IS GOD A MATHEMATICIAN?

is clearly even more complex because of a multiplicity of interrelated elements, many of which are highly uncertain at best. The researchers of the seventeenth century realized soon enough that a search for precise universal social principles of the type of Newton's law of gravitation was doomed from the start. For a while, it seemed that when the intricacies of human nature are brought into the equation, secure predictions become virtually impossible. The situation appeared to be even more hopeless when the minds of an entire population were involved. Rather than despairing, however, a few ingenious thinkers developed a fresh arsenal of innovative mathematical tools—*statistics* and *probability theory.*

The Odds Beyond Death and Taxes

The English novelist Daniel Defoe (1660–1731), best known for his adventure story *Robinson Crusoe,* also authored a work on the supernatural entitled *The Political History of the Devil.* In it, Defoe, who saw evidence for the devil's actions everywhere, wrote: "Things as certain as death and taxes, can be more firmly believed." Benjamin Franklin (1706–90) seems to have subscribed to the same perspective with respect to certainty. In a letter he wrote at age eighty-three to the French physicist Jean-Baptiste Leroy, he said: "Our Constitution is in actual operation. Everything appears to promise that it will last; but in this world nothing can be said to be certain but death and taxes." Indeed, the courses of our lives appear to be unpredictable, prone to natural disasters, susceptible to human errors, and affected by pure happenstance. Phrases such as "[- - - -] happens" have been invented precisely to express our vulnerability to the unexpected and our inability to control chance. In spite of these obstacles, and maybe even because of these challenges, mathematicians, social scientists, and biologists have embarked since the sixteenth century on serious attempts to tackle uncertainties methodically. Following the establishment of the field of statistical mechanics, and faced with the realization that the very foundations of physics—in the form of quantum mechanics—are based on uncertainty, physicists of the twentieth and twenty-first centuries have enthusiastically joined the battle. The weapon researchers use to combat the lack of precise determinism is the ability to calculate the odds of a particular outcome. Short of being capable of actually predicting a result, computing the likelihood of different consequences is the next best thing. The tools that have been fashioned to improve on mere guesses and speculations—statistics and probability theory—provide the underpinning of not just much of modern science, but also a wide range of social activities, from economics to sports.

We all use probabilities and statistics in almost every decision we make, sometimes subconsciously. For instance, you probably don't know that the number of fatalities from automobile accidents in the U.S. was 42,636 in 2004. However, had that number been, say, 3 million, I'm sure you would have known about it. Furthermore, this knowledge would have probably caused you to think twice before getting into the car in the morning. Why do these precise data on road fatalities give us some confidence in our decision to drive? As we shall see shortly, a key ingredient to their reliability is the fact that they are based on very large numbers. The number of fatalities in Frio Town, Texas, with a population of forty-nine in 1969 would hardly have been equally convincing. Probability and statistics are among the most important arrows for the bows of economists, political consultants, geneticists, insurance companies, and anybody trying to distill meaningful conclusions from vast amounts of data. When we talk about mathematics permeating even disciplines that were not originally under the umbrella of the exact sciences, it is often through the windows opened by probability theory and statistics. How did these fruitful fields emerge?

Statistics—a term derived from the Italian *stato* (state) and *statista* (a person dealing with state affairs)—first referred to the simple collection of facts by government officials. The first important work on statistics in the modern sense was carried out by an unlikely researcher—a shopkeeper in seventeenth century London. John Graunt (1620–74) was trained to sell buttons, needles, and drapes. Since his job afforded him a considerable amount of free time, Graunt studied Latin and French on his own and started to take interest in the Bills of Mortality—weekly numbers of deaths parish by parish—that had been published in London since 1604. The process of issuing these reports was established mainly in order to provide an early warning signal for devastating epidemics. Using those crude numbers, Graunt started to make interesting observations that he eventually published in a small, eighty-five-page book entitled *Natural and Political Observations Mentioned in a Following Index, and Made upon the Bills of Mortality.* Figure 32 presents an example of a table from Graunt's book, where no fewer than sixty-three diseases and casualties were listed alphabetically. In a dedication to the president of the Royal Society, Graunt points out that since his work concerns "the Air, Countries, Seasons, Fruitfulness, Health, Diseases, Longevity, and the proportion between the Sex and Ages of Mankind," it is really a treatise in

(9)

The Diſeaſes, and Caſualties this year being 1632.

ABortive, and Stilborn	445	Jaundies	43
Affrighted	1	Jawfaln	8
Aged	628	Impoſtume	74
Ague	43	Kil'd by ſeveral accidents	46
Apoplex, and Meagrom	17	King's Evil	38
Bit with a mad dog	1	Lethargie	2
Bleeding	3	Livergrown	87
Bloody flux, ſcowring, and flux	348	Lunatique	5
Bruſed, Iſſues, ſores, and ulcers,	28	Made away themſelves	15
Burnt, and Scalded	5	Meaſles	80
Burſt, and Rupture	9	Murthered	7
Cancer, and Wolf	10	Over-laid, and ſtarved at nurſe	7
Canker	1	Palſie	25
Childbed	171	Piles	1
Chriſomes, and Infants	2268	Plague	8
Cold, and Cough	55	Planet	13
Colick, Stone, and Strangury	56	Pleuriſie, and Spleen	36
Conſumption	1797	Purples, and ſpotted Feaver	38
Convulſion	241	Quinſie	7
Cut of the Stone	5	Riſing of the Lights	98
Dead in the ſtreet, and ſtarved	6	Sciatica	1
Dropſie, and Swelling	267	Scurvey, and Itch	9
Drowned	34	Suddenly	62
Executed, and preſt to death	18	Surfet	86
Falling Sickneſs	7	Swine Pox	6
Fever	1108	Teeth	470
Fiſtula	13	Thruſh, and Sore mouth	40
Flocks, and ſmall Pox	531	Tympany	13
French Pox	12	Tiſſick	34
Gangrene	5	Vomiting	1
Gout	4	Worms	27
Grief	11		

Chriſtened	Males—4994 Females-4590 In all—9584	Buried	Males—4932 Females—4603 In all—9535	Whereof, of the Plague-8

Increaſed in the Burials in the 122 Pariſhes, and at the Peſthouſe this year 993
Decreaſed of the Plague in the 122 Pariſhes, and at the Peſthouſe this year, 266

C 7 In

Figure 32

natural history. Indeed, Graunt did much more than merely collect and present the data. By examining, for instance, the average numbers of christenings and burials for males and females in London and in the country parish Romsey in Hampshire, he demonstrated for the first time the stability of the sex ratio at birth. Specifically, he found that in London there were thirteen females born for every fourteen males and in Romsey fifteen females for sixteen males. Remarkably, Graunt had the foresight to express the wish that "travellers would enquire whether it be the same in other countries." He also noted that "it is a blessing to Man-kind, that by this overplus of *Males* there is this natural Bar to *Polygamy*: for in such a state Women could not live in that parity, and equality of expence with their Husbands, as now, and here they do." Today, the commonly assumed ratio between boys and girls at birth is about 1.05. Traditionally the explanation for this excess of males is that Mother Nature stacks the deck in favor of male births because of the somewhat greater fragility of male fetuses and babies. Incidentally, for reasons that are not entirely clear, in both the United States and Japan the proportion of baby boys has fallen slightly each year since the 1970s.

Another pioneering effort by Graunt was his attempt to construct an age distribution, or a "life table," for the living population, using the data on the number of deaths according to cause. This was clearly of great political importance, since it had implications for the number of fighting men—men between sixteen and fifty-six years of age—in the population. Strictly speaking, Graunt did not have sufficient information to deduce the age distribution. This is precisely where, however, he demonstrated ingenuity and creative thinking. Here is how he describes his estimate of childhood mortality:

> Our first Observation upon the Casualties shall be, that in twenty Years there dying of all diseases and Casualties, 229,250, that 71,124 dyed of the Thrush, Convulsion, Rickets, Teeths, and Worms; and as Abortives, Chrysomes, Infants, Livergrown, and Overlaid; that is to say, that about ⅓ of the whole died of those diseases, which we guess did all light upon Children under four or five Years old. There died also of the Small-Pox,

> Swine-Pox, and Measles, and of Worms without Convulsions, 12,210, of which number we suppose likewise that about ½ might be Children under six Years old. Now, if we consider that 16 of the said 229 thousand died of that extraordinary and grand Casualty the Plague, we shall finde that about thirty six percentum of all quick conceptions, died before six years old."

In other words, Graunt estimated the mortality before age six to be $(71{,}124 + 6{,}105) \div (229{,}250 - 16{,}000) = 0.36$. Using similar arguments and educated guesses, Graunt was able to estimate the old-age mortality. Finally, he filled the gap between ages six and seventy-six by a mathematical assumption about the behavior of the mortality rate with age. While many of Graunt's conclusions were not particularly sound, his study launched the science of statistics as we know it. His observation that the percentages of certain events previously considered purely a matter of chance or fate (such as deaths caused by various diseases) in fact showed an extremely robust regularity, introduced scientific, quantitative thinking into the social sciences.

The researchers who followed Graunt adopted some aspects of his methodology, but also developed a better mathematical understanding of the use of statistics. Surprisingly perhaps, the person who made the most significant improvements to Graunt's life table was the astronomer Edmond Halley—the same person who persuaded Newton to publish his *Principia.* Why was everybody so interested in life tables? Partly because this was, and still is, the basis for life insurance. Life insurance companies (and indeed gold diggers who marry for money!) are interested in such questions as: If a person lived to be sixty, what is the probability that he or she would also live to be eighty?

To construct his life table, Halley used detailed records that were kept at the city of Breslau in Silesia since the end of the sixteenth century. A local pastor in Breslau, Dr. Caspar Neumann, was using those lists to suppress superstitions in his parish that health is affected by the phases of the Moon or by ages that are divisible by seven and nine. Eventually, Halley's paper, which had the rather long title of "An Estimate of the Degrees of the Mortality of Mankind, drawn from curious Tables of the Births and Funerals at the City of Breslaw; with an Attempt to ascertain the Price of Annuities upon Lives," became the basis for the mathematics of life insurance. To get an idea of how insurance companies may assess their odds, examine Halley's life table below:

Halley's Life Table

AGE CURRENT	PERSONS	AGE CURRENT	PERSONS	AGE CURRENT	PERSONS
1	1000	11	653	21	592
2	855	12	646	22	586
3	798	13	640	23	579
4	760	14	634	24	573
5	732	15	628	25	567
6	710	16	622	26	560
7	692	17	616	27	553
8	680	18	610	28	546
9	670	19	604	29	539
10	661	20	598	30	531
AGE CURRENT	PERSONS	AGE CURRENT	PERSONS	AGE CURRENT	PERSONS
31	523	41	436	51	335
32	515	42	427	52	324
33	507	43	417	53	313
34	499	44	407	54	302
35	490	45	397	55	292
36	481	46	387	56	282
37	472	47	377	57	272
38	463	48	367	58	262
39	454	49	357	59	252
40	445	50	346	60	242
AGE CURRENT	PERSONS	AGE CURRENT	PERSONS	AGE CURRENT	PERSONS
61	232	71	131	81	34
62	222	72	120	82	28
63	212	73	109	83	23
64	202	74	98	84	20
65	192	75	88		
66	182	76	78		
67	172	77	68		
68	162	78	58		
69	152	79	49		
70	142	80	41		

The table shows, for instance, that of 710 people alive at age six, 346 were still alive at age fifty. One could then take the ratio of 346/710 or 0.49 as an estimate of the probability that a person of age six would live to be fifty. Similarly, of 242 at age sixty, 41 were alive at age eighty. The probability of making it from sixty to eighty could then be estimated to be 41/242, or about 0.17. The rationale behind this procedure is simple. It relies on past experience to determine the probability of various future events. If the sample on which the experience is predicated is sufficiently large (Halley's table was based on a population of about 34,000), and if certain assumptions hold (such as that the mortality rate is constant over time), then the calculated probabilities are fairly reliable. Here is how Jakob Bernoulli described the same problem:

> What mortal, I ask, could ascertain the number of diseases, counting all possible cases, that afflict the human body in every one of its many parts and at every age, and say how much more likely one disease is to be fatal than another . . . and on that basis make a prediction about the relationship between life and death in future generations?

After concluding that this and similar forecasts "depend on factors that are completely obscure, and which constantly deceive our senses by the endless complexity of their interrelationships," Bernoulli also suggested a statistical/probabilistic approach:

> There is, however, another way that will lead us to what we are looking for and enable us at least to ascertain *a posteriori* what we cannot determine *a priori,* that is, to ascertain it from the results observed in numerous similar instances. It must be assumed in this connection that, under similar conditions, the occurrence (or nonoccurrence) of an event in the future will follow the same pattern as was observed for like events in the past. For example, if we have observed that out of 300 persons of the same age and with the same constitution as a certain *Titius,* 200 died within ten years while the rest survived, we can with reasonable certainty conclude that there are twice

> as many chances that Titius also will have to pay his debt to nature within the ensuing decade as there are chances that he will live beyond that time.

Halley followed his mathematical articles on mortality with an interesting note that had more philosophical overtones. One of the passages is particularly moving:

> Besides the uses mentioned in my former, it may perhaps not be an unacceptable thing to infer from the same Tables, how unjustly we repine at the shortness of our lives, and think our selves wronged if we attain not Old Age; whereas it appears hereby, that the one half of those that are born are dead in Seventeen years time, 1238 being in that time reduced to 616. So that instead of murmuring at what we call an untimely Death, we ought with Patience and unconcern to submit to that Dissolution which is the necessary Condition of our perishable Materials, and of our nice and frail Structure and Composition: And to account it as Blessing that we have survived, perhaps by many Years, that Period of Life, whereat the one half of the whole Race of Mankind does not arrive.

While the situation in much of the modern world has improved significantly compared to Halley's sad statistics, this is unfortunately not true for all countries. In Zambia, for instance, the mortality for ages five and under in 2006 has been estimated at a staggering 182 deaths per 1,000 live births. The life expectancy in Zambia remains at a heartbreaking low of thirty-seven years.

Statistics, however, are not concerned only with death. They penetrate into every aspect of human life, from mere physical traits to intellectual products. One of the first to recognize the power of statistics to potentially produce "laws" for the social sciences was the Belgian polymath Lambert-Adolphe-Jacques Quetelet (1796–1874). He, more than anyone else, was responsible for the introduction of the common statistical concept of the "average man," or what we would refer to today as the "average person."

The Average Person

Adolphe Quetelet was born on February 22, 1796, in the ancient Belgian town of Ghent. His father, a municipal officer, died when Adolphe was seven years old. Compelled to support himself early in life, Quetelet started to teach mathematics at the young age of seventeen. When not on duty as an instructor, he composed poetry, wrote the libretto for an opera, participated in the writing of two dramas, and translated a few literary works. Still, his favorite subject remained mathematics, and he was the first to graduate with the degree of doctor of science from the University of Ghent. In 1820, Quetelet was elected as a member of the Royal Academy of Sciences in Brussels, and within a short time he became the academy's most active participant. The next few years were devoted mostly to teaching and to the publication of a few treatises on mathematics, physics, and astronomy.

Quetelet used to open his course on the history of science with the following insightful observation: "The more advanced the sciences become, the more they have tended to enter the domain of mathematics, which is a sort of center towards which they converge. We can judge of the perfection to which a science has come by the facility, more or less great, with which it may be approached by calculation."

In December of 1823, Quetelet was sent to Paris at the state's expense, mostly to study observational techniques in astronomy. As it turned out, however, this three-month visit to the then mathematical capital of the world veered Quetelet in an entirely different direction—the theory of probability. The person who was mostly responsible for igniting Quetelet's enthusiastic interest in this subject was Laplace himself. Quetelet later summarized his experience with statistics and probability:

> Chance, that mysterious, much abused word, should be considered only a veil for our ignorance; it is a phantom which exercises the most absolute empire over the common mind, accustomed to consider events only as isolated, but which is reduced to naught before the philosopher, whose eye embraces a long series of events and whose penetration is not led astray

by variations, which disappear when he gives himself sufficient perspective to seize the laws of nature.

The importance of this conclusion cannot be overemphasized. Quetelet essentially denied the role of chance and replaced it with the bold (even though not entirely proven) inference that even social phenomena have causes, and that the regularities exhibited by statistical results can be used to uncover the rules underlying social order.

In an attempt to put his statistical approach to the test, Quetelet started an ambitious project of collecting thousands of measurements related to the human body. For instance, he studied the distributions of the chest measurements of 5,738 Scottish soldiers and of the heights of 100,000 French conscripts by plotting separately the frequency with which each human trait occurred. In other words, he represented graphically how many conscripts had heights between, say, five feet and five feet two inches, and then between five feet two inches and five feet four inches, and so on. He later constructed similar curves even for what he called "moral" traits for which he had sufficient data. The latter qualities included suicides, marriages, and the propensity to crime. To his surprise, Quetelet discovered that all the human characteristics followed what is now known as the *normal* (or *Gaussian,* named somewhat unjustifiably after the "prince of mathematics" Carl Friedrich Gauss), bell-shaped frequency distribution (figure 33). Whether it was heights, weights, measurements of limb lengths, or even intellectual qualities determined by what were then pioneering psychological tests, the same type of curve appeared again and again. The curve itself was not new to Quetelet—mathematicians and physicists recognized it from the mid-eighteenth century, and Quetelet was familiar with it from his astronomical work—it was just

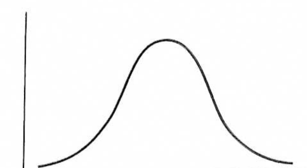


Figure 33

the association of this curve with human characteristics that came as somewhat of a shock. Previously, this curve had been known as the *error curve*, because of its appearance in any type of errors in measurements.

Imagine, for instance, that you are interested in measuring very accurately the temperature of a liquid in a vessel. You can use a high-precision thermometer and over a period of one hour take one thousand consecutive readings. You will find that due to random errors and possibly some fluctuations in the temperature, not all measurements will give precisely the same value. Rather, the measurements would tend to cluster around a central value, with some measurements giving temperatures that are higher and others that are lower. If you plot the number of times that each measurement occurred against the value of the temperature, you will obtain the same type of bell-shaped curve that Quetelet found for the human characteristics. In fact, the larger the number of measurements performed on any physical quantity, the closer will the obtained frequency distribution approximate the normal curve. The immediate implication of this fact for the question of the unreasonable effectiveness of mathematics is quite dramatic in itself—even human errors obey some strict mathematical rules.

Quetelet thought that the conclusions were even more far-reaching. He regarded the finding that human characteristics followed the error curve as an indication that the "average man" was in fact a type that nature was trying to produce. According to Quetelet, just as manufacturing errors would create a distribution of lengths around the average (correct) length of a nail, nature's errors were distributed around a preferred biological type. He declared that the people of a nation were clustered about their average "as if they were the results of measurements made on one and the same person, but with instruments clumsy enough to justify the size of the variation."

Clearly, Quetelet's speculations went a bit too far. While his discovery that biological characteristics (whether physical or mental) are distributed according to the normal frequency curve was extremely important, this could neither be taken as proof for nature's intentions nor could individual variations be treated as mere mistakes. For instance, Quetelet found the average height of the French conscripts to be five feet four inches. At the low end, however, he found a man of one foot five inches. Obviously one could not make an error of almost four feet in measuring the height of a man five feet four inches tall.

Even if we ignore Quetelet's notion of "laws" that fashion humans in a single mold, the fact that the distributions of a variety of traits ranging from weights to IQ levels all follow the normal curve is in itself pretty remarkable. And if that is not enough, even the distribution of major-league batting averages in baseball is reasonably normal, as is the annual rate of return on stock indexes (which are composed of many individual stocks). Indeed, distributions that deviate from the normal curve sometimes call for a careful examination. For instance, if the distribution of the grades in English in some school were found not to be normal, this could provoke an investigation into the grading practices of that school. This is not to say that all distributions are normal. The distribution of the lengths of words that Shakespeare used in his plays is not normal. He used many more words of three and four letters than words of eleven or twelve letters. The annual household income in the United States is also represented by a non-normal distribution. In 2006, for instance, the top 6.37% of households earned roughly one third of all income. This fact raises an interesting question in itself: If both the physical and the intellectual characteristics of humans (which presumably determine the potential for income) are normally distributed, why isn't the income? The answer to such socioeconomic questions is, however, beyond the scope of the present book. From our present limited perspective, the amazing fact is that essentially all the physically measurable particulars of humans, or of animals and plants (of any given variety) are distributed according to just one type of mathematical function.

Human characteristics served historically not only as the basis for the study of the statistical frequency distributions, but also for the establishment of the mathematical concept of *correlation*. The correlation measures the degree to which changes in the value of one variable are accompanied by changes in another. For instance, taller women may be expected to wear larger shoes. Similarly, psychologists found a correlation between the intelligence of parents and the degree to which their children succeed in school.

The concept of a correlation becomes particularly useful in those situations in which there is no precise functional dependence between the two variables. Imagine, for example, that one variable is the maximal daytime temperature in southern Arizona and the other is the number of forest fires in that region. For a given value of the temperature, one cannot predict precisely the number of forest fires that will break out, since the latter depends on other variables such as the humidity and the number of fires started by people. In other words, for any value of the temperature, there could be many corresponding numbers of forest fires and vice versa. Still, the mathematical concept known as the *correlation coefficient* allows us to measure quantitatively the strength of the relationship between two such variables.

The person who first introduced the tool of the correlation coefficient was the Victorian geographer, meteorologist, anthropologist, and statistician Sir Francis Galton (1822–1911). Galton—who was, by the way, the half-cousin of Charles Darwin—was not a professional mathematician. Being an extraordinarily practical man, he usually left the mathematical refinements of his innovative concepts to other mathematicians, in particular to the statistician Karl Pearson (1857–1936). Here is how Galton explained the concept of correlation:

> The length of the cubit [the forearm] is correlated with the stature, because a long cubit usually implies a tall man. If the correlation between them is very close, a very long cubit would usually imply a very tall stature, but if it were not very close, a very long cubit would be on the average associated with only a tall stature, and not a very tall one; while, if it were *nil,* a very long cubit would be associated with no especial stature, and therefore, on the average, with mediocrity.

Pearson eventually gave a precise mathematical definition of the correlation coefficient. The coefficient is defined in such a way that when the correlation is very high—that is, when one variable closely follows the up-and-down trends of the other—the coefficient takes the value of 1. When two quantities are *anticorrelated,* meaning that when one increases the other decreases and vice versa, the coefficient is equal to −1. Two variables that each behave as if the other didn't even exist have a correlation coefficient of 0. (For instance, the behavior of some governments unfortunately shows almost zero correlation with the wishes of the people whom they supposedly represent.)

Modern medical research and economic forecasting depend crucially on identifying and calculating correlations. The links between smoking and lung cancer, and between exposure to the Sun and skin cancer, for instance, were established initially by discovering and evaluating correlations. Stock market analysts are constantly trying to find and quantify correlations between market behavior and other variables; any such discovery can be enormously profitable.

As some of the early statisticians readily realized, both the collection of statistical data and their interpretation can be very tricky and should be handled with the utmost care. A fisherman who uses a net with holes that are ten inches on a side might be tempted to conclude that all fish are larger than ten inches, simply because the smaller ones would escape from his net. This is an example of *selection effects*—biases introduced in the results due to either the apparatus used for collecting the data or the methodology used to analyze them. Sampling presents another problem. For instance, modern opinion polls usually interview no more than a few thousand people. How can the pollsters be sure that the views expressed by members of this sample correctly represent the opinions of hundreds of millions? Another point to realize is that correlation does not necessarily imply causation. The sales of new toasters may be on the rise at the same time that audiences at concerts of classical music increase, but this does not mean that the presence of a new toaster at home enhances musical appreciation. Rather, both effects may be caused by an improvement in the economy.

In spite of these important caveats, statistics have become one of the most effective instruments in modern society, literally putting the "science" into the social sciences. But why do statistics work at all? The answer is given by the mathematics of *probability*, which reigns over many facets of modern life. Engineers trying to decide which safety mechanisms to install into the Crew Exploration Vehicle for astronauts, particle physicists analyzing results of accelerator experiments, psychologists rating children in IQ tests, drug companies evaluating the efficacy of new medications, and geneticists studying human heredity all have to use the mathematical theory of probability.

Games of Chance

The serious study of probability started from very modest beginnings—attempts by gamblers to adjust their bets to the odds of success. In particular, in the middle of the seventeenth century, a French nobleman—the Chevalier de Méré—who was also a reputed gamester, addressed a series of questions about gambling to the famous French mathematician and philosopher Blaise Pascal (1623–62). The latter conducted in 1654 an extensive correspondence about these questions with the other great French mathematician of the time, Pierre de Fermat (1601–65). The theory of probability was essentially born in this correspondence.

Let's examine one of the fascinating examples discussed by Pascal in a letter dated July 29, 1654. Imagine two noblemen engaged in a game involving the roll of a single die. Each player has put on the table thirty-two pistoles of gold. The first player chose the number 1, and the second chose the number 5. Each time the chosen number of one of the players turns up, that player gets one point. The winner is the first one to have three points. Suppose, however, that after the game has been played for some time, the number 1 has turned up twice (so that the player who had chosen that number has two points), while the number 5 has turned up only once (so the opponent has only one point). If, for whatever reason, the game has to be interrupted at that point, how should the sixty-four pistoles on the table be divided between the two players? Pascal and Fermat found the mathematically logical answer. If the player with two points were to win the next roll, the sixty-four pistoles would belong to him. If the other player were to win the next roll, each player would have had two points, and so each would have gotten thirty-two pistoles. Therefore, if the players separate without playing the next roll, the first player could correctly argue: "I am certain of thirty-two pistoles even if I lose this roll, and as for the other thirty-two pistoles perhaps I shall have them and perhaps you will have them; the chances are equal. Let us then divide these thirty-two pistoles equally and give me also the thirty-two pistoles of which I am certain." In other words, the first player should get forty-eight pistoles and the other sixteen pistoles. Unbelievable, isn't it, that a new, deep mathematical discipline could have emerged from this type of apparently trivial discussion? This is, however, precisely the reason why the effectiveness of mathematics is as "unreasonable" and mysterious as it is.

The essence of probability theory can be gleaned from the following simple facts. No one can predict with certainty which face a fair coin tossed into the air will show once it lands. Even if the coin has just come up heads ten times in a row, this does not improve our ability to predict with certainty the next toss by one iota. Yet we can predict with certainty that if you toss that coin ten million times, very close to half the tosses will show heads and very close to half will show tails. In fact, at the end of the nineteenth century, the statistician Karl Pearson had the patience to toss a coin 24,000 times. He obtained heads in 12,012 of the tosses. This is, in some sense, what probability theory is really all about. Probability theory provides us with accurate information about the collection of the results of a large number of experiments; it can never predict the result of any specific experiment. If an experiment can produce n possible outcomes, each one having the same chance of occurring, then the probability for each outcome is $1/n$. If you roll a fair die, the probability of obtaining the number 4 is 1/6, because the die has six faces, and each face is an equally likely outcome. Suppose you rolled the die seven times in a row and each time you got a 4, what would be the probability of getting a 4 in the next throw? Probability theory gives a crystal-clear answer: The probability would still be 1/6—the die has no memory and any notions of a "hot hand" or of the next roll making up for the previous imbalance are only myths. What is true is that if you were to roll the die a million times, the results will average out and 4 would appear very close to one-sixth of the time.

Let's examine a slightly more complex situation. Suppose you simultaneously toss three coins. What is the probability of getting two tails and one head? We can find the answer simply by listing all the possible outcomes. If we denote heads by "H" and tails by "T," then there are eight possible outcomes: TTT, TTH, THT, THH, HTT, HTH, HHT, HHH. Of these, you can check that three are favorable to the event "two tails and one head." Therefore, the probability for this event is 3/8. Or more generally, if out of n outcomes of equal chances, m are favorable to the event you are interested in, then the probability for that event to happen is m/n. Note that this means that the probability always takes a value between zero and one. If the event you are interested in is in fact impossible, then $m = 0$ (no outcome is favorable) and the probability would be zero. If, on the other hand, the event is absolutely certain, that means that all n events are favorable ($m = n$) and the probability is then simply $n/n = 1$. The results of the three coin tosses demonstrate yet another important result of probability theory—if you have several events that are entirely *independent* of each other, then the probability of all of them happening is the product of the individual probabilities. For instance, the probability of obtaining three heads is 1/8, which is the product of the three probabilities of obtaining heads in each of the three coins: $1/2 \times 1/2 \times 1/2 = 1/8$.

OK, you may think, but other than in casino games and other gambling activities, what additional uses can we make of these very basic probability concepts? Believe it or not, these seemingly insignificant probability laws are at the heart of the modern study of genetics—the science of the inheritance of biological characteristics.

The person who brought probability into genetics was a Moravian priest. Gregor Mendel (1822–84) was born in a village near the border between Moravia and Silesia (today Hynčice in the Czech Republic). After entering the Augustinian Abbey of St. Thomas in Brno, he studied zoology, botany, physics, and chemistry at the University of Vienna. Upon returning to Brno, he began an active experimentation with pea plants, with strong support from the abbot of the Augustinian monastery. Mendel focused his research on pea plants because they were easy to grow, and also because they have both male and female reproductive organs. Consequently, pea plants can be either self-pollinated or cross-pollinated with another plant.

By cross-pollinating plants that produce only green seeds with plants that produce only yellow seeds, Mendel obtained results that at first glance appeared to be very puzzling (figure 34). The first offspring generation had only yellow seeds. However, the following generation consistently had a 3:1 ratio of yellow to green seeds! From these surprising findings, Mendel was able to distill three conclusions that became important milestones in genetics:

1. The inheritance of a characteristic involves the transmittance of certain "factors" (what we call *genes* today) from parents to offspring.
2. Every offspring inherits one such "factor" from each parent (for any given trait).
3. A given characteristic may not manifest itself in an offspring but it can still be passed on to the following generation.

But how can one explain the quantitative results in Mendel's experiment? Mendel argued that each of the parent plants must have had two identical "factors" (what we would call alleles, varieties of a gene), either two yellow or two green (as in figure 35). When the two were mated, each offspring inherited two different alleles, one from each parent (according to rule 2 above). That is, each offspring seed contained a yellow allele and a green allele. Why then were the peas of this generation all yellow? Because, Mendel explained, yellow was the dominant color and it masked the presence of the green allele in this generation (rule 3 above). However (still according to rule 3), the dominant yellow did not prevent the recessive green from being

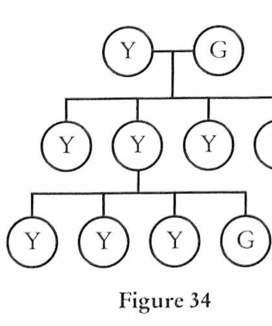


Figure 34

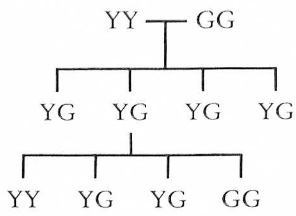


Figure 35

passed on to the next generation. In the next mating round, each plant containing one yellow allele and one green allele was pollinated with another plant containing the same combination of alleles. Since the offspring contain one allele from each parent, the seeds of the next generation may contain one of the following combinations (figure 35): green-green, green-yellow, yellow-green, or yellow-yellow. All the seeds with a yellow allele become yellow peas, because yellow is dominant. Therefore, since all the allele combinations are equally likely, the ratio of yellow to green peas should be 3:1.

You may have noticed that the entire Mendel exercise is essentially identical to the experiment of tossing two coins. Assigning heads to green and tails to yellow and asking what fraction of the peas would be yellow (given that yellow is dominant in determining the color) is precisely the same as asking what is the probability of obtaining at least one tails in tossing two coins. Clearly that is 3/4, since three of the possible outcomes (tails-tails, tails-heads, heads-tails, heads-heads) contain a tails. This means that the ratio of the number of tosses that do contain at least one tails to the number of tosses that do not should be (in the long run) 3:1, just as in Mendel's experiments.

In spite of the fact that Mendel published his paper "Experiments on Plant Hybridization" in 1865 (and he also presented the results at two scientific meetings), his work went largely unnoticed until it was rediscovered at the beginning of the twentieth century. While some questions related to the accuracy of his results have been raised, he is still regarded as the first to have laid the mathematical foundations of modern genetics. Following in the path cleared by Mendel, the influential British statistician Ronald Aylmer Fisher (1890–1962) established the field of population genetics—the mathematical branch that centers on modeling the distribution of genes within a population and on calculating how gene frequencies change over time. Today's geneticists can use statistical samplings in combination with DNA studies to forecast probable characteristics of unborn offspring. But still, how exactly are probability and statistics related?

Facts and Forecasts

Scientists who try to decipher the evolution of the universe usually try to attack the problem from both ends. There are those who start from the tiniest fluctuations in the cosmic fabric in the primordial universe, and there are those who study every detail in the current state of the universe. The former use large computer simulations in an attempt to evolve the universe forward. The latter engage in the detective-style work of trying to deduce the universe's past from a multitude of facts about its present state. Probability theory and statistics are related in a similar fashion. In probability theory the variables and the initial state are known, and the goal is to predict the most likely end result. In statistics the outcome is known, but the past causes are uncertain.

Let's examine a simple example of how the two fields supplement each other and meet, so to speak, in the middle. We can start from the fact that statistical studies show that the measurements of a large variety of physical quantities and even of many human characteristics are distributed according to the *normal frequency curve*. More precisely, the normal curve is not a single curve, but rather a family of curves, all describable by the same general function, and all being fully characterized by just two mathematical quantities. The first of these quantities—the *mean*—is the central value about which the distribution is symmetric. The actual value of the mean depends, of course, on the type of variable being measured (e.g., weight, height, or IQ). Even for the same variable, the mean may be different for different populations. For instance, the mean of the heights of men in Sweden is probably different from the mean of the heights of men in Peru. The second quantity that defines the normal curve is known as the *standard deviation*. This is a measure of how closely the data are clustered around the mean value. In figure 36, the normal curve (a) has the largest standard

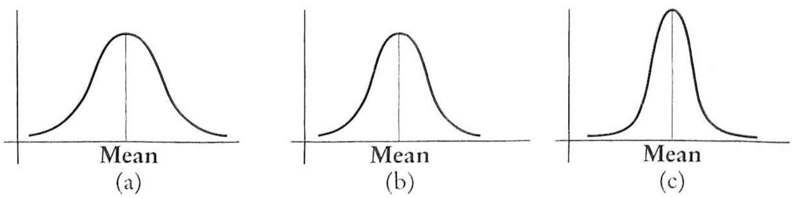


Figure 36

deviation, because the values are more widely dispersed. Here, however, comes an interesting fact. By using integral calculus to calculate areas under the curve, one can prove mathematically that irrespective of the values of the mean or the standard deviation, 68.2 percent of the data lie within the values encompassed by one standard deviation on either side of the mean (as in figure 37). In other words, if the mean IQ of a certain (large) population is 100, and the standard deviation is 15, then 68.2 percent of the people in that population have IQ values between 85 and 115. Furthermore, for all the normal frequency curves, 95.4 percent of all the cases lie within two standard deviations of the mean, and 99.7 percent of the data lie within three standard deviations on either side of the mean (figure 37). This implies that in the above example, 95.4 percent of the population have IQ values between 70 and 130, and 99.7 percent have values between 55 and 145.

Suppose now that we want to predict what the probability would be for a person chosen at random from that population to have an IQ value between 85 and 100. Figure 37 tells us that the probability would be 0.341 (or 34.1 percent), since according to the laws of probability, the probability is simply the number of favorable outcomes divided by the total number of possibilities. Or we could be interested in finding out what the probability is for someone (chosen at random) to have an IQ value higher than 130 in that population. A glance at figure 37 reveals that the probability is only about 0.022, or 2.2 percent. Much in the same way, using the properties of the normal distribution and the tool of integral calculus (to calculate areas), one can calculate the probability of the IQ value being in any given range. In other words, probability theory and its complementary helpmate, statistics, combine to give us the answer.

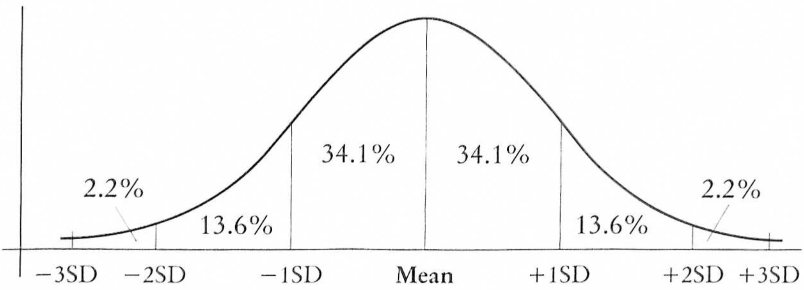
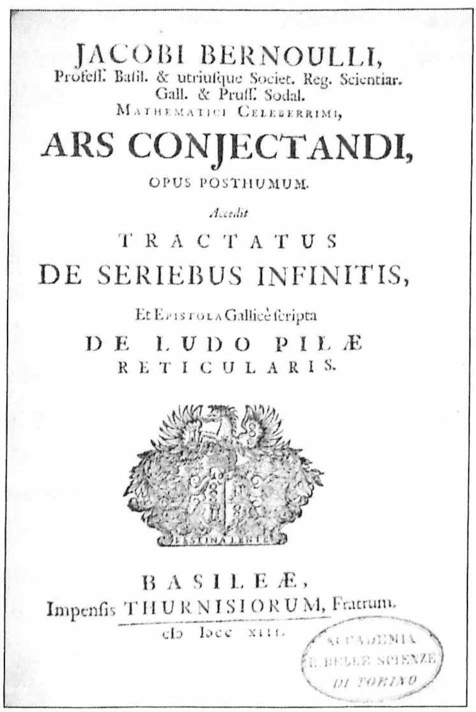


Figure 37

As I have noted several times already, probability and statistics become meaningful when one deals with a large number of events—never individual events. This cardinal realization, known as the *law of large numbers,* is due to Jakob Bernoulli, who formulated it as a theorem in his book *Ars Conjectandi* (*The Art of Conjecturing;* figure 38 shows the frontispiece). In simple terms, the theorem states


JACOBI BERNOULLI,
Profeff. Bafil. & utriufque Societ. Reg. Scientiar. Gall. & Pruff. Sodal.
MATHEMATICI CELEBERRIMI,
ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè fcripta
DE LUDO PILÆ
RETICULARIS.

FESTINA LENTE

BASILEÆ,
Impenfis THURNISIORUM, Fratrum.
clↄ Iↄcc XIII.

ACCADEMIA R. DELLE SCIENZE DI TORINO

Figure 38

that if the probability of an event's occurrence is p, then p is the most probable proportion of the event's occurrences to the total number of trials. In addition, as the number of trials approaches infinity, the proportion of successes becomes p with certainty. Here is how Bernoulli introduced the law of large numbers in *Ars Conjectandi:* "What is still to be investigated is whether by increasing the number of observations we thereby also keep increasing the probability that the recorded proportion of favorable to unfavorable instances will approach the true ratio, so that this probability will finally exceed any desired degree of certainty." He then proceeded to explain the concept with a specific example:

> We have a jar containing 3000 small white pebbles and 2000 black ones, and we wish to determine empirically the ratio of white pebbles to the black—something we do not know—by drawing one pebble after another out of the jar, and recording how often a white pebble is drawn and how often a black. (I remind you that an important requirement of this process is that you put back each pebble, after noting the color, before drawing the next one, so that the number of pebbles in the urn remains constant.) Now we ask, is it possible by indefinitely extending the trials to make it 10, 100, 1000, etc., times more probable (and ultimately "morally certain") that the ratio of the number of drawings of a white pebble to the number of drawings of a black pebble will take on the same value (3:2) as the actual ratio of white to black pebbles in the urn, than that the ratio of the drawings will take on a different value? If the answer is no, then I admit that we are likely to fail in the attempt to ascertain the number of instances of each case (i.e., the number of white and of black pebbles) by observation. But if it is true that we can finally attain moral certainty by this method [and Jakob Bernoulli proves this to be the case in the following chapter of *Ars Conjectandi*] . . . then we can determine the number of instances *a posteriori* with almost as great accuracy as if they were known to us *a priori.*

Bernoulli devoted twenty years to the perfection of this theorem, which has since become one of the central pillars of statistics. He concluded with his belief in the ultimate existence of governing laws, even in those instances that appear to be a matter of chance:

> If all events from now through eternity were continually observed (whereby probability would ultimately become certainty), it would be found that everything in the world occurs for definite reasons and in definite conformity with law, and that hence we are constrained, even for things that may seem quite accidental, to assume a certain necessity and, as it were, fatefulness. For all I know that is what Plato had in mind when, in the doctrine of the universal cycle, he maintained that after the passage of countless centuries everything would return to its original state.

The upshot of this tale of the science of uncertainty is very simple: Mathematics is applicable in some ways even in the less "scientific" areas of our lives—including those that appear to be governed by pure chance. So in attempting to explain the "unreasonable effectiveness" of mathematics we cannot limit our discussion only to the laws of physics. Rather, we will eventually have to somehow figure out what it is that makes mathematics so omnipresent.

The incredible powers of mathematics were not lost on the famous playwright and essayist George Bernard Shaw (1856–1950). Definitely not known for his mathematical talents, Shaw once wrote an insightful article about statistics and probability entitled "The Vice of Gambling and the Virtue of Insurance." In this article, Shaw admits that to him insurance is "founded on facts that are inexplicable and risks that are calculable only by professional mathematicians." Yet he offers the following perceptive observation:

> Imagine then a business talk between a merchant greedy for foreign trade but desperately afraid of being shipwrecked or eaten by savages, and a skipper greedy for cargo and passen-

gers. The captain answers the merchant that his goods will be perfectly safe, and himself equally so if he accompanies them. But the merchant, with his head full of the adventures of Jonah, St. Paul, Odysseus, and Robinson Crusoe, dares not venture. Their conversation will be like this:

Captain: Come! I will bet you umpteen pounds that if you sail with me you will be alive and well this day a year.

Merchant: But if I take the bet I shall be betting you that sum that I shall die within the year.

Captain: Why not if you lose the bet, as you certainly will?

Merchant: But if I am drowned you will be drowned too; and then what becomes of our bet?

Captain: True. But I will find you a landsman who will make the bet with your wife and family.

Merchant: That alters the case of course; but what about my cargo?

Captain: Pooh! The bet can be on the cargo as well. Or two bets: one on your life, the other on the cargo. Both will be safe, I assure you. Nothing will happen; and you will see all the wonders that are to be seen abroad.

Merchant: But if I and my goods get through safely I shall have to pay you the value of my life and of the goods into the bargain. If I am not drowned I shall be ruined.

Captain: That also is very true. But there is not so much for me in it as you think. If you are drowned I shall be drowned first; for I must be the last man to leave the sinking ship. Still, let me persuade you to venture. I will make the bet ten to one. Will that tempt you?

Merchant: Oh, in that case—

The captain has discovered insurance just as the goldsmiths discovered banking.

For someone such as Shaw, who complained that during his education "not a word was said to us about the meaning or utility of mathematics," this humorous account of the "history" of the mathematics of insurance is quite remarkable.

With the exception of Shaw's text, we have so far followed the development of some branches of mathematics more or less through the eyes of practicing mathematicians. To these individuals, and indeed to many rationalist philosophers such as Spinoza, Platonism was obvious. There was no question that mathematical truths existed in their own world and that the human mind could access these verities without any observation, solely through the faculty of reason. The first signs of a potential gap between the perception of Euclidean geometry as a collection of universal truths and other branches of mathematics were uncovered by the Irish philosopher George Berkeley, Bishop of Cloyne (1685–1753). In a pamphlet entitled *The Analyst; Or a Discourse Addressed to An Infidel Mathematician* (the latter presumed to be Edmond Halley), Berkeley criticized the very foundations of the fields of calculus and analysis, as introduced by Newton (in *Principia*) and Leibniz. In particular, Berkeley demonstrated that Newton's concept of "fluxions," or instantaneous rates of change, was far from being rigorously defined, which in Berkeley's mind was sufficient to cast doubt on the entire discipline:

> The method of fluxions is the general key, by help whereof the modern Mathematicians unlock the secrets of Geometry, and consequently of Nature . . . But whether this Method be clear or obscure, consistent or repugnant, demonstrative or precarious, as I shall inquire with the utmost impartiality, so I submit my inquiry to your own Judgement, and that of every candid Reader.

Berkeley certainly had a point, and the fact is that a fully consistent theory of analysis was only formulated in the 1960s. But mathematics was about to experience a more dramatic crisis in the nineteenth century.