and terms 1, 2, …, *l* places distant by T_1, T_2, … T_l, then to this point De Moivre has obtained:

$$\frac{T_0}{2^n} \approx \frac{2}{\sqrt{2\pi n}}$$

---

1 Published by John Wiley & Sons, 1990, pages 473 to 476.

and

$$\log(\frac{T_l}{T_0}) \approx \frac{-2ll}{n}$$

Then $T_l = T_0 e^{-2ll/n}$, or as De Moivre would write:

$$T_l = T_0(1 - 2ll/n + 4l^2/2nn - 8l^6/6n^3 + ...)$$

Note that

$$T_0 \approx \frac{2}{\sqrt{2\pi n}} \cdot 2^n$$

for a binomial (1+1) to the exponent n very large. If we consider the binomial $(1/2+1/2)^n$, which is $1^n = 1$, the middle term would be

$$\binom{n}{n/2}(1/2)^n$$

Since in the binomial $(1+1)^n$, the middle term is

$$\binom{n}{n/2}$$

Then this is equivalent to T_0, and we can write

$$\frac{2}{\sqrt{2\pi n}} \cdot 2^n \cdot 2^{-n} = \frac{2}{\sqrt{2\pi n}}$$

for the middle term of the binomial $(1/2+1/2)^n$.

Using the previously defined symbols T_0, T_1,…, T_l, then sums of terms between the middle term and one l places distant can be obtained as

$T_0 + T_1 + T_2 + \ldots + T_l = T_0\,(1+\exp(-2 \cdot 1^2/n) + \exp(-2 \cdot 2^2/n)+\ldots+\exp(-2 \cdot l^2/n))$

This is essentially what De Moivre does in Corollary 2. Using the "hyperbolic logarithm" series expansions for each of the terms, he develops a sum of the binomial terms from the middle, to a term *l* places distant for the case of a binomial $(1/2 + 1/2)^n$:

$$\frac{2}{\sqrt{2\pi n}} \cdot (l - 2l^3/1 \cdot 3n + 4l^5/2 \cdot 5n^2 - ...)$$

Setting $l = s\sqrt{n}$, with s = 1/2, he gets

$$\frac{2}{\sqrt{2\pi n}} \cdot (1/2 - 1/3 \cdot 4 + 1/2 \cdot 5 \cdot 8 - ...)$$

Which he observes converges very quickly, and after using a few terms obtains the estimate 0.341344 for the sum of terms from the middle term to a term, which is about $1/2\sqrt{n}$ terms distant. Having obtained this result, De Moivre states in Corollary 3:

> *"And therefore, if it was possible to take an infinite number of Experiments, the Probability that an Event which has an equal number of Chances to happen or fail, shall neither appear more frequently than $1/2n + 1/2\sqrt{n}$ times, not more rarely than $1/2n - 1/2\sqrt{n}$ times, will be expressed by the double Sum of the number exhibited in the second Corollary, that is by 0.682688…"*

We have here, in 1733, a result about what would later be called a "normal distribution". In his book, *The Life and Times of the Central Limit Theorem1*, William Adams states:

> *"De Moivre did not name $\sqrt{n}/2$, which is what we would today call standard deviation within the context considered, but in Corollary 6 he referred to $\sqrt{n}$ as the Modulus by which we are to regulate our estimation."*

De Moivre generalizes the results for the binomial $(a+b)^n$ acquiring as William Adams indicates in modern notation

$$T_l = T_0 e^{-(a+b)^2 l^2 / 2abn}$$

using the symbols T_l and T_0 defined earlier.

De Moivre's intention was to develop a method for approximating binomial sums or probabilities when the number of trials was very large. He was able to do this with the aid of mathematics relating to series

1 Kaedmon Publishing Company, New York 1974, page 24.

expansions for logarithms and exponentials, as well as approximation methods for factorials. The approximation itself is a normal distribution. In addition to being an early occurrence of this distribution, for practical application, the approximation is an early illustration of a central limit theorem. For large n, the middle term (average value) is associated with a normal distribution, and this value can be limited by a measure related to $\sqrt{n}$.

Review[1] of Friedrich Robert Helmert's *The Calculation of the Probable Error from the Squares of the Adjusted Direct Observations of Equal Precision and Fechner's Formula2*

1. Biographical Notes

Friedrich Robert Helmert was born at Freiberg in 1841. He studied engineering sciences at the technical university in Dresden. He was as a lecturer and professor of geodesy at the technical university in Aachen, where he wrote "The Mathematical and Physical Theories of Higher Geodesy" (2 volumes, Leipzig 1884). In 1886 Helmert became the director of the Prussian Geodetic Institute and the International Earth Measurement Central Office, as well as professor at the university in Berlin. He died at Potsdam in 1917.

2. Review of *The Calculation of Probable Error…*

In this 1876 article by F.R. Helmert, there is apparently for the first time[3] a demonstration that, given $X_1,\ldots,X_n$ independent $N(\mu,\sigma^2)$ random variables, then

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$$

is distributed as (what would be called) a chi-square distribution with n-1 degrees of freedom. The present review is concerned primarily with how Helmert effectively shows this, although his purpose was to apply the result to the calculation of the mean squared error for $\hat{\sigma}$, in order to estimate "the probable error"[4].

To begin, Helmert considers $e_1,\ldots,e_n$, the "true errors" of a set of observations $X_1,\ldots,X_n$. The true errors are defined as $e_i = X_i - \mu$, where μ is the (true) mean for the population from which the X_i are observed. The joint probability "volume" (referred to as the "future probability"[5]) of the e_i, given that $Xi \sim N(\mu,\sigma^2)$ is presented as:

$$\left[\frac{h}{\sqrt{\pi}}\right]^n e^{-h^2[\mathrm{ee}]}d\mathrm{e}_1\cdots d\mathrm{e}_n$$

where h (referred to as "the precision") is equal to $1/\sigma\sqrt{2}$ and [ee] is the sum of squares of the e_i: $e^2_1+e^2_2+\ldots+e^2_n$.

---

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.
2 An abridged version of the article is found in *Annotated Readings in the History of Statistics* pages 109 to 113, by H.A. David and A.W.F. Edwards (Springer-Verlag, New York 2001). The section relating to Fechner's formula has been deleted.
3 Refer to *Annotated Readings in the History of Statistics* page 103, by H.A. David and A.W.F. Edwards (Springer-Verlag, New York 2001).
4 Defined as $\sigma\Phi^{-1}(0.75)$, where Φ is the standard normal distribution function. See *Annotated Readings,* page 103.
5 The likelihood function from the set of observations.

Helmert notes that [ee] is not known, since the parameter μ is not known (assuming that only sampling of the population is possible or feasible).

As the true mean can only be estimated by $\bar{X}$, then the true errors are estimated by X_i-$\bar{X}$, written λ_i by Helmert, and referred to as the "deviations" (from the arithmetic mean of the sample). Noting that $\lambda_1 + \lambda_2 + \ldots + \lambda_n = 0$, then $\lambda_n = -\lambda_1 - \lambda_2 - \ldots - \lambda_{n-1}$, and with $\bar{e} = \bar{X} - \mu$, the true errors are related to the deviations as:

$$e_1 = \lambda_1 + \bar{e},$$

$$e_2 = \lambda_2 + \bar{e}$$
$$e_{n-1} = \lambda_{n-1} + \bar{e}$$
$$e_n = -\lambda 1 - \ldots - \lambda_{n-1} + \bar{e}$$

Helmert identifies the following matrix with the above transformations:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 1 \\ & & & \ddots & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 \\ -1 & -1 & -1 & \cdots & & -1 & 1 \end{pmatrix}$$

which will be referred to as H. In matrix equation form, the transformation may be written:

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_{n-1} \\ e_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 1 \\ & & & \ddots & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 \\ -1 & -1 & -1 & \cdots & & -1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_{n-1} \\ \bar{e} \end{pmatrix}$$

The determinant of the matrix H is n. This can be shown using two properties relating to determinants:

1) If a matrix M is formed by adding a multiple of one column to another column in H, then the determinant of M equals that of H.
2) The determinant of a triangular matrix is equal to the product of diagonal elements.

New matrices can be formed by consecutively adding –1 times columns 1 to n-1, to the last column, resulting in:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ -1 & -1 & -1 & \cdots & & -1 & n \end{pmatrix}$$

The determinant is then $1 \cdot 1 \cdot \ldots \cdot n = n$.

Observing that the Jacobian of the transformation of variables is equivalent to the determinant of H, the joint probability for the change of variables becomes:

$$n\left[\frac{h}{\sqrt{\pi}}\right]^{n} e^{-h^2[\lambda\lambda]+h^2 n\bar{\mathrm{e}}^2} d\lambda_1 \cdots d\lambda_{n-1} d\bar{\mathrm{e}}$$

Helmert then notes that integrating the above expression over all possible values of ē results in the probability of the set $\lambda_1,\ldots,\lambda_n$:

$$\sqrt{n}\left[\frac{h}{\sqrt{\pi}}\right]^{n-1} e^{-h^2[\lambda\lambda]} d\lambda_1 \cdots d\lambda_{n-1}$$

Then

$$\sqrt{n}\left[\frac{h}{\sqrt{\pi}}\right]^{n-1} \int \cdots \int e^{-h^2[\lambda\lambda]} d\lambda_1 \cdots d\lambda_{n-1}$$

is the probability that $[\lambda\lambda]$ lies between values u and u+du.

Next, a transformation is devised for n-1 new variables t , i = 1,…,n-1, such that [tt] is equivalent to the sum of n-1 true errors. The transformation in matrix form, is given by:

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ \vdots \\ t_{n-1} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \cdots & \frac{\sqrt{2}}{2} \\ 0 & \sqrt{\frac{3}{2}} & \sqrt{\frac{3}{2}}\cdot\frac{1}{3} & \sqrt{\frac{3}{2}}\cdot\frac{1}{3} & \cdots & \sqrt{\frac{3}{2}}\cdot\frac{1}{3} \\ 0 & 0 & \sqrt{\frac{4}{3}} & \sqrt{\frac{4}{3}}\cdot\frac{1}{4} & \cdots & \sqrt{\frac{4}{3}}\cdot\frac{1}{4} \\ & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & \sqrt{\frac{n}{n-1}}\cdot\frac{1}{n} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_{n-1} \end{pmatrix}$$

To illustrate this transformation, consider the two variable case, using:

$$t_1 = \sqrt{2}(\lambda_1 + \sqrt{2}\frac{\lambda_2}{2})$$
$$t_2 = \sqrt{3/2}\,\lambda_2$$

Then $t_1^2 + t_2^2 = \lambda_1^2 + \lambda_2^2 + \lambda_1^2 + 2\lambda_1\lambda_2 + \lambda_2^2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$, since $\lambda_3 = -\lambda_1 - \lambda_2$.

The determinant of the transformation is √n, noting that the associated matrix is upper triangular, with product of the diagonal terms:

$$\begin{aligned} &\sqrt{2}\cdot\sqrt{\frac{3}{2}}\cdot\sqrt{\frac{4}{3}}\cdots\cdots\sqrt{\frac{n}{n-1}} \\ &= \sqrt{2}\cdot\frac{1}{\sqrt{2}}\cdot\sqrt{3}\cdot\frac{1}{\sqrt{3}}\cdots\cdots\sqrt{n-1}\cdot\frac{1}{\sqrt{n-1}}\cdot\sqrt{n} \\ &= \sqrt{n} \end{aligned}$$

Then the probability that [tt] is between u and u + du is given by

$$\left[\frac{h}{\sqrt{\pi}}\right]^{n-1}\int\cdots\int e^{-h^2[tt]}dt_1\cdots dt_{n-1}$$

Helmert then refers to a result he obtained in 1875: The probability that the sum [tt] of n-1 true errors equals u, is given by:

$$\frac{h^{n-1}}{\Gamma\left(\frac{n-1}{2}\right)}\cdot u^{\frac{n-3}{2}}\cdot e^{-h^2u}du$$

where again, h is “the precision”. Since [tt] = [λλ], the density applies to the sum of squares of n deviations. The above density is Gamma($^{n-1}\!/_2$,h) for variable hu. This can seen, recalling that if a random variable v ~ Gamma (α,β), then the probability density function can be written as:

$$\frac{\beta^{\alpha}}{\Gamma(\alpha)} v^{\alpha-1} e^{-\beta v}$$

Substituting hu for v, $\frac{n-1}{2}$ for α, and h for β:

$$= \frac{h^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \cdot (hu)^{\frac{n-3}{2}} \cdot e^{-h^2 u}$$

$$= \frac{h^{(\frac{n-1}{2}+\frac{n-3}{2})}}{\Gamma\left(\frac{n-1}{2}\right)} \cdot u^{\frac{n-3}{2}} \cdot e^{-h^2 u}$$

$$= \frac{h^{n-2}}{\Gamma\left(\frac{n-1}{2}\right)} \cdot u^{\frac{n-3}{2}} \cdot e^{-h^2 u}$$

Then the probability associated with volume d(hu) is:

$$= \frac{h^{n-2}}{\Gamma\left(\frac{n-1}{2}\right)} \cdot u^{\frac{n-3}{2}} \cdot e^{-h^2 u} d(hu)$$

$$= \frac{h^{n-1}}{\Gamma\left(\frac{n-1}{2}\right)} \cdot u^{\frac{n-3}{2}} \cdot e^{-h^2 u} du$$

Since hu ~ Gamma($^{n-1}\!/_2$,h), then u ~ Gamma($^{n-1}\!/_2$,h)$/$h. For h = $1/\sigma\sqrt{2}$, it can be shown that the probability density function for u is:

$$\frac{(\frac{u}{\sigma^2})^{\frac{n-1}{2}} \cdot e^{-\frac{1}{2}\frac{u}{\sigma^2}}}{2^{\frac{n-1}{2}-1}\Gamma\left(\frac{n-1}{2}\right)}$$

Then $\frac{u}{\sigma^2} \sim \chi^2_{n-1}$, so that u ~ $\frac{\chi^2_{n-1}}{\sigma^2}$. Recalling that u = [λλ], and that [λλ] = $\sum_{i=1}^{n}(X_i - \bar{X})^2$, we have the result:

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Helmert does not assign a name to the distribution of the sum of squares [λλ]. His objective is to use the result to estimate the probable error.

Also of interest in the article is the estimation related to the precision h (and thereby σ) in a maximum likelihood manner in section 2, such that:

$$\frac{1}{2h^2} = \sigma = \frac{[\lambda\lambda]}{n-1}$$

Review[1] of R.A. Fisher's *Inverse Probability*

1. Biographical Notes

Ronald Aylmer Fisher was born in London in 1890. He received scholarships to study mathematics, statistical mechanics and quantum theory at Cambridge University, where he also studied evolutionary theory and biometrics. After graduation, he worked for an investment company and taught mathematics and physics at public schools from 1915 to 1919. From 1919 to 1943, he was associated with the Rothamsted (Agricultural) Experimental Station, contributing to experimental design theory and the development of a Statistics Department. In 1943 he became professor of genetics at Cambridge, remaining there until his retirement in 1957. He died at Adelaide in Australia, in 1962.

2. Review of *Inverse Probability*

In this article, published in 1930, R.A. Fisher cautions against the application of prior probability densities for parameter estimation using inverse probability[2], when *a priori* knowledge of the distribution of the parameters is not available (e.g., from known frequency distributions). Fisher indicates in the first paragraph, that the subject had been controversial for some time, suggesting that:

> "Bayes, who seems to have first attempted to apply the notion of probability, not only to effects in relation to their causes but also to causes in relation to their effects, invented a theory[3], and evidently doubted its soundness, for he did not publish it during his life."

Fisher describes the manner in which a (known) prior density can be used in the calculation of probabilities:

> *"Suppose that we know that a population from which our observations were drawn had itself been drawn at random from a super-population…that the probability that* $\theta_1,\theta_2,\theta_3,\ldots$ *shall lie in any defined infinitesimal range* $d\theta_1 d\theta_2 d\theta_3\ldots$ *is given by*
>
> $$dF = \Psi(\theta_1,\theta_2,\theta_3,\ldots)d\theta_1 d\theta_2 d\theta_3\ldots,$$
>
> *then the probability of successive events (a) drawing from the super-population a population with parameters having the particular values* $\theta_1,\theta_2,\theta_3,\ldots$ *and (b) drawing from such a population the sample values* $x_1,\ldots,x_n$*, will have a joint probability*

1 Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.

2 According to H.A. David and A.W.F. Edwards in *Annotated Readings in the History of Statistics*, Springer-Verlag 2001, page 189:

> *"…"Inverse probability"…would not necessarily have been taken to refer exclusively to the Bayesian method (which in the paper Fisher calls "inverse probability strictly speaking") but to the general problem of arguing "inversely" from sample to parameter…"*

3 In the mid-1700s.

$$\Psi(\theta_1,\theta_2,\theta_3,...)d\theta_1 d\theta_2 d\theta_3...\times \prod_{p=1}^{n}\{\phi(x_p,\theta_1,\theta_2,\theta_3,...)dx_p\}.$$

> *If we integrate this over all possible values of $\theta_1,\theta_2,\theta_3,...$ and divide the original expression by the integral we shall then have a perfectly definite value for the probability…that $\theta_1,\theta_2,\theta_3,...$ shall lie in any assigned limits."*

It is noted that this is a direct argument, which provides the frequency distribution of the population parameters $\boldsymbol{\theta}$. Fisher's caution relates to cases in which the function Ψ is not known, and is then taken to be constant. He argues that this assumption is as arbitrary as any other, and will have inconsistent results. While an example is not given in the *Inverse Probability* paper, it is helpful to consider an illustration provided elsewhere by Fisher, related by Anders Hald in *A History of Mathematical Statistics From 1750 to 1930*[1].

Consider the posterior probability element:

$$\mathrm{P}(\theta|\mathrm{a,n})\mathrm{d}\theta \;\propto\; \theta^{\mathrm{a}}(1-\theta)^{n-a}\mathrm{d}\theta,\quad 0\le\theta\le 1$$

Then if the parameter ς is defined by

$$\sin\varsigma = 2\theta\text{-}1,\quad -\frac{1}{2}\pi \le \varsigma \le \frac{1}{2}\pi$$

such that

$$\varsigma = \arcsin(2\theta-1)$$

and ς is assumed to be uniformly distributed, the posterior probability element becomes:

$$\mathrm{P}(\varsigma|\mathrm{a,n})\mathrm{d}\varsigma \;\propto\; (1+\sin\varsigma)^{\mathrm{a}}(1-\sin\varsigma)^{n-a}\mathrm{d}\varsigma$$

Since

$$\begin{aligned}\frac{d\varsigma}{d\theta} &= \frac{d\arcsin(2\theta-1)}{d\theta}\\ &= \frac{1}{\sqrt{1-(2\theta-1)^2}}\cdot\frac{d(2\theta-1)}{d\theta}\\ &= \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}\end{aligned}$$

$$d\varsigma = \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}d\theta,$$

it follows that:

1 *A History of Mathematical Statistics From 1750 to 1930,* John Wiley and Sons Inc., 1998, page 277.

$$P(\varsigma|a,n)d\varsigma \propto \theta^{a-\frac{1}{2}}(1-\theta)^{n-a-\frac{1}{2}}d\theta .$$

However, since under the assumptions we can show that $P(\theta|a,n)d\theta \propto P(\varsigma|a,n)d\varsigma$, then $P(\varsigma|a,n)d\varsigma \propto \theta^{a-\frac{1}{2}}(1-\theta)^{n-a-\frac{1}{2}}d\theta$ is inconsistent with $P(\theta|a,n)d\theta \propto \theta^{a}(1-\theta)^{n-a}d\theta$.

To Fisher, the use of prior densities (not based on known frequencies) implies that nothing can be known about the parameters, regardless of the amount of information available in the observations. What is needed, according to Fisher, is a "rational theory of learning by experience".

It is noted that (for continuous distributions) the likelihood (function) is not a probability, however it is a measure of "rational belief". He writes:

> *"Knowing the population we can express our incomplete knowledge of, or expectation of, the sample in terms of probability; knowing the sample we can express our incomplete knowledge of the population in terms of likelihood."*

Next the concept of *fiducial* distribution is introduced, which is an example of the *confidence concept* described by H.A. David and A.W.F. Edwards as:

> *"…the idea that a probability statement may be made about an unknown parameter (such as limits between which it lies, or a value which it exceeds) will be correct under repeated sampling from the same population."*[1]

Fisher writes:

> "In many cases the random sampling distribution of a statistic, T, calculable directly from the observations, is expressible solely in terms of a single parameter, of which T is the estimate found by the method of maximum likelihood. If T is a statistic of continuous variation, and P the probability that T should be less than any specified value, we have then a relation of the form
>
> $$P = F(T,\theta)$$
>
> If we now give to P any particular value such as .95, we have a relationship between the statistic T and the parameter θ, such that T is the 95 per cent. Value corresponding to a given θ…"

In *Principles of Statistics*, by M.G. Bulmer[2,] an illustration from a 1935 paper by Fisher is described, with sampling from a normal distribution. If a sample of size n is taken, the quantity

1 *Annotated Readings in the History of Statistics*, page 187.
2 *Principles of Statistics* by M.G. Bulmer, Dover Publications, Inc., New York 1979, page 177.

$$\frac{\bar{x}-\mu}{s/\sqrt{n}}$$

(with notation as usually defined in contemporary statistics) follows a *t* distribution with n-1 degrees of freedom. Then 100P per cent of those values would be expected to be less than t_p, or the probability that

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \leq t_P$$

is equal to P. Fisher notes that the above inequality is equivalent to

$$\mu \geq \bar{x} - st_P/\sqrt{n}$$

and reasons that the probability that $\mu \geq \bar{x} - st_P/\sqrt{n}$ is also P. In this case, μ is a random variable and $\bar{x}$ and s are constants. By varying t_P, the probability that μ is greater than specific values may be obtained, establishing a *fiducial* distribution for μ, from which *fiducial* intervals may be constructed. Such intervals would correspond to the confidence intervals (as defined in contemporary statistics). The interpretation however, is different[1]. In his paper, Fisher provides a table, associated with correlations derived from four pairs of observations.

H.A. David and A.W.F. Edwards suggest that *Inverse Probability* is the first paper clearly identifying the confidence concept (although similar approximate constructs such as "probable error" had been in use for some time). It is also suggested that Student (W.S. Gosset) first expressed the notion (in an exact way) remarking in his 1908 paper:

> *"…if two observations have been made and we have no other information, it is an even chance that the mean of the (normal) population will lie between them."*[2]

1 In a confidence interval, μ is a constant, with a certain probability of being contained in a random interval.
2 *Annotated Readings*…page 187.