An Historical Survey of the Development of Probability and Statistics based on the Appearance of Fundamental Concepts [1]

Submitted by Daniel McFadyen, September 8, 2003

In an essay entitled *Tradition in Science*[2], the quantum physicist Werner Heisenberg, speaking about progress in science generally, suggests that

"...our present problems, our methods, our scientific concepts are, at least partly, the results of a scientific tradition which accompanies or leads the way of science through the centuries. It is therefore natural to ask to what extent our present work is determined or influenced by tradition...Looking back upon history...we see that we apparently have little freedom in the selection of our problems. We are bound up with the historical process, our lives are parts of this process, and our choice seems to be restricted to the decision whether or not we want to participate in a development that takes place in our time...one may say that a fruitful period is characterized by the fact that the problems are given, that we need not invent them."

Philosopher of science Ian Hacking, more specifically addressing the advent of ideas contributing to probability theory, writes in his book *The Emergence of Probability*[3],

"I am inclined to think that the preconditions for the emergence of our concept of probability determined the very nature of this intellectual object 'probability'...The preconditions for the emergence of probability determined the space of possible theories about probability. That means that they determined, in part, the space of possible interpretations of quantum mechanics, of statistical inference, and of inductive logic."

It is clear of course, that new ideas are necessarily expressed using the methods of communication which precede them. Even new symbols must be explained, and to do so, established forms must be used. It is elementary, however very important. Combined,

[1] Submitted for STA 4000H under the direction of Professor Jeffrey Rosenthal.
[2] *Tradition in Science* by Werner Heisenberg, Seabury Press, New York 1983.
[3] Cambridge University Press, Cambridge 1975

Heisenberg and Hacking are proposing even more than this. A science such as probability has as its provenance a set of problems, which are recognized and studied at some point in time, and in some way define the very nature and development of that science. The possibility that this is so, is of itself substantial motivation for the historical review of probability theory. From a practical perspective, ideas and concepts may be found which are of relevance to on-going research or suggest avenues which, have not yet been investigated. Because of this, an historical overview of the subject is highly relevant. Additionally, such a review is both interesting and of pedagogical value to the exposition of fundamental concepts.

As the volume of material relating to the field of probability and statistics is quite substantial, a methodology for presenting an overview must be proposed. First, the basic subject matter must be identified. A survey of the Table of Contents of books relating to the elementary theory of probability and statistics would present a list of topics similar to the following:

- random events and the concept of probability
- random variables and probability distributions
- the binomial distribution (and other discrete distributions)
- mathematical expectation
- the normal distribution
- the $\chi 2$, t and F distributions
- point and interval estimation
- statistical inference

The historical development of the theory of probability and statistics does to a large degree follow the order of topics just listed, and it is possible, based upon a review of historical commentary on the development of probability and statistics, to assemble a set of writings in which those concepts appear for the first time. A review of those works may then provide some insight into the problems being investigated at the time.

The brief historical overview presented in this paper relates to the development of probability in Europe, from the 16th century to the 20th century. Ian Hacking cites evidence that probability concepts are found in the eastern world (India and the Arabias) at a much earlier date. In Europe the first “calculations on chance” are recorded in the 15th century in books such as the one by Fra Luca Pacioli on bookkeeping.

That a mathematical theory of probability begins to develop in Europe in the 16th and 17th centuries, despite considerable evidence that people centuries earlier were aware of random processes and played games of chance remains a mystery. One suggestion, favoured by Ian Hacking, is based on the simple observation that even elementary probability requires the use of arithmetic:

"These [probabilities] are calculations with numbers; when you can do them, you know quite a lot about probability...these calculations require some facility with figures. The Greeks...lacked a perspicuous notation for numerals. So did their heirs. Perhaps a symbolism that makes addition and multiplication easy is a pre-requisite for any rich concept of probability."

Evidence of the first systematic presentation of a subject related to probability appears to date back to the mid-sixteenth century. Gerolamo Cardano, an Italian medical doctor with an interest in mathematics and gambling, wrote a treatise on games of chance entitled *De Ludo Aleae*. Although this work was not published until 1663, almost 100 years after his death, it may be indicative of the kind of scientific dialogue already underway in the 16th century. Cardano’s treatise provides both advice to gamblers and a theoretical consideration of outcomes in dice and card games.

In the first Chapter of his book, entitled *On Kinds of Games*, Cardano categorizes games based on whether they depend on chance or not. Again, an apparently obvious distinction, known by people for centuries (Cardano cites evidence that persons in ancient times played games of chance), however highly relevant. This is a statement recognizing that there are processes, which have random[4] outcomes and those, which do not. The identification of random processes is fundamental to the development of probability theory. It is precisely this category of phenomena, restricted in this case to games of chance, which is the object of Cardano's study.

In Chapter 9 of his book, Cardano begins to present the results of his theory on dice, and with respect to the roll of a die writes:

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

The use of *"should turn up"* in the articulation of this principle suggests that it is based upon the symmetry of the die, with six sides, each as likely as the other to occur. It is (for now) an intuitive concept, and is even used to introduce elementary probability calculations today. Also, Cardano recognises (likely confirmed by experience), that the principle is an ideal, and that in practice we will not have each point turn up once in every six casts. There is clearly a "long range relative frequency" interpretation of *"in six casts each point should turn up once".* In the contemporary language of probability theory, we would say that we *expect* in six casts each point should turn up once.

Another important pre-requisite for calculating probabilities is the facility to conduct counting based on permutations and combinations. In Chapter 11 Cardano discusses the case of two dice, enumerating 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."

A statement mirroring the concept of probability can then be made:

[4] Leaving aside problems in defining "randomness", and trusting intuition for the present.

"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."

The "circuit" is Cardano's expression of the size of the sample space. The use of "odds" and "favourable" convey the sense in which probability is interpreted.

It is even possible that Cardano uses the word "probability" as it would be used today:

"In comparison where the probability is one half, as of even faces with odd…"

Although the use of this sentence perhaps merits more investigation, considering the treatise was published in 1663, after the word "probability" appears in a text by Christiaan Huygens. Whether used by Cardano or not, it is clear that important concepts have been expressed in *De Ludo Aleae*, and may have been known within the societies of learning in 16th century Italy. It would not be surprising that probability mathematics made its appearance in Europe by way of Italy, as Italy maintained close contacts with the eastern world[5].

The earliest published article devoted to explaining observations from a random process appears to have been written by Galileo Galilei likely between 1613 and 1623[6]. He wrote *Sopra Le Scoperte dei Dadi* in 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."

[5] The story of Leonardo of Pisa is an interesting example of Italian contact with Arabian society. See *Leonard of Pisa* by Joseph and Francis Gies, New Classics Library, Gainesville GA, text copyright 1969.
[6] Refer to F.N. David's *Games, Gods and Gambling – A History of Probability and Statistical Ideas*, Dover Publications 1962, page 62.

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. He is able to show 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).

Galileo's explanation combines intuition with observation. Empirical evidence suggests that the sum 10 is more frequently realized than 9. As there are more combinations of numbers summing to 10, this is cited as the likely reason.

Most commonly associated with the advent of probability theory are the ideas presented in the correspondence between the French mathematicians Pierre de Fermat and Blaise Pascal between 1654 and 1660. Of these, Pascal's letter to Fermat on July 29, 1654 is especially important. The famous "problem of points" is solved. The problem relates to the equitable distribution of the proceeds for a wager in a game of chance (for example dice) when the gamblers agree to discontinue play. For the development of probability theory this was a significant problem – precisely the kind of problem Werner Heisenberg alludes to in *Tradition in Science*. While the means for computing probabilities were available much earlier, and as Cardano and Galileo demonstrated, it was possible to identify more probable events – these were expressions *of* probability. Pascal and Fermat would *use* probabilistic reasoning to solve more complicated problems. H.A David and A.W.F. Edwards note in *Annotated Readings in the History of Statistics*[7]:

[7] *Annotated Readings in the History of Statistics*, Springer-Verlag 2001.

"The notion of the expected value of a gamble or of an insurance is as old as those activities themselves, so that in seeking the origin of "expectation" as it is nowadays understood it is important to be clear about what is being sought. Straightforward enumeration of the fundamental probability set suffices to establish the expected value of a throw at dice, for example, and Rennaisance gamblers were familiar enough with the notion of a fair game, in which the expectation of each player is the same so that each should stake the same amount. But when more complicated gambles were considered, as in the Problem of Points, no one was quite sure how to compute the expectation…Pascal clarified the basic notion [expectation] and used it to solve the problem of points."

Pascal 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).

Pascal then generalizes the result for distributing a wager of 2W (each player providing W), when one of the players requires one more point, as $2W-W/2^n$, 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."

If each player wagers W, then the distribution of the wager after the first throw is:

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

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

Pascal is able to relate the above formula to results involving binomial expansions[8]. This will be an important field of investigation and application in the development of probability theory.

Christiaan Huygens' *On Reasoning in Games of Chance*, is cited in the literature as the first published mathematical treatise on the subject of probability[9]. 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[10].

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

---

[8] Note that $1/2 \cdot 3/4 \cdot 5/6 \cdot \ldots \cdot (2n-1)/(2n) = (2n-1)!/n!(n-1)! \cdot 1/2^{2n-1}$; $(2n-1)!/n!(n-1)! = 1/2 \cdot (2n!/n!n!)$; $2^{2n-1} = 1/2 \cdot (1+1)^{2n}$ *and* $1/2 \cdot (1+1)^{2n} = 1/2 \cdot \sum_{i=0}^{2n} \binom{2n}{i}$.

9 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.

[10] Huygens visited Paris and had contact with associates of Pascal and Fermat.

itself[11], is observed in writings prior to Huygens', the 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.

Huygens' Propositions I, II and III summarizes the principles for evaluating expectations (values of wagers):

Proposition I:

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

Proposition II 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:

*"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**."*

In this context "chances" is equivalent to the number of ways a specific outcome can occur. Although the quotients $\frac{p}{p+q}$, and $\frac{q}{p+q}$ are not named, they are probabilities, and the proposition describes the additive principle.

[11] Refer to Gerolamo Cardano's *De Ludo Aleae*.

The multiplicative property of probability is not as clearly seen, however the use of Propositions I to III do account for this principle. Consider the solution presented by Huygens to the following problem

"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 $(1 \cdot \mathbf{a} + 5 \cdot \mathbf{0}) / (1+5) = 1/6\mathbf{a}$. 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 $(5/6) \cdot (1/6)$. In Huygens system, expectations for simpler cases are combined using Proposition 3, to solve more complex problems.

By 1657, the basic characteristics of probability as we apply them today had been established. The word "probability" has now been introduced, with a meaning comparable to today's, however the principal interest remains the computation of expectations. The elementary laws of probability are clearly at work in Huygen's Propositions, however the probability ratio is a means for computation rather than an object for investigation itself.

In Huygen's book solutions to a variety of problems regarding expectations are solved, and problems are also provided for the reader. Indeed, this stage of development of probability theory may be characterized by an investigation into the variety of problems (principally relating to games of chance) which can be addressed using methods developed. Perhaps representative of this stage of development is Pierre Remond de Montmort's *Essai d'Analyse sur les Jeux de Hazard*, which presents a theory of combinatorics, discusses certain games of chance with cards and with dice and provides solutions to various problems including the five problems proposed by Huygens. Included in this book is an early solution to a problem relating to coincidences or matches. Montmort describes a method for computing the expectation of a card being drawn with value and order drawn considered equivalent.

In 1710, Dr. John Arbuthnott, a friend of Jonathan Swift and Isaac Newton, wrote a paper entitled *An Argument for Divine Providence, taken from the constant Regularity observed in the Births of both Sexes*. In a history of the appearance of fundamental concepts of probability and statistics, this paper is important for the following reasons:

- It is one of the first applications of probability to phenomena other than games of chance.
- It possesses an argument, which has been referred to as the first published test of significance[12].

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

[12] Ian Hacking, *The Emergence of Probability*, page 168.

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.

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, 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. James Bernoulli will present a theorem relevant to this in the next few years.

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[13]), 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), Arbuthnott 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.

By the late 17th century, the nature of probability and rules for computing results in processes requiring probability had become quite familiar to the scientific community. Mathematics relating to binomial expansions was being used, and the methods developed were being applied to subjects other than gambling. Dr. Arbuthnott even suggests in an earlier paper[14],

"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."

This period also coincides with advances in other fields of mathematics, including the calculus, limits, sequences, series and series expansions of functions, and approximation methods for factorials. These would be of considerable importance to the development of probability theory.

In Dr. Arbuthnott's article there was a foreshadowing of an idea that would become central to probability and statistics, and would profit from recent advances made in the mathematical sciences. This is the notion of the large sample. It became apparent that certain statements could be made about sets of data, which were very large, and it became necessary to make statements about very large data sets.

[13] The christenings would likely have been a reasonable proxy for births during that period. Arbuthnott presents the data in his paper.

[14] In a 1692 translation of Huygens' *De ratiociniis in ludo aleae* (1657) entitled *Of the Laws of Chance.*

In 1713 James Bernoulli, a mathematician from Bâle published a book on probability entitled *Artis Conjectandi*, which discusses a number of problems requiring probability mathematics, and considers the nature of probability:

"...probability is a degree of certainty and differs from absolute certainty as a part differs from the whole. If, for example, the whole and absolute certainty – which we designate by the letter a or by the unity symbol 1 – is supposed to consist of five probbilities or parts, three of which stand for the existence or future existence of some event, the remaining two standing against its existence or future existence, this event is said to have 3/5a or 3/5 certainty."

Bernoulli also suggests that probability is a consequence of uncertainty:

"…those data which are supposed to determine later events (and especially such data which are in nature) have nevertheless not been learned well enough by us."

This view is consistent with those of the adherents of "determinism" in the 18^{th} century. In Chapter IV of Part IV of his text on probability, *Artis Conjectandi* (published in 1713), 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…"[15]

The proposed solution to this inquiry, is what would be called the law of large numbers today. Bernoulli's proposition is as follows:

[15] Translation by Bing Sung (1966). *Translations from James Bernoulli*. Department of Statistics, Harvard University, Cambridge, Massachusetts.

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.

This theorem is proven using lemmas relating to properties of binomial expansions, and the limiting characteristics of quotients of binomial terms and sums. It is significant as the first mathematical formalization of the "intuitive" notion relating to long range relative frequency.

Following this important theorem, in 1718 Abraham DeMoivre published his *Doctrine of Chances*, which again addresses a number of practical probability problems. In an addendum to his book after 1733, referred to as the *Approximatio*, DeMoivre derives a formula for approximating binomial sums or probabilities when the number of trials is very large. He was able to do this with the aid of mathematics relating to series expansions for logarithms and exponentials, as well as approximation methods for factorials. DeMoivre shows for a very large number of trials denoted by n,

"… 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√n times, not more rarely than 1/2n – 1/2√n times, will be expressed by the double Sum of the number exhibited in the second Corollary, that is by 0.682688…"

This is a result about what would later be called 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 √n. In his book, *The Life and Times of the Central Limit Theorem*[16], William Adams states:

"De Moivre did not name√n/2, which is what we would today call standard deviation within the context considered, but … he referred to √n as the Modulus by which we are to regulate our estimation."

Again, it is noteworthy that the first appearance of the normal distribution was in relation to an approximation problem. In the late 18th century and early 19th century, the normal distribution would be more explicitly recognized and appreciated for its facility in describing errors of observation. Here also, the problems were presented to the inquirers, in this case in the field of astronomy, where errors of observation were of great importance. Carl F. Gauss and Pierre S. Laplace contributed substantially in this stage of the growth of probability theory. Laplace is credited with the first proof of a central limit theorem. Jerzy Neyman, in his text *First Course in Probability and Statistics*[17], presents Laplace's Theorem (in contemporary notation) as follows:

Whatever be two numbers $t_1 < t_2$, and whatever be the fixed value of the probability of success p, 0<p<1, if the number n of completely independent trials is indefinitely increased, then the probability that the corresponding binomial variable X(n) will satisfy the inequalities

$$t_1 < \frac{X(n) - np}{\sqrt{np(1-p)}} < t_2$$

tends to the limit

$$\frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{x^2}{2}} dx$$

[16] Kaedmon Publishing Company, New York 1974, page 24.
[17] *First Course in Probability and Statistics*, Henry Holt and Company, New York 1950.

The normal distribution and central limit theorem have considerable application in the contemporary practice of probability and statistics.

In the 19th century, the precision of estimates was commonly measured using the "probable error"[18]. In an 1876 article by Friedrich Robert Helmert, which was written with the intention of reporting on an improved method for estimating probable error, it is shown that u, the sum of squares of observed deviations from the mean is distributed according to

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

where h (referred to as "the precision") is equal to $1/\sigma\sqrt{2}$. The result was obtained with the aid of matrix mathematics and the methods of (multivariate) calculus. From this result, it is effectively shown that given $X_1,\ldots,X_n$ independent $N(\mu,\sigma^2)$ random variables, then

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

that is, the sum of squares of the observed deviations from the mean, are distributed as (what would be called) a chi-square distribution with n-1 degrees of freedom. Once again, a result relating to a distribution widely applied in the practice of probability and statistics, has been obtained in the process of the investigation of another matter.

By the end of the 19th century, a considerable amount of the theory of probability and statistics as presented in contemporary elementary texts, had been developed. Statistics, which involves the inferring of information relating to population parameters from sample observations, makes considerable application of results relating to the law of large numbers, normal distribution, central limit theorem and chi-square distributions. Not

[18] Defined as $\sigma\Phi^{-1}(0.75)$, where Φ is the standard normal distribution function. See *Annotated Readings…*, page 103.

reviewed in this paper, however also of considerable importance, is the t distribution, attributed to W.S. Gosset. This distribution has application in the use of small samples, the opposite to the problem presented in the early 18th century.

The last major topic from the list given at the beginning of this paper relates to interval estimation. In 1930, Ronald Aylmer Fisher wrote a paper entitled *Inverse Probability*, which presents a critique of some practices in what is referred to nowadays as Bayesian estimation. The second half of the paper introduces the notion of *fiducial* inference. The concept is explained in *Principles of Statistics*, by M.G. Bulmer[19] using an example with sampling from a normal distribution. If a sample of size n is taken, the quantity

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

(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

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

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

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

and reasons that the probability that $\mu \ge \bar{x} - st_P/\sqrt{n}$ 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), however are interpreted with μ as a random variable

[19] *Principles of Statistics* by M.G. Bulmer, Dover Publications, Inc., New York 1979, page 177.

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). The appearance of *fiducial* intervals is a response to the Bayesian approach to estimation.

The brief survey of developments in the history of probability and statistics given in this paper appears to confirm the statement of Werner Heisenberg. In each of the centuries since the advent of a probability theory, problems have been proposed, which have contributed to advances, even if not always recognized at the time. The sudden appearance (in Europe) of probability in the 16th or 17th century may have been related to modifications in numeric notation, facilitating arithmetic calculations. Also, it is clear that the theory of probability has its origins in questions on gambling. 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. The availability of observable and measurable "fundamental probability sets" is perhaps one of the preconditions alluded to by Ian Hacking, and determines the very nature of probability. It may also be noted that probability has been defined over the centuries in the sense of long range relative frequency, and associated at the same time with a degree of belief. James Bernoulli discusses both of these characteristics in the *Artis Conjectandi,* and provides the first limit theorem, mathematically formalizing the notion of long range relative frequency in the case of binomial data. This theorem is itself a probability statement about probabilities. Very generally, from the mid-eighteenth century the development of concepts in probability and statistics proceeds as follows: the need for approximating binomial sums for very large numbers of trials lead to the discovery of the normal distribution; the study of astronomical data and errors of observation provided for an application of the normal distribution, and discovery of a central limit theorem; research to improve the estimate of the probable error identifies the χ^2 distribution, and questions relating to the application of inverse probability lead to a definition of the confidence interval.

In the introduction of D.A.S. Fraser's book *Probability and Statistics*[20], an observation is made about the probability of an event:

"...it is a physical phenomenon that the proportion of occurrences for the event tends to a limit as the number of repetitions increases indefinitely; the limit is called the probability of the event. A system with this property is called a stable system or a random system. In cases where the proportion behaves systematically or does not tend to a limit, experience indicates that the input variables have not been controlled or kept constant; in other words, the system is not random."

It is perhaps surprising to see random systems described as stable systems. However, this is again an important precondition. The methods of probability and statistics are only applicable to systems, which have such a nature. In contemporary categories, probability is usually defined in the following manner[21]:

Given a sample space Ω with σ field of events Φ, the probability of the event $A \in \Phi$, written $P(A)$, is a set function onto the closed interval $[0,1]$ such that

i) $0 \leq P(A) \leq 1$, *all* $A \in \Phi$,

ii) $P(\Omega) = 1$,

iii) $P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$

Such an axiomatic definition (attributable to the Russian mathematician A. Kolmogorov), allows for considerable versatility in practice and research. It is frequently advisable however, to check the characteristics of the model with those of the object being modeled. In this regard again, an appreciation for the development of probability theory

[20] *Probability and Statistics, Theory and Applications*, DAI Press, University of Toronto 1976.
[21] From C.R. Heathcote's *Probability, Elements of the Mathematical Theory*, Dover Publications Inc. Mineola 1971.

may be useful in ensuring consistency with the historical understanding, which has the benefit of the contribution of many minds through the centuries.

Bibliography

Primary Sources (Chronological – based on date of writing)

Gerolamo Cardano. *Liber De Ludo Aleae* (mid-sixteenth century).

Galileo Galilei. *Sopra Le Scoperte dei Dadi* (between 1613 and1623).

Pierre de Fermat and Blaise Pascal. *Correspondence* (1654 to 1660).

Christiaan Huygens. *De Ratiociniis in Ludo Aleae* (1657).

John Arbuthnott. *An Argument for Divine Providence, taken from the constant Regularity observed in the Births of both Sexes* (1710).

Pierre Remond De Monmort. *On the Game of Thirteen* (1713).

James Bernoulli. *Artis Conjectandi* (1713).

Abraham DeMoivre. *The Doctrine of Chances* (1733).

Friedrich Robert Helmert. *The Calculation of Probable Error from the Squares of the Adjusted Direct Observations of Equal Precision and Fechner's Formula* (1876).

Ronald Aylmer Fisher. *Inverse Probability* (1930).

<u>Secondary Sources</u>

Adams, William. *The Life and Times of the Central Limit Theorem.* Kaedmon Publishing Company, New York 1974.

Bing Sung. *Translations from James Bernoulli*. Department of Statistics, Harvard University, Cambridge, Massachusetts 1966.

Bulmer, M.G. *Principles of Statistics.* Dover Publications, Inc., New York 1979.

David, F.N. *Games, Gods and Gambling – A History of Probability and Statistical Ideas*, Dover Publications 1962.

David, H.A. and Edwards, A.W.F. *Annotated Readings in the History of Statistics*. Springer-Verlag 2001.

Fraser, D.A.S. *Probability and Statistics, Theory and Applications*. DAI Press, University of Toronto 1976.

Gies, Joseph and Francis. *Leonard of Pisa.* New Classics Library, Gainesville GA, text copyright 1969.

Hacking, Ian. *The Emergence of Probability.* Cambridge University Press, Cambridge 1975.

Hald, Anders. *A History of Probability and Statistics and Their Applications before 1750.* John Wiley and Sons, 1990.

Hald, Anders. *A History of Mathematical Statistics from 1750 to 1930.* John Wiley and Sons, 1998.

Heathcote, C.R. *Probability, Elements of the Mathematical Theory*. Dover Publications Inc. Mineola 1971.

Heisenberg ,Werner. *Tradition in Science*. Seabury Press, New York 1983.

Maistrov, L.E. *Probability Theory – A Historical Sketch*. Academic Press Inc., New York 1974.

Neyman, J. *First Course in Probability and Statistics*, Henry Holt and Company, New York 1950.

Oystein, Ore. *Cardano, the Gambling Scholar*. Dover Publications, New York 1965.

Todhunter, Isaac. *A History of the Mathematical Theory of Probability.* Chelsea Publishing Co., New York (1965 unaltered reprint of the First Edition, Cambridge 1865).