Introduction

Like most fields in mathematics, the world of probability and statistics is one of mankind's own creation. The numbers, principles, and equations are not simply out there in nature, waiting to be found. What lies out in nature is an entropic cocktail of order and chaos, sometimes condensing in just such a way that allows for a curious researcher to notice a pattern—a pattern that can then be looked at meticulously and subsequently translated into the language of numbers. The numbers are conceptual too, of course, but they are a powerful tool that allows for an observer to describe the world. Over generations of refinement, humans arrive at new types of math that allow them to describe the world more accurately than before. When we look at things that seem uncertain, perhaps even random, it may be difficult to describe with an algebraic equation. Thus, there is a need for new math: probability and statistics.

From a real-world observation, one can analyze past occurrences and refine that data into a meaningful model. This is the strength of statistics. And from that model, one can extrapolate beyond the data of the past and into the future to make predictions about events that have yet to occur. With statistics, this is what probability enables. In tandem, these tools grant what many have dreamed of: the power to glimpse into the future.

This volume is divided into three sections, with the belief that readers aiming for a more complete understanding of probability and statistics will pursue entries in the progression in which they are provided. Just as how mathematics is designed, we must stand upon the shoulders of giants in the field to reach the heights of study. Each section builds upon and applies the knowledge of the previous section, and thus it is recommended that the volume be read in the order provided. Within each section is a collection of entries that addresses key facets of probability and statistics in a clear and easy-to-digest manner. Piece by piece, these articles demystify the world of probability and statistics, and unlock important foundations of knowledge for greater understanding in the future.

The first section, Mathematical Foundations of Probability and Statistics, begins with a focus on the mathematical principles essential for understanding these frequently overlapping concepts. In this section, many key concepts will be discussed, each providing readers with a solid grasp of the mathematical frameworks that are invaluable in moving forward with the study of probability and statistics. We will explore permutations and combinations, which help describe the many ways events such as a series of coin flips can occur. Notable types of probability distributions, which accurately depict the probabilities of different outcomes of events and are extremely important going forward in research, will be considered. We examine the concept of randomness in this section, and many other terms that readers may be familiar with. Since probability and statistics is meant to describe the real world, it is the bridge between math and language. As such, it is imperative that every term is understood strictly to its definition, since the slightest variation in verbiage can completely change the statement that is being made.

In the second section, Research Applications of Probability and Statistics, we shift our gaze to the applications of probability and statistics in research. Experts explore how scientists and researchers employ probabilistic reasoning in testing hypotheses, designing experiments, and navigating the intricacies of sampling methods and data analysis. Research is a battlefield of intellectual rigor, where steel sharpens steel. Studies on the same subject may come to opposite conclusions, and both have merit in the conversation. Proper methodology must be applied with excruciating attention to detail to collect results that are subject to as little bias as possible. Ego has no say here. After all, the goal of research is not for the researchers to prove themselves right; it is to collect data and subsequently offer an explanation for why the data might behave that way. This section of the volume reviews the tools and strategies that researchers need to know in order to deal with many variables in a multitude of data collecting situations while avoiding a preponderance of ways for the data to be skewed.

The final section, Real-World Applications of Probability and Statistics, bridges the gap between theory and everyday life, showcasing the tangible impact of probability and statistics in diverse practical settings. From the odds of sharing a birthday with another person to the statistical models shaping epidemiological studies, this section sheds light on just how widely applicable probability and statistics is in society. The emphasis here is on real-world scenarios, illustrating how mathematical concepts come to life in the hands of those seeking to unravel the mysteries of the world around them. Of course, since probability and statistics is such a widely useful tool, this volume offers just a selection of the innumerable ways that it can be applied to the real world. In the future, any curious person may go on to use probability and statistics in a way not previously done before in a new field. Therein lies the true beauty of math.

This volume aims to be an accessible compendium, inviting readers to engage with essays that balance mathematical rigor with real-world relevance. It serves as a valuable resource for those seeking a fundamental understanding of probability and statistics and its many applications without getting bogged down in complex mathematical derivations. Each section serves as a window into these disciplines, offering a practical and approachable exploration of the many ways that we use two of the most powerful mathematical tools: probability and statistics.

Students of business, economics, and mathematics in particular will find a roadmap for navigating the applications of probability and statistics that are contained in this volume. This volume presents interested scholars with an opportunity to explore the primary concepts that are necessary to employ statistical analysis in a broad variety of fields. Mathematicians with interest in probability and statistics find elegance and power embedded in the consideration of uncertainties, and are inspired to find clarity in complexity, and inspiration in the patterns that emerge from the seemingly random tapestry of data and uncertainty.

—Jake D. Nicosia, MS
—James F. Nicosia, PhD

Bayes' Theorem

Fields of Study: Algebra, Calculus, Connections, Data Analysis and Communication, Data Analysis and Probability, Experimental Design, Games and Gaming, Hypothesis Testing, Mathematics, Numbers and Operations, Probability Analysis, Quasi-experiment, Relational Analysis, Research Design, Research Theory, Sampling Design, Software Engineering, Statistical Analysis, Statistics, Variance Analysis

ABSTRACT

Bayes' theorem, developed in the 1700s by English mathematician Thomas Bayes, is a fundamental concept in the field of probability theory. Probability theory is a branch of mathematics that examines the likelihood that certain events will occur. Probability theorists use this theorem to examine a situation with possible outcomes and try to calculate how likely it is that one particular outcome will happen.

PRINCIPAL TERMS

Bayesian filtering: a means for monitoring the state of linear dynamical systems, Bayesian filtering relates the conditional probability of two events, thus establishing reasonable connections between terms and spam

outcomes: the total number of mathematically possible combinations resulting from a certain process or experiment

probability: in its simplest definition, the likelihood of a specific event happening, although determining a probability can be a very complex problem as the number of independent variables increases

INTRODUCTION

Bayes' theorem is named after its creator, Thomas Bayes, a noted English theologian and mathematician. Born in 1702, Bayes embarked on a life of both religion and science that led him to become one of the major founders of the mathematical discipline of statistics.

His most impactful work in this area was "Essay Towards Solving a Problem in the Doctrine of Chances," which was published posthumously in *Philosophical Transactions of the Royal Society*, two years after Bayes' 1761 death. In his work, Bayes demonstrates a mathematical technique to estimate the likelihood of a particular event occurring based on events that have occurred previously.

His work revolutionized math studies of the era, although later generations of mathematicians found some weak points in his theories and revised them into new forms. However, even in the twenty-first century, mathematicians acknowledge the important role Bayes played in bringing about the field of statistics and modernizing understanding of probability.

Equation 1

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Equation 2

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A^-)P(A^-) + P(B|A^+)P(A^+)}$$

OVERVIEW

Bayes' theorem is an important concept in probability theory. Probability theory is an area of mathematics that deals with probability, or the likelihood, that a phenomenon will lead to a particular outcome. This phenomenon, usually a random event, will have more than one, and sometimes many, possible outcomes.

In probability theory, mathematicians use rules and equations to determine the chance that a possible outcome will occur. This chance is described in a numerical way, typically the amount of a particular kind of outcome divided by the total number of potential outcomes. For example, flipping a coin may have two possible outcomes: heads or tails. The probability of getting heads is therefore one out of two, which may be written mathematically as 1/2 or 0.5.

This calculation always leads to a numeric ranking between zero and one. Zero indicates that a result has no likelihood of occurring, and one indicates that a result is definitely expected to occur. Results cannot be higher than one because the particular outcome being calculated cannot exceed the total number of possible outcomes.

Mathematicians may use probability theory in different ways, the main two of which are theoretical and experimental. In theoretical probability, mathematicians reach their conclusions mainly through logical reasoning. In experimental probability, mathematicians use experiments, usually several repeated tests, to reach their conclusions about results and likelihoods.

Probability theory has several different areas of specialization. The area most related to Bayes' theorem is known as conditional probability. This area focuses on events that may occur after an earlier

Portrait of Reverend Thomas Bayes, for whom the theorem was named. Image via Wikimedia Commons. [Public domain.]

event has already taken place. For example, it might study the possible outcomes of new government policies, such as calculating the likelihood of property values rising in a city if the mayor decides to bring in a new business.

Mathematicians use the equation $P(A|B)$ to show conditional probability. The vertical line used in math and logic represents the phrase "such that" or "it is true that." In this equation, P represents probability. A represents the possible outcome being calculated. B represents the important motivating event that has already taken place. Here, both A and B are independent events, meaning that the probability of one outcome is not dependent on the probability of the other outcome.

Using the previous example, A might be described as "property values rising" and B might be described as "the mayor brings in a new business." Therefore, the equation might be read as: "the probability that property values will rise if it is true that the mayor brings in a new business."

Bayes' theorem expands upon this basic equation, using the same elements of P, A, and B. This theorem may be represented as (see Equation 1). On the left is the original equation, "the probability of A occurring if B is true." This equation is equal to the more complex equation on the right. The equation on the right shows "the probability of B occurring if A is true," multiplied by "the probability of A occurring," with the result divided by "the probability of B occurring."

Complications may arise in the application of Bayes' theorem that require adjustments to the equation. One complication is that one of the events may be a binary variable, meaning it can only have one of two results instead of several or many possible outcomes. Some common examples of binary variables are "on or off," or "true or false." If the A event in the equation were a binary variable, a new and more complicated expression of the equation becomes necessary (see Equation 2).

Law of Large Numbers

Fields of Study: Algebra, Business, Calculus, Computer Science, Data Analysis and Communication, Data Analysis and Probability, Experimental Design, Hypothesis Testing, Mathematics, Numbers and Operations, Probability Analysis, Research Design, Sampling Design, Statistical Analysis, Statistics, Variance Analysis

ABSTRACT

The law of large numbers is a concept in probability theory stating that the more times a random action is repeated, the closer the average of the action's numerous outcomes will get to the true predicted average of those outcomes. The important point in the law of large numbers is that only through frequent repetition of a random action can someone observe the real probability of each of the action's outcomes. This is because an action repeated only a few times can appear to result in one-sided outcomes due to short-term deviations from the real average. The law of large numbers is often used to debunk beliefs that gamblers can predict the next outcome in games of chance based on what has already occurred. According to the law, an outcome's true probability of occurring will always determine what happens.

PRINCIPAL TERMS

probability: in its simplest definition, the likelihood of a specific event happening, although determining a probability can be a very complex problem as the number of independent variables increases

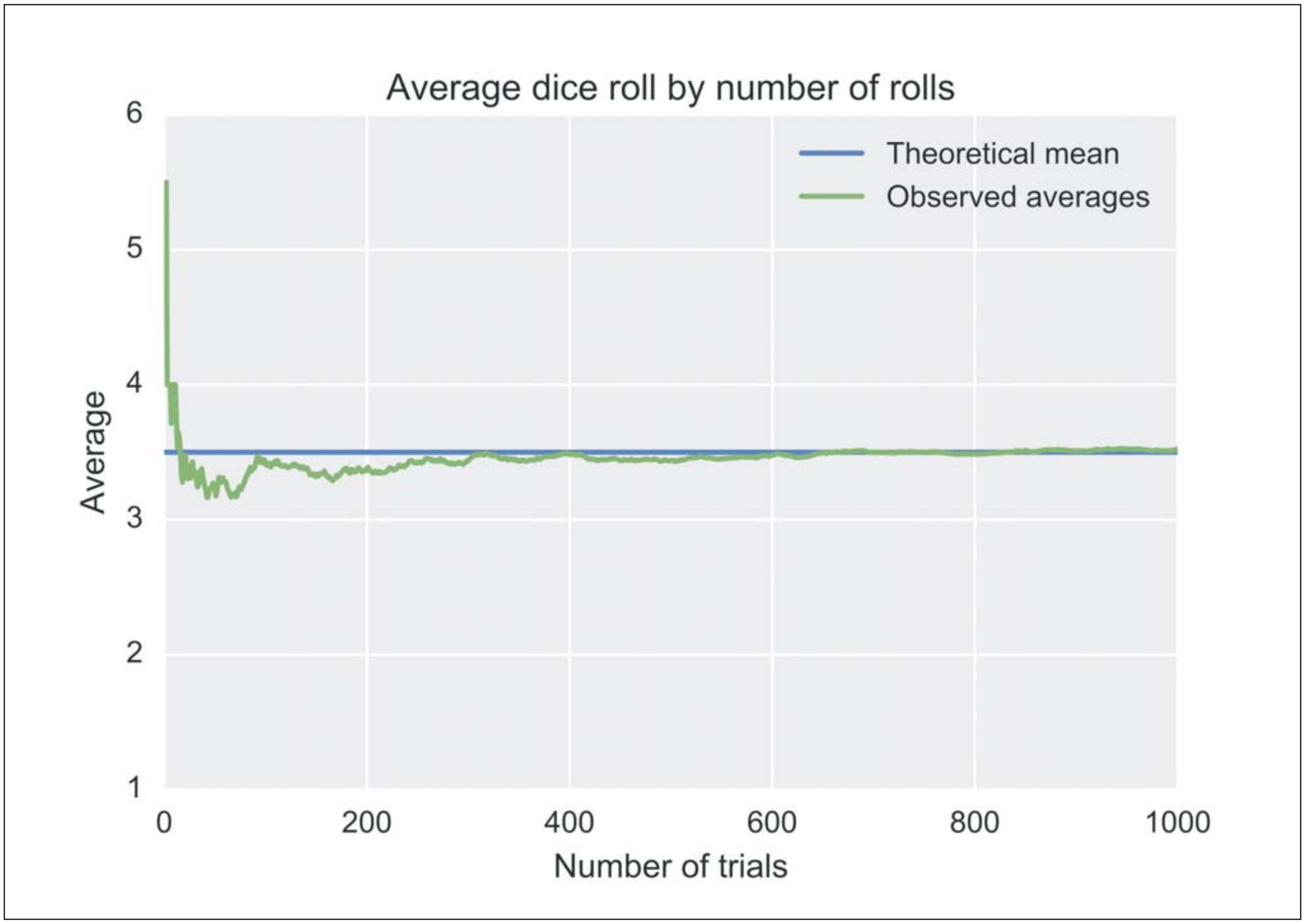
standard deviation: a measurement of the degree of dispersion present in a distribution; greater dispersion results in a larger standard deviation

INTRODUCTION

The law of large numbers states that the outcomes in a series of random events essentially have more time to reach the true probability of the outcomes occurring if the events are repeated a large number of times. For instance, the heads and tails sides of a flipped coin each have a 50 percent chance of landing face up. If someone flips a coin five or ten times, the coin may land with tails up every time. However, this only appears to violate the coin's 50 percent probability of landing with heads up. The same coin flipped one hundred times would yield a distribution of heads and tails much closer to the predicted 50 percent average. The distribution would be even closer after one thousand flips. Five or ten flips are not enough to indicate the true probability of the coin's landing on heads or tails.

As its name suggests, the law of large numbers applies only to a large number of random event repetitions. The pattern of an event's true probability cannot be observed in the short term. Many people misunderstand this aspect of the law by subscribing to the law of averages. This is the assumption that the outcomes of future random events will even out any recent deviations from the event's true average. The law of averages might lead someone to assume that a coin that has landed with tails up ten consecutive times is "due" to land on heads next. However, each coin flip is its own random event and is not influenced by any previous flips. The law of large numbers dictates that an outcome's probability of occurring is revealed only as the number of event repetitions increases.

The "gambler's fallacy" is a particular application of the law of averages. It refers to the wishful thinking of gamblers who assume that because they are on a losing streak, it is "time" for the game to turn in their favor. They misunderstand the law of large numbers by thinking that a losing streak must end so the event's average of outcomes can even out. But the law of large numbers states only that this average evens out as the number of event repetitions increases. Deviations can still occur in the short term. Slot machines and roulette wheels have the same odds of success every time they are played. The law



An illustration of the law of large numbers using a particular run of rolls of a single die. As the number of rolls in this run increases, the average of the values of all the results approaches 3.5. Although each run would show a distinctive shape over a small number of throws (at the left), over a large number of rolls (to the right) the shapes would be extremely similar. Image via Wikimedia Commons. [Public domain.]

of large numbers always applies, but slot machines theoretically could pay out jackpots two, three, or more times consecutively.

—*Michael Ruth*

Further Reading

"Gambler's Fallacy/Monte Carlo Fallacy." *Investopedia*, www.investopedia.com/terms/g/gamblersfallacy.asp.

Hebner, Mark. "The Gambler's Fallacy and the Misuse of the Law of Large Numbers." *Index Fund Advisors*, 21 Feb. 2014, www.ifa.com/articles/gambler_fallacy_misuse_large_numbers/.

Kiersz, Andy. "Apple CEO Tim Cook Said Something That Would Make Statisticians Cringe." *Business Insider*, 11 Feb. 2015, www.businessinsider.com/law-of-large-numbers-tim-cook-2015-2.

"Law of Large Numbers." *Investopedia*, www.investopedia.com/terms/l/lawoflargenumbers.asp.

Mass, A.J. "A Look at the Gambler's Fallacy." *ESPN*, 3 May 2010, www.espn.com/fantasy/baseball/flb/story?id=5157660.

Mauro, John. *Statistical Deception at Work*. John Wiley & Sons, 2008, p. 281.

Sehlinger, Bob, et al. *The Unofficial Guide to Las Vegas 2008*. Routledge, 1992, pp. 66–67.

Vedantam, Shankar. "How the Bias Known as Gambler's Fallacy Affects Our Lives." *NPR*, 29 Dec. 2015, www.npr.org/2015/12/29/461352879/how-the-bias-known-as-gamblers-fallacy-effects-our-lives.

Further Reading

Castagnoli, Erio. *Probability: A Brief Introduction*. Chicago: Independent Publishers, 2017.

Devlin, Keith. *The Unfinished Game: Pascal, Fermat, and the Seventeenth-Century Letter That Made the World Modern*. Basic Books, 2008.

Dunnington, Guy Waldo. *Carl Friedrich Gauss: Titan of Science*. Mathematical Association of America, 2004.

Gigerenzer, Gerd. *The Empire of Chance: How Probability Changed Science and Everyday Life*. Cambridge UP, 1990.

Hacking, Ian. *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*. Cambridge UP, 2006.

_____. *An Introduction to Probability and Inductive Logic*. Cambridge UP, 2001.

Hald, Anders. *A History of Probability and Statistics and Their Applications Before 1750*. Wiley-Interscience, 2003.

Salsburg, David. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. Holt, 2002.

Probability and Statistics

Fields of Study: Chemical Kinetics, Climate Modeling, Earth System Modeling, Environmental Studies, Fluid Dynamics, Mathematics, Physical Chemistry, Process Modeling, Statistics, Thermodynamics, Waste Management,

ABSTRACT

Probability and statistics are two related fields covering the science of collecting, measuring, and analyzing information in the form of numbers. Both probability and statistics are branches of applied mathematics. Probability focuses on using numeric data to predict a future outcome. Statistics incorporates theory into the gathering of numerical data and the drawing of accurate conclusions. Because nearly all fields in applied science rely on the analysis of numbers in some way, probability and statistics are one of the most diverse areas in terms of subjects and career paths. Statisticians also practice in areas of the academic world outside of science and throughout industry.

PRINCIPAL TERMS

confidence: a value assigned to a statistical result that indicates how close the calculated result is expected to be to the true value of the property

probabilism: the doctrine that probability is a sufficient basis for belief and action, since certainty in knowledge is unattainable

probability: in its simplest definition, the likelihood of a specific event happening, although determining a probability can be a very complex problem as the number of independent variables increases

statistics: the mathematical treatment of a set of collected values in order to determine essential ratios and trends

INTRODUCTION

Probability and statistics are two interconnected fields within applied mathematics. In both fields, principles of scientific theory are applied to the analysis of groups of data in the form of numbers. The main objective of probability and statistics is to ask and answer questions about data with as much accuracy as possible.

Defining "probability" can be a challenge, as multiple schools of thought exist. In one view, held by a group of scholars known as frequentists, probability is defined as the likelihood that a statement about a set of data will be true in the long run. Frequentists focus on the big picture, specifically at the collective outcome of multiple experiments conducted over time, rather than on specific data items or outcomes. In contrast, scholars known as Bayesians prefer to start with a probability-based assumption about a set of data, then test to see how close the actual data come to the initial assumption. On both sides of the debate, probabilists are seeking to understand patterns in data to predict how a population might behave in the future.

Statistics is a field with a broader scope than probability, but in some ways, it is easier to define. The academic discipline of statistics is based on the study of groups of numbers in three stages: collection, measurement, and analysis. At the collection stage, statistics involves issues such as the design of experiments and surveys. Statisticians must answer questions such as whether to examine an entire population or to work from a sample. Once the data are collected, statisticians must determine the level of measurement to be used and the types of questions that can be answered with validity based on the numbers.

BACKGROUND AND HISTORY

Probability has been a subject of interest since dice and card games were first played for money. Gambling inspired the first scholarly discussions of probability in the sixteenth and early seventeenth centuries. The Italian mathematician Gerolamo Cardano wrote *Libellus de ratiociniis in ludo aleae* (*The Value of All Chances in Games of Fortune*, 1714) in about 1565, although the work was not published until 1663. In the mid-1600s, French mathematicians Blaise Pascal and Pierre de Fermat discussed principles of probability in a series of letters about the gambling habits of a mutual friend.

The earliest history of a statistical study is less clear, but it is generally thought to involve demographics. British scholar John Graunt studied causes of mortality among residents of London and published his findings in 1662. Graunt found that statistical data could be biased by social factors, such as relatives' reluctance to report deaths due to syphilis. In 1710, John Arbuthnot analyzed the male-female ratio of babies born in Britain since 1629. His findings—that there were more males than females—were used to support his argument in favor of the existence of a divine being.

A third branch of statistics, the design of experiments and the problem of observational error, has its roots in the eighteenth-century work of German astronomer Tobias Mayer. However, a paper by British theologian Thomas Bayes published in 1764 after his death is considered a turning point in the history of probability and statistics. Bayes dealt with the question of how much confidence could be placed in the predictions of a mathematical model based on probability. The convergence between probability and statistics has increased over time. The development of modern computers has led to major advances in both fields.

Bernoulli's Ars Conjectandi *was the first work that dealt with probability theory as currently understood. Image via Wikimedia Commons. [Public domain.]*

HOW IT WORKS

In terms of scope, probability and statistics are some of the widest, most diverse fields in applied mathematics. If a research project involves items that must be counted or measured in some way, statistics will be part of the analysis. It is common to associate statistics with research in the sciences, but an art historian tracking changes in the use of color in painting from one century to another is just as likely to use a

Betting and Fairness

Fields of Study: Algebra, Business, Calculus, Data Analysis and Probability, Economics, Games and Gaming, Hypothesis Testing, Mathematics, Numbers and Operations, Probability Analysis, Recreation, Sports, Statistics

ABSTRACT
Mathematics is used to analyze betting and probabilities for games of chance and for investing in the stock market.

PRINCIPAL TERMS

expected value: the long-term average of a variable calculated over a large number of arithmetic measurements

probability: in its simplest definition, the likelihood of a specific event happening, although determining a probability can be a very complex problem as the number of independent variables increases

INTRODUCTION

A pivotal moment in the early development of probability occurred in 1654, as the French mathematicians Blaise Pascal and Pierre de Fermat exchanged a series of letters. Pascal and Fermat were wrestling with questions involving the fair payoff for a gambler who is forced to quit in the middle of a game. In modern language, they were calculating the "expected value" of the game's payoff (the average payoff under the various possible outcomes, weighted according to the likelihood of those outcomes). A bet is said to be "fair" if the price of placing it is equal to the expected value of the payoff. Betting plays an integral part in our modern society. People place bets in casinos and at sporting events, as well as by buying lottery tickets. They are also placing bets when purchasing insurance or investing in the stock market. Some of these bets are fair, some are unfair, and some cannot be objectively categorized.

Coin flip. Photo via iStock/alexeys. [Used under license.]

The primary problem that Pascal and Fermat solved (each employing a different method) can be used to illustrate some important ideas on fairness. In the problem, two gamblers are playing a game in which a coin is repeatedly tossed. The game is interrupted at a point where 2 more heads are required for Player A to win and 3 more tails are required for Player B to win (whichever occurs first). How should the potential winnings be divided at this stage of the game?

Fermat solved the problem by observing that at most 4 tosses remain in order to identify the winner, and that there are 16 equally likely ways in which 4 tosses could occur:

HHHH, HHHT, HHTH, HTHH, THHH, HHTT, HTHT, HTTH, THHT, THTH, TTHH, HTTT, THTT, TTHT, TTTH, and TTTT.

In 11 of these possibilities (the first 11 items on the list), Player A would win, because 2 heads occur before 3 tails; in the other 5 possibilities, Player B would win, because 3 tails occur first. Therefore, Fermat reasoned that Player A should receive 11/16 of the winnings, and Player B should receive 5/16 of the winnings. In modern language, Player A would win the game with probability 11/16 and Player B would win with probability 5/16; Fermat was calculating the "expected value" of the winnings for each player.

Suppose that up to this point in their game, neither Player A nor B has paid any money for the opportunity to play, but that they are now required to pay a total of $1, altogether, and that this dollar will constitute the winnings. Fermat's solution to the previous problem allows for a fair method of dividing the payment: Player A should pay 11/16 of the dollar and Player B should pay 5/16, so that the payments match the expected winnings. In other words, if the game is being played for a $1 payoff, then the price for a fair bet is 11/16 of a dollar for Player A and 5/16 of a dollar for Player B.

Blaise Pascal, portrait. Image by Palace of Versailles, via Wikimedia Commons.

Pierre de Fermat, portrait. Image via Wikimedia Commons. [Public domain.]

LOTTERIES AND CASINOS

State-run lotteries are unfair to the player who purchases a ticket, because some of the revenue goes to the state and is not redistributed to the winner(s). Of course, even if all of the ticket revenue were paid to the winner(s)—so that the bets were fair—a lottery would be unfavorable to almost every player. None-