Statistics may not sound like the most heroic of pursuits. But if not for statisticians, a Long Island fisherman might have died in the Atlantic Ocean after falling off his boat early one morning last summer.

The man owes his life to a once obscure field known as Bayesian statistics — a set of mathematical rules for using new data to continuously update beliefs or existing knowledge.

The method was invented in the 18th century by an English Presbyterian minister named Thomas Bayes — by some accounts to calculate the probability of God's existence. In this century, Bayesian statistics has grown vastly more useful because of the kind of advanced computing power that did not exist even 20 years ago.

It is proving especially useful in approaching complex problems, including searches like the one the Coast Guard used in 2013 to find the missing fisherman, John Aldridge (though not, so far, in the hunt for Malaysia Airlines Flight 370).

Now Bayesian statistics are rippling through everything from physics to cancer research, ecology to psychology. Enthusiasts say they are allowing scientists to solve problems that would have been considered impossible just 20 years ago. And lately, they have been thrust into an intense debate over the reliability of research results.

When people think of statistics, they may imagine lists of numbers — batting averages or life-insurance tables. But the current debate is about how scientists

turn data into knowledge, evidence and predictions. Concern has been growing in recent years that some fields are not doing a very good job at this sort of inference. In 2012, for example, a team at the biotech company Amgen announced it had analyzed 53 cancer studies and found it could not replicate 47 of them.

Similar follow-up analyses have cast doubt on so many findings in fields such as neuroscience and social science that researchers talk about a "replication crisis"

Some statisticians and scientists are optimistic that Bayesian methods can improve the reliability of research by allowing scientists to crosscheck work done with the more traditional or "classical" approach, known as frequentist statistics. The two methods approach the same problems from different angles.

The essence of the frequentist technique is to apply probability to data. If you suspect your friend has a weighted coin, for example, and you observe that it came up heads nine times out of 10, a frequentist would calculate the probability of getting such a result with an unweighted coin. The answer (about 1 percent) is not a direct measure of the probability that the coin is weighted; it's a measure of how improbable the nine-in-10 result is — a piece of information that can be useful in investigating your suspicion.

By contrast, Bayesian calculations go straight for the probability of the hypothesis, factoring in not just the data from the coin-toss experiment but any other relevant information — including whether you have previously seen your friend use a weighted coin.

Scientists who have learned Bayesian statistics often marvel that it propels them through a different kind of scientific reasoning than they had experienced using classical methods.

"Statistics sounds like this dry, technical subject, but it draws on deep philosophical debates about the nature of reality," said the Princeton University astrophysicist Edwin Turner, who has witnessed a widespread conversion to Bayesian thinking in his field over the last 15 years.

Countering Pure Objectivity

Frequentist statistics became the standard of the 20th century by promising

just-the-facts objectivity, unsullied by beliefs or biases. In the 2003 statistics primer "Dicing With Death," Stephen Senn traces the technique's roots to 18th- century England, when a physician named John Arbuthnot set out to calculate the ratio of male to female births.

Arbuthnot gathered christening records from 1629 to 1710 and found that in London, a few more boys were recorded every year. He then calculated the odds that such an 82-year run could occur simply by chance, and found that it was one in trillions. This frequentist calculation can't tell them what is causing the sex ratio to be skewed. Arbuthnot proposed that God skewed the birthrates to balance the higher mortality that had been observed among boys, but scientists today favor a biological explanation over a theological one.

Later in the 1700s, the mathematician and astronomer Daniel Bernoulli used a similar technique to investigate the curious geometry of the solar system, in which planets orbit the sun in a flat, pancake-shaped plane. If the orbital angles were purely random — with Earth, say, at zero degrees, Venus at 45 and Mars at 90 — the solar system would look more like a sphere than a pancake. But Bernoulli calculated that all the planets known at the time orbited within seven degrees of the plane, known as the ecliptic.

What were the odds of that? Bernoulli's calculations put them at about one in 13 million. Today, this kind of number is called a p-value, the probability that an observed phenomenon or one more extreme could have occurred by chance. Results are usually considered "statistically significant" if the p-value is less than 5 percent.

But there is a danger in this tradition, said Andrew Gelman, a statistics professor at Columbia. Even if scientists always did the calculations correctly — and they don't, he argues — accepting everything with a p-value of 5 percent means that one in 20 "statistically significant" results are nothing but random noise.

The proportion of wrong results published in prominent journals is probably even higher, he said, because such findings are often surprising and appealingly counterintuitive, said Dr. Gelman, an occasional contributor to Science Times.

Looking at Other Factors

Take, for instance, a study concluding that single women who were ovulating were 20 percent more likely to vote for President Obama in 2012 than those who were not. (In married women, the effect was reversed.)

Dr. Gelman re-evaluated the study using Bayesian statistics. That allowed him to look at probability not simply as a matter of results and sample sizes, but in the light of other information that could affect those results.

He factored in data showing that people rarely change their voting preference over an election cycle, let alone a menstrual cycle. When he did, the study's statistical significance evaporated. (The paper's lead author, Kristina M. Durante of the University of Texas, San Antonio, said she stood by the finding.)

Dr. Gelman said the results would not have been considered statistically significant had the researchers used the frequentist method properly. He suggests using Bayesian calculations not necessarily to replace classical statistics but to flag spurious results.

A famously counterintuitive puzzle that lends itself to a Bayesian approach is the Monty Hall problem, in which Mr. Hall, longtime host of the game show "Let's Make a Deal," hides a car behind one of three doors and a goat behind each of the other two. The contestant picks Door No. 1, but before opening it, Mr. Hall opens Door No. 2 to reveal a goat. Should the contestant stick with No. 1 or switch to No. 3, or does it matter?

A Bayesian calculation would start with one-third odds that any given door hides the car, then update that knowledge with the new data: Door No. 2 had a goat. The odds that the contestant guessed right — that the car is behind No. 1 — remain one in three. Thus, the odds that she guessed wrong are two in three. And if she guessed wrong, the car must be behind Door No. 3. So she should indeed switch.

In other fields, researchers are using Bayesian statistics to tackle problems of formidable complexity. The New York University astrophysicist David Hogg credits Bayesian statistics with narrowing down the age of the universe. As recently as the late 1990s, astronomers could say only that it was eight billion to 15 billion years; now, factoring in supernova explosions, the distribution of galaxies and patterns

seen in radiation left over from the Big Bang, they have concluded with some confidence that the number is 13.8 billion years.

Bayesian reasoning combined with advanced computing power has also revolutionized the search for planets orbiting distant stars, said Dr. Turner, the Princeton astrophysicist.

In most cases, astronomers can't see these planets; their light is drowned out by the much brighter stars they orbit. What the scientists can see are slight variations in starlight; from these glimmers, they can judge whether planets are passing in front of a star or causing it to wobble from their gravitational tug.

Making matters more complicated, the size of the apparent wobbles depends on whether astronomers are observing a planet's orbit edge-on or from some other angle. But by factoring in data from a growing list of known planets, the scientists can deduce the most probable properties of new planets.

One downside of Bayesian statistics is that it requires prior information — and often scientists need to start with a guess or estimate. Assigning numbers to subjective judgments is "like fingernails on a chalkboard," said physicist Kyle Cranmer, who helped develop a frequentist technique to identify the latest new subatomic particle — the Higgs boson.

Others say that in confronting the so-called replication crisis, the best cure for misleading findings is not Bayesian statistics, but good frequentist ones. It was frequentist statistics that allowed people to uncover all the problems with irreproducible research in the first place, said Deborah Mayo, a philosopher of science at Virginia Tech. The technique was developed to distinguish real effects from chance, and to prevent scientists from fooling themselves.

Uri Simonsohn, a psychologist at the University of Pennsylvania, agrees. Several years ago, he published a paper that exposed common statistical shenanigans in his field — logical leaps, unjustified conclusions, and various forms of unconscious and conscious cheating.

He said he had looked into Bayesian statistics and concluded that if people misused or misunderstood one system, they would do just as badly with the other.

Bayesian statistics, in short, can't save us from bad science.

At Times a Lifesaver

Despite its 18th-century origins, the technique is only now beginning to reveal its power with the advent of 21st-century computing speed.

Some historians say Bayes developed his technique to counter the philosopher David Hume's contention that most so-called miracles were likely to be fakes or illusions. Bayes didn't make much headway in that debate — at least not directly.

But even Hume might have been impressed last year, when the Coast Guard used Bayesian statistics to search for Mr. Aldridge, its computers continually updating and narrowing down his most probable locations.

The Coast Guard has been using Bayesian analysis since the 1970s. The approach lends itself well to problems like searches, which involve a single incident and many different kinds of relevant data, said Lawrence Stone, a statistician for Metron, a scientific consulting firm in Reston, Va., that works with the Coast Guard.

At first, all the Coast Guard knew about the fisherman was that he fell off his boat sometime from 9 p.m. on July 24 to 6 the next morning. The sparse information went into a program called Sarops, for Search and Rescue Optimal Planning System. Over the next few hours, searchers added new information — on prevailing currents, places the search helicopters had already flown and some additional clues found by the boat's captain.

The system could not deduce exactly where Mr. Aldridge was drifting, but with more information, it continued to narrow down the most promising places to search.

Just before turning back to refuel, a searcher in a helicopter spotted a man clinging to two buoys he had tied together. He had been in the water for 12 hours; he was hypothermic and sunburned but alive.

Even in the jaded 21st century, it was considered something of a miracle.

Citations:
In-text: (Flam 2015)
Bibliography: Flam, F. D. 29 September 2014. "The Odds, Continually Updated." *New York Times*. Accessed 2 September 2015. http://www.nytimes.com/2014/09/30/science/the-odds-continually-updated.html.