The Early Development of Mathematical Probability

Glenn Shafer

This article is concerned with the development of the mathematical theory of probability, from its founding by Pascal and Fermat in an exchange of letters in 1654 to its early nineteenth-century apogee in the work of Laplace. It traces how the meaning, mathematics, and applications of the theory evolved over this period.

1. Summary

Blaise Pascal and Pierre Fermat are credited with founding mathematical probability because they solved the problem of points, the problem of equitably dividing the stakes when a fair game is halted before either player has enough points to win. This problem had been discussed for several centuries before 1654, but Pascal and Fermat were the first to give the solution we now consider correct. They also answered other questions about fair odds in games of chance. Their main ideas were popularized by Christian Huygens, in his *De ratiociniis in ludo aleae*, published in 1657.

During the century that followed this work, other authors, including James and Nicholas Bernoulli, Pierre Rémond de Montmort, and Abraham De Moivre, developed more powerful mathematical tools in order to calculate odds in more complicated games. De Moivre, Thomas Simpson, and others also used the theory to calculate fair prices for annuities and insurance policies.

James Bernoulli's *Ars conjectandi*, published in 1713, laid the philosophical foundations for broader applications. Bernoulli brought the philosophical idea of probability into the mathematical theory, formulated rules for combining the probabilities of arguments, and proved his famous theorem: the probability of an event is morally certain to be approximated by the frequency with which it occurs. Bernoulli advanced this theorem (later called the law of large numbers by Poisson) as a justification for using observed frequencies as probabilities, to be combined by his rules to settle practical questions. Bernoulli's ideas attracted philosophical and mathematical attention, but gambling, whether on games or on lives, remained the primary source of new ideas for probability theory during the first half of the eighteenth century.

In the second half of the eighteenth century, a new set of ideas came into play. Workers in astronomy and geodesy began to develop methods for reconciling observations, and students of probability theory began to seek probabilistic grounding for such methods. This work inspired Pierre Simon Laplace's invention of the method of inverse probability, and it benefitted from the.evolution of Bernoulli's law of large numbers into what we now call the central limit theorem. It culminated in Adrien Marie Legendre's publication of the method of least squares in 1805 and in the probabilistic rationales for least squares developed by Laplace and by Carl Friedrich Gauss from 1809 to 1811. These ideas were brought together in Laplace's great treatise on probability, *Théorie analytique des probabilités*, published in 1812.

2. Games of Chance

Games of chance and the drawing of lots are discussed in a number of ancient texts, including the Talmud. A number of medieval authors, both mystics and mathematicians, enumerated the ways various games can come out. But most of these enumerations were not enumerations of equally likely cases, and they were not used to calculate odds (Kendall *1956*, David *1962*).

One genuine precursor of probability theory was Cardano's *Liber de ludo aleae*. This book was probably written in the 1560s (Cardano died in 1576), but it was not published until after the work of Huygens. Cardano formulated the principle that the stakes in an equitable wager should be in proportion to the number of ways in which each player can win, and he applied this principle to find fair odds for wagering with dice. Another precursor was Galileo, who enumerated the possible outcomes for a throw of three dice in order to show that the faces add to ten more easily and hence more frequently than they add to nine. This work also remained unpublished until after the work of Huygens.

Galileo did not comment on the problem of points, but it was the most salient mathematical problem concerning games of chance during this period. It was discussed by Italian mathematicians as early as the fourteenth century, and Pacioli, Tartaglia, Peverone, and Cardano all tried to solve it. Tartaglia concluded it had no solution: "the resolution of the question is judicial rather than mathematical, so that in whatever way the division is made there will be cause for litigation." But Pascal and Fermat found a solution, and they gave several convincing arguments for it.

Fermat preferred to solve the problem by listing the ways the play might go. Suppose, for example, that Peter and Paul have staked equal money on being the first to win three points, and they want to stop the game and divide the stakes when Peter lacks two points and Paul lacks only one. Paul should get more than Peter because he is ahead, but how much more? If two more games were to be played, there would be four possibilities:

> Paul wins the first and the second.
> Paul wins the first, and Peter wins the second.
> Peter wins the first, and Paul wins the second.
> Peter wins the first and the second.

In the first three cases, Paul wins the game (in the first two, it is not even necessary to play for the second point), and in the fourth case Peter wins the game. By Cardano's principle, the stakes should be in the same proportion: three for Paul and one for Peter. If the stakes are divided now, Paul gets three-fourths.

Pascal preferred another method, the "method of expectations." This method relies on principles of equity instead of Cardano's principle, and it enabled Pascal to solve the problem when the players lack so many points that listing all the possibilities is impractical. Since the game is fair, Pascal said, both players have an equal right to expect to win the next point. If Paul wins it, he wins the whole stakes; this entitles him to half of the stakes. If Peter wins, the two are tied; so both are entitled to half of what is left. So Paul is entitled to three-fourths altogether. Extending this backwards reasoning by mathematical induction, and using the recursive properties of what we now call Pascal's triangle, Pascal was able to solve the problem for any number of points the players might lack. He found, for example, that if Peter lacks four points and Paul lacks two, then their shares are found by adding the numbers in the base of the triangle in Figure 1: Paul's share is to Peter's as 1+5+10+10 to 5+1, or 13 to 3.

1	1	1	1	1	1
1	2	3	4	5	
1	3	6	10		
1	4	10			
1	5				
1					

Figure 1. Pascal's arithmetic triangle

The combinatorial knowledge that underlay Pascal's reasoning was richer and less streamlined than the combinatorics now taught as a foundation for probability. Pascal organized much of this knowledge, together with his general solution of the problem of points, in his *Traité du triangle arithmétique*, which was published in 1665, three years after his death (Edwards *1987*).

Huygens heard about Pascal and Fermat's ideas but had to work out the details for himself. His treatise *De ratiociniis in ludo aleae*, published in 1657, essentially followed Pascal's method of expectations. In one respect, he went farther than Pascal. He justified the very idea that there was a fair price for a position in a game of chance by showing that by paying that price, the player would end up with the same chances for the same net payoffs that could be obtained in a fair game with side payments.

At the end of his treatise, Huygens listed five problems about fair odds in games of chance, some of which had already been solved by Pascal and Fermat. These problems, together with similar questions inspired by other card and dice games popular at the time, set an agenda for research that continued for nearly a century. The most important landmarks of this work were James Bernoulli's *Ars conjectandi*, published in 1713, Montmort's *Essai d'analyse sur les jeux de hasard*, with editions in 1708 and 1711, and De Moivre's *Doctrine of Chances*, with editions in 1718, 1738, and 1756. These authors investigated many of the problems still studied under the heading of discrete probability, including gambler's ruin, duration of play, handicaps, coincidences, and runs. In order to solve these problems, they improved Pascal and Fermat's combinatorial reasoning, summed infinite series, developed the method of inclusion and exclusion, and developed methods for solving the linear difference equations that arise from Pascal's method of expectations. Perhaps the most important technical mathematical development in this work was the invention of generating functions. Though Laplace gave them their name much later, De Moivre invented them in 1733 in order to find the odds for different sums of the faces of a large number of dice. This was the first step in the development of what we now call Laplace and Fourier transforms.

From the time of Huygens onward there was one important practical application of this work: the pricing of annuities and life insurance policies. Huygens's ideas were first applied to pricing annuities by Jan de Witt in 1671 and Edmund Halley in 1694. De Moivre, in his *Annuities on Lives*, with editions in 1725, 1743, 1750, and 1752, and Simpson, in his *Doctrine of Annuities and Reversions*, published in 1742, extended this work to more complicated annuities, reversions, and insurance and assurance policies. This work involved the rate of interest, but otherwise it remained conceptually close to the work on games of chance.

3. Probability

Probability—the weighing of evidence and opinion—was an important topic in the seventeenth century, but it was not discussed in the letters between Pascal and Fermat or in Huygens's treatise. The word “probability” did not appear in this work. For these authors, the number between zero and one that we now call probability was only the proportion of the stakes due a player. They did not isolate it as a measure of belief.

The ambition to apply the new ideas to problems in the domain of probability developed almost immediately, however. Pascal himself used these ideas in his famous argument for believing in God, and the Port Royal *Logic*, published by Pascal's religious colleagues in 1662, argued for using the theory of games of chance to weigh probabilities in everyday life. Soon the idea of probability as a number between zero and one was in the air. The English cleric George Hooper, writing in 1689, gave rules for combining such numbers to assess the reliability of concurring testimony and chains of testimony.

James Bernoulli formalized the connection between probability and Huygens's theory. In *Ars conjectandi*, published in 1713, eight years after his death, Bernoulli contended that an argument is worth a portion of complete certainty, just as a position in a game of chance is worth a share of the total stakes. Take complete certainty to be a unit, he said, and then probability is a number between zero and one. By Huygens's rules, it will be the number of favorable cases to the total number of cases.

Bernoulli tried to draw the traditional ideas of probability into the mathematical theory by formulating rules for combining the probabilities of arguments, rules that were similar to Hooper's rules for testimony but more general. Bernoulli hoped that these rules would make probability a widely used practical tool. Probabilities would be found from observed frequencies, and then combined to make judicial, business, and personal decisions.

Bernoulli buttressed this program with his famous theorem, which says that it is morally certain (99.9% certain, say) that the frequency of an event in a large number of trials will approximate its probability. Though Bernoulli proved this theorem rigorously, he gave only a disappointingly large upper bound on the number of trials needed for moral certainty that the frequency would be within a given distance of the probability.

Nicholas Bernoulli improved the upper bound, and De Moivre, in 1733, accurately estimated the number of trials needed, using a series expansion of the integral of what we now call the normal density. In retrospect, this can be seen as a demonstration that the binomial distribution can be approximated by the normal distribution, but De Moivre did not think of it in this way. He did not have our modern concept of a probability distribution; he was merely improving on Bernoulli in finding the number of trials needed in order to be sure that the observed frequency would approximate the probability.

This work of Bernoulli and De Moivre is also seen today as a precursor of the modern theory of confidence intervals: De Moivre's approximation allows us to state a degree of confidence, based on the observations, that the true probability lies between certain bounds. This is not the way Bernoulli and De Moivre saw the matter. They were aiming for moral certainty, not for some middling degree of confidence.

Bernoulli's theorem and his rules for combining probabilities did not achieve his goal of making probability a tool for everyday life and judicial affairs. His rules for combining probabilities were discussed in textbooks, but not used in practice. His theorem was mainly discussed in the context of speculation about the ratio of girl births to boy births. Yet his program had great philosophical impact. It put probability, rather than equity, into the center of the mathematical theory.

The increasing autonomy of probability from equity in the early eighteenth century can be seen in the contrasting attitudes of Nicholas Bernoulli and his younger cousin Daniel Bernoulli. In the 1730s, they discussed what we now call the St. Petersburg problem. A person throws a die repeatedly, and he wins a prize when he first gets a 6. The prize is doubled every time he fails to get a 6; he wins one crown if he gets a 6 on the first throw, two crowns if his first 6 is on the second throw, four crowns if it is on the third throw, and so on. How much should he pay to play this game? By Huygens's rules, he should pay an infinite amount, yet no one would be willing to pay more than a few crowns. Daniel explained this by formulating the idea of expected utility. If a person's utility for money is only proportional to its logarithm, then the expected utility of the game is finite. Nicholas could not see the point of this explanation, because for him the theory of probability was based on equity. If we are talking about what a person wants to pay rather than what is fair, then there is no basis for calculation (Shafer *1988*).

4. The Combination of Observations

Though probability had found a theoretical place in the mathematical theory by 1750, the applications of the theory were still only to questions of equity. No one had learned how to use probability in data analysis. This was true even in the work on annuities and life insurance. De Moivre used theoretical mortality curves, Simpson used mortality statistics, but no one used probabilistic methods to fit models in the way modern demographers do.

It was work on combining observations in astronomy and geodesy that finally brought data analysis and probability together. In the eighteenth century, combining observations meant reconciling inconsistent equations. An observation typically yielded numbers that could serve as coefficients in a linear equation relating unknown quantities.

A few such observations would give enough equations to find the unknowns. More observations would mean more equations than unknowns, and since the measurements contained errors, these equations would be inconsistent.

Galileo had struggled with this problem, and had vaguely formulated the principle that the most likely values of the unknowns would be those that made the errors in all the equations reasonably small, but the first formal method for estimating the unknown quantities was developed by Tobias Mayer, in his study of the libration of the moon in 1750. Mayer generalized the averaging of observations under identical circumstances to the averaging of groups of similar equations. Another early contributor was Roger Boscovich, who used the idea of minimizing the sum of absolute deviations in his work on the shape of the earth, in 1755. Laplace, in an investigation of the motions of Saturn and Jupiter in 1787, generalized Mayer's method by using different linear combinations of a single group of equations. Then, in 1805, Adrien-Marie Legendre showed how to estimate the unknown quantities by minimizing the sum of squared deviations. This method of least squares was an immediate success, because of its conceptual and computational simplicity, its generality, and its clear relationship to other methods that had been used in practice. It was similar to Boscovich's principle of minimizing absolute deviations, but easier to implement. Like Laplace's generalization of Mayer's method, it entailed solving a set of linear combinations of the original equations, but it gave a generalizable rationale for the choice of these linear combinations.

Though practical work on combining observations was not influenced by probability theory until after Legendre's publication of least squares, attempts to base methods for combining observations on probability theory began very early. Simpson, in 1755, made some arbitrary assumptions about the distribution of errors and deduced that the average of a set of observations of a single quantity is likely to be less in error than the individual observations. A number of authors, including Daniel Bernoulli, Joseph Louis Lagrange, and Laplace, derived methods for combining observations from various assumptions about error distributions.

The most important fruit of this probabilistic work was the invention by Laplace of the method of inverse probability—what we now call the Bayesian method of statistical inference. Laplace discovered inverse probability in the course of his work on the theory of errors in the 1770s. Laplace realized that probabilities for errors, once the observations are fixed, translate into probabilities for the unknown quantities being estimated. He called them posterior probabilities, and he justified using them by adopting the principle that after an observation, the probabilities of its possible causes are proportional to the probabilities given the observation by the causes. This was called the method of inverse probability in the nineteenth century. We now call it Bayes's theorem, in honor of Thomas Bayes, who had enunciated a similar but more obscure principle in a paper published 1764 (Shafer *1982*).

Inverse probability was recognized by Laplace's contemporaries as an important contribution to the theory of probability. During the 1770s and 1780s, moreover, Laplace made great strides in developing numerical methods for evaluating posterior probabilities. But this work did not immediately bring probability theory into contact with the practical problem of combining observations. When applied to the error distributions proposed by Laplace and his contemporaries, inverse probability produced methods for combining observations that were intractable or unattractive in comparison to the established methods.

This changed after Legendre's discovery of the least squares. In 1809, Carl Friedrich Gauss, who considered himself the inventor of the method, published his own account of it. Gauss provided a probabilistic justification for least squares: the least squares estimate of a quantity is the value with greatest posterior probability if the errors have what we now call a normal distribution: a distribution with probability density

$$f(\Delta) = \frac{h}{\sqrt{\pi}} e^{-h^2\Delta^2} \tag{1}$$

for some constant h. Why should we expect the errors to have this distribution? Gauss gave a rather weak argument. It is the only error distribution that makes the arithmetic mean the most likely value when we have observations of a single unknown quantity. Thus the consensus in favor of the arithmetic mean must be a consensus in favor of this error distribution.

Laplace saw Gauss's work in 1810, just after he had completed a paper rounding out several decades of work on approximating the probability distributions of averages. Laplace had originally worked on this problem, in the 1770s, in connection with the question of whether the average inclination of the planetary orbits to the ecliptic was too small to have happened by chance. In the 1780s, he found ways to use generating functions and the asymptotic analysis of integrals to approximate probabilities for averages of many observations, and by 1810 he had succeeded in showing that no matter what the probability distribution of the original observations, probabilities for their average could be found by integrating functions of the form (1). This is a generalization of De Moivre's normal approximation for the binomial; it is the first statement of what we now call the central limit theorem (central meaning fundamental).

With this recent work in mind, Laplace saw immediately that he could improve on Gauss's argument for using (1) as an error distribution. Each individual error would have the distribution (1) if it itself was an average—the resultant of many independent additive influences. The following year, 1811, Laplace also pointed out that Legendre's least square estimates, being weighted averages of the observations, would be approximately normal if there were many observations, no matter what the distributions of the individual errors. Moreover, Laplace showed, these estimates would have the least expected error of any estimates that were weighted averages of the observations. As Gauss pointed out in 1823, this last statement is true even if there are only a few observations.

5. Laplace's synthesis

The work on the combination of observations brought into probability theory the main idea of modern mathematical statistics: data analysis by fitting models to observations. It also established the two main methods for such fitting: linear estimation and Bayesian analysis.

Laplace's great treatise on probability, *Théorie analytique des probabilités*, appeared in 1812, with later editions in 1814 and 1820. Its picture of probability theory was entirely different than the picture that had been available in 1750. On the philosophical side was Laplace's interpretation of probability as rational belief, with inverse probability as its underpinning. On the mathematical side was the method of generating functions, the central limit theorem, and Laplace's techniques for evaluating posterior probabilities. On the applied side, games of chance were still in evidence, but they were dominated by problems of data analysis and by Bayesian methods for combining probabilities of judgments, which replaced the earlier non-Bayesian methods of Hooper and Bernoulli (Shafer *1978*, Zabell *1988*).

Laplace's views dominated probability for a generation, but in time they gave way in turn to a different understanding of probability. The error distributions at the base of Laplace's approach seemed to the empiricists of the nineteenth century to have a frequency interpretation incompatible with Laplace's philosophy, and Laplace's own rationale for least squares based on the central limit theorem seemed to make his method of inverse probability unnecessary. In the empiricist view, which became dominant in the late nineteenth and early twentieth centuries, frequency is the real foundation of probability.

Laplace's synthesis also came apart in a different way. Laplace's powerful mathematics of probability was aimed directly at applications. Mathematics and applications constituted a single topic of study. Today probability has evolved into too vast a topic for such unity to be sustained. It is now a rich branch of pure mathematics, and its role as a foundation for mathematical and applied statistics is only one of its many roles in the sciences.

Bibliography

Shapiro (*1983*) surveys non-mathematical probability in seventeenth-century England, while Hacking (*1975*) surveys the broader intellectual antecedents of mathematical probability. Daston (*1988*) traces the interpretation of probability up to 1840. Hald (*1990*), updating Todhunter (*1865*), catalogs the mathematics and applications of probability before 1750. Stigler (*1986*) traces the development of error theory from 1750 onwards, as well as the further development of mathematical statistics in the nineteenth century.

Daston, Lorraine *1988*. *Classical Probability in the Enlightenment.* Princeton (Princeton University Press).

David, F.N. *1962*. *Games, Gods, and Gambling.* London (Griffin).

Edwards, A.W.F. *1987*. *Pascal's Arithmetic Triangle.* London (Griffin).

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

Hald, Anders *1990*. *A History of Probability and Statistics and their Applications before 1750.* New York (Wiley).

Kendall, Maurice G. *1956*. ‘The beginnings of a probability calculus.’ *Biometrika 43,* 1-14. Reprinted in Pearson and Kendall (*1970*).

Kendall, Maurice G., and R. L. Plackett, eds. *1977*. *Studies in the History of Statistics and Probability*. Volume 2. New York (Macmillan).

Pearson, Egon S., and Maurice G. Kendall, eds. *1970*. *Studies in the History of Statistics and Probability.* Volume 1. Darien, Connecticut (Hafner).

Shafer, Glenn *1978*. ‘Non-additive probabilities in the work of Bernoulli and Lambert.’ *Archive for History of Exact Sciences 19,* 309-370.

Shafer, Glenn *1982*. ‘Bayes's two arguments for the rule of conditioning.’ *Annals of Statistics 10*, 1075-1089.

Shafer, Glenn *1988*. ‘The St. Petersburg paradox.’ *Encyclopedia of Statistical Sciences*, Vol. 8, pp. 865-870. Samuel Kotz and Norman L. Johnson, editors. New York (Wiley).

Shapiro, Barbara J. *1983*. *Probability and Certainty in Seventeenth-Century England.* Princeton (Princeton University Press).

Stigler, Stephen M. *1986*. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, Massachusetts (Harvard University Press).

Todhunter, Isaac *1865*. *A History of the Mathematical Theory of Probability*. London (Macmillan).

Zabell, Sandy L. *1988*. ‘The probabilistic analysis of testimony.’ *Journal of Planning and Inference 20,* 327-354.