"Pinning Down Outliers: 19th Century Stabs at Exact Probabilities for Rare Events"

Byron Wall,
York University, Toronto

Abstract: In the late 19th century, statistics emerged as a discipline from probability theory. Statistics made predictions of future events based upon the past frequency of such events under similar circumstances. When the events were commonplace aspects of human experience, such as average longevity for males in the population, a lot of data supported the predicted likelihood of the unknown, future event. But the less frequent the event in question, the smaller the sample of data on which one could make predictions. In some cases, there were no previous outcomes of the kind contemplated. Yet these too were assigned probabilities. The question is, on what basis was a probability assigned for such events? It would not be surprising to find out that many of the assigned probabilities were not based on data at all, but instead were extrapolations based upon dubious assumptions about the symmetry of vast unknowns in Nature. A more disturbing thought is that such irrational probability assignments may have become the norm and have entered standard statistical practices and are still with us today. This paper explores some of the relevant cases from that period.

………

I am a native of New Orleans, a city now famous for having been devastated by the floodwaters from Hurricane Katrina in 2005. It was an unprecedented natural disaster for the United States. The city has always been in danger of flooding, and on that account has a very elaborate and powerful pumping system that rapidly and efficiently drains the streets of standing water and pumps it away.

The chief danger that had been prepared for was flooding either from torrential rain, which the city gets plenty of, or from an overflow of the

Mississippi River which snakes its way around the city of New Orleans to the south of the main city and indeed carries dangerous amounts of water down from northern and central states. By the time the Mississippi reaches New Orleans it is over 2000 miles long from its origin in Minnesota, and has been emptied into by the Minnesota River, the Illinois River, the Chippewa, Black, Wisconsin, Saint Croix, Iowa, Des Moines, and Rock Rivers to the north, and at St. Louis, it is joined by the Missouri River, which drains the Great Plains to the west. And at Cairo, Illinois, it is joined from the east by the Ohio River. When it hits New Orleans, it is about a mile wide, and moving rapidly. It does not take much extra rain or extra snow from the north melting in the spring to overwhelm the river and become a significant peril for people living downstream. This much has been known from the time of the earliest settlement of New Orleans. In fact, the French engineers sent by King Louis XIV to lay out the plan of the city in the early 18th century advised against founding a city in a place so prone to flooding, and trapped between the mighty Mississippi River and the vast lake just to the north. But there were other considerations that made the site of *Nouvelle Orleans* ideal, so it was built where it is now.

The site chosen was already in use in 1718 by French merchants and trappers as a meeting place to do business with the native population and from there to transport furs out to ocean-going ships that had travelled up the river from the Gulf of Mexico. It had easy access by water from several directions since it was essentially swampland. The original part of the city, now known as the French Quarter, was built on the highest land around, at a point where the Mississippi meanders around in a crescent shape, and through centuries of flooding its banks had laid down more silt at this turn than at other nearby places on its shores, leading to the somewhat higher elevation. Even so, the French Quarter is about a foot below sea level. The rest of the city, which was developed over the next 300 years, was on land that was not even up to that level. Newer parts of the city are up to 20 feet below the level of the Mississippi River and Lake Pontchartrain to the north.

Growing up in New Orleans, as I did, we all knew that we lived below sea level, surrounded by a river and a lake that could drown us anytime its levees were breached. It was a matter of considerable pride, or, I should say, hubris, to be nonchalant about the palpable dangers faced by the residents. But of course we were protected by the powerful pumps that drained our streets so efficiently during any rainstorm. As a kid, I can remember sort of bragging to visiting relatives from out of town, or enlightening wide-eyed younger children, on the vagaries of living below sea level. I would gesture toward the river, that was perhaps a mile from my home, and describe the rolling hill of mud that was the levee that kept the mighty Mississippi from flooding the street where I lived. And, as a precursor of my quantitative interests that followed in later years, I would put a figure on the odds of surviving to manhood while living under this imminent threat. If I remember correctly, I suggested to my gullible audience that "9 times out of 10" the levees will do their job as intended, but there is always that time that the system will break down. This would invariably produce the wide-eyed look of panic among the younger members of my audience, which was the desired effect. Now that I am older and a little more numerically literate, I have come to wonder just what those chances really were, and how anyone would come up with a figure to begin with.

In the actual case of New Orleans and its levee and pumping system, the place the system failed was not where it was expected to. The pumps were designed to evacuate flood waters to the large lake that lies to the north of the city, Lake Pontchartrain, which connects to the Gulf of Mexico. It is very effective. The pumps were doubtless working very well in 2005 when Hurricane Katrina hit the city with torrential rain. And had it not been for something quite unexpected, Katrina would have remained just another hurricane of many hurricanes that passed over New Orleans and did what hurricanes normally do in the way of wind and water damage. But alas, something unexpected did happen: two levees that protect the city from rising water levels not on the Mississippi but on Lake Pontchartrain were the ones that broke—under the pressure of the storm surge that came not from the river, but via the lake, where it had not been expected.

Because the levees that broke were along the lake, the pumps that dutifully sent flood waters out to the lake were totally ineffective. Whatever was pumped out returned immediately, not to mention that once the flooding really got underway, electricity in the city was wiped out, which stopped the pumps from working. In a matter of hours, 80% of the city was underwater.

Thus the ultimate cause of the disaster falls into the category of what Nassim Nicholas Taleb has so famously called a "Black Swan." This, in the book of that title, he defines as an event having three attributes:

> First, it is an *outlier*, as it lies outside the realm of regular expectations, because nothing in the past can convincingly point to its possibility. Second, it carries an extreme impact. Third, in spite of its outlier status, human nature makes us concoct explanations for its occurrence *after* the fact, making it explainable and predictable. (Taleb, pp. xvii-xviii.)

Now, in fact the system that was in place in New Orleans to deal with possible flooding was very complex, very expensive, and for the most part, very effective. The city had been developing and refining its protection system throughout its 300 year history. All along the way, estimates had been made of the reliability of the precautions taken to achieve the desired effect, and, in effect, that meant that assessments had been made of the probability of failure.

Though I have not (yet) uncovered the precise estimate of the probability of failure of the lakeside levees, it seems highly likely to me that in this day of the preemptive status of quantitative over other non-quantitative arguments, that at some critical point, a decisive case was made for how important the strength of lakeside levees were in New Orleans to stave off disaster. My guess is that a calculation was produced that argued that the chance of failure of those levees on Lake Pontchartrain was negligible.

What interests me is the reasoning that went into and supports a calculation of the chances of such an event. My very strong suspicion is that the numbers are based upon an extrapolation from the frequency of other outliers that are very dissimilar in character to events such as levee failure, or, more likely, are extrapolations based upon notions of the symmetry of probability distributions that were derived from idealized games of chance rather than from actual frequency distributions of relevant statistical data.

This takes me out of the recent past and into an investigation of the foundations of statistical theory that was being developed in the late nineteenth century, for it was then that general notions were developed on how to apply probability theory to the interpretation and prediction of events in the world outside of the casino.

One of the raging debates in the foundations of probability theory at that time was over the meaning of the probability of an event. The dominant school of thought, especially on the European continent, was that probability referred to the rational assessment of the degree of belief that the next instance of a certain class of events would be of the specified type, e.g., "heads" for a coin toss, or "rain" for a weather prediction. The opposing point of view, most popular among British empiricist philosophers, was that the probability was simply the ratio of successful outcomes to all outcomes in potentially infinitely repeated iterations of the event, or events, that were for all practical purposes nearly identical. Thus, to assert that the probability of heads is ½ in a coin toss entails that there had been, somewhere, somehow, a sufficiently long number of repetitions of actual coin tosses in which the coin came up heads nearly half the time. Or, for the weather prediction, a sufficient number of days with nearly identical conditions in which it rained the percentage of times that is asserted to be the probability of rain in the prediction.

For many instances of ordinary, everyday events, it didn't really matter which of these viewpoints one subscribed to, the asserted probability would come out with the same value. Hence, except for the purists who wanted to get their assertions clarified—typically, these would be philosophers—, the degree of belief school and the frequentist school were saying the same thing, though in different ways. Out of this cacophony of different formulations, statistics emerged as a distinct discipline from probability.

My contention is that regardless of how pervasive the degree of belief conception had been among philosophers and certain groups of mathematicians, it was the frequency viewpoint that provided the framework for statistical analysis. The failing of the degree-of-belief formulation was that rational assessment of degree of belief demanded knowledge of causes, and, where that failed, it required rational assumptions of expectations in the face of ignorance of causes.

On the other hand, a frequency argument came down to faith in the continuation of correlations among events that could be judged similar, without the necessity to know the details.

I have spoken about this at greater length on other occasions, but let me recap a bit of that here to make my point. The mathematical theory of probability had its origin in analyses by mathematicians of the expected outcomes of various games of chance. The overriding mathematical calculation used in probability theory is one of permutations and combinations. What is the chance of drawing a Full House in straight poker? The answer is calculated by counting up all possible hands that contain three cards of one denomination and two cards of another, divided by the total number of all possible hands of five cards from a standard deck. The trick that makes this calculation straightforward and manageable is that for any deal of the cards, the likelihood of any one card appearing is considered identical to the likelihood of any other. And the same principle applies to most games of chance. The likelihood of any one face of a die coming up is deemed to be identical to any other; the same applied to the roulette wheel, the coin toss, a lottery, etc. And, barring dishonest manipulation of the equipment, the physical symmetry of construction of the gaming devices makes such an assumption seem reasonable.

Of course, if one takes a determinist view of the laws of Nature, as would have been almost universal in the 19th century, there really is only one outcome that is possible, and that is the one that occurs. The question is not then one of calculating the chance of an event occurring, but rather, in the face of our ignorance of all the manifold minute factors that go into determining what cards appear in what order in a shuffled deck, or which face will land uppermost on a thrown die, etc., what degree of expectation of a given result is it reasonable to have, taking into account all of our ignorance? That basically is the position of the "degree of belief" interpretation of probability. And, again, note that this viewpoint leads to useful results only if each fundamental event, such as the position of one cast die, or one dealt card, is exactly as likely as any other outcome. A good review of many of the facets of this sort of reasoning can be found in the collection by Sandy Zabell titled *Symmetry and Its Discontents:*

Essays on the History of Inductive Probability, which consists of various papers by Zabell published over the years in a variety of journals.

What I would like to emphasize here is a particular application of probability theory that dovetails perfectly with the determinist view and with the interpretation of probability as representing a degree of belief. That is the interpretation of the normal probability distribution as errors from an ideal or from a true value. In particular, I call attention to the theoretical treatment given by such mathematicians as Pierre Simon de Laplace of astronomical observations. In Laplace's time, much attention was paid to the recorded observations of star transits by astronomers. The first important realization was that they did not entirely agree with each other. Regardless of what measures might be taken to resolve systematic differences between the observations of one astronomer and another, the simple fact is that even the best of astronomers did not report consistent data of the exact time at which a given star passed specified crosshairs in certain telescopes. Laplace and others took the view that there indeed was a "true" value, and the observations, made by fallible human astronomers, represented "errors" from this true value. Much of the analysis of probability distributions was aimed at resolving these errors in a systematic way in order to give the best possible estimate of the true position. Thus the theory that supported the familiar bell-shaped normal probability distribution was called "error theory."

For Laplace, probability theory allowed one to come up with rational measures of human ignorance of the true state of affairs, either of events already having occurred, such as a star position, or of future events, that were, in his view, fully determined, though we could not calculate them with certainty. Note that in the example I have given, that of star transits, the best estimate, the closest approximation to the true value, would be represented by the apex of the probability distribution. It would be the value that had the least deviation from all of the observations taken collectively.

Laplace was among those who eagerly ventured to apply "error theory" to all manner of questions of human judgments, going way beyond matters of simple measurement. Among these, as I have discussed on other occasions, was the matter of the optimal size of juries in criminal cases, and the optimal level of agreement within a jury panel that should be required for a conviction that would minimize the likelihood of a miscarriage of justice. I mention this again here, because the reasoning used made manifest the usage of the principle that in the absence of evidence to the contrary, all possible outcomes are to be viewed as equally probable. This is called, by Zabell and others, the *principle of indifference.* With a confidence in his methodology characteristic of the Enlightenment, Laplace calculated, for example, that a unanimous jury panel of n members has a chance of being wrong equal to $(\frac{1}{2})^{n+1}$. Ian Hacking has commented that "no tidier example of an *a priori* rabbit out of a hat can be imagined." (Hacking, p. 92)

Another application of this astronomical notion of "error theory" to terrestrial matters, one that led much more directly to the conversion of probability theory into a tool of statistical inference, was the groundbreaking study by Adolphe Quetelet translated into English in 1842 as *A Treatise on Man and the Development of his Faculties.* The assumptions about Nature made by Quetelet that justified his application of error theory to the study of human faculties are pertinent. Just as the astronomer was primarily concerned with getting the best estimate of the true value of a star position that can be wrested from a series of differing observations, Quetelet had the idea that the statistics that he collected on human characteristics would uncover God's model for humanity. There was, in Quetelet's mind, an ideal form of a human being that was Nature's template, as it were. Actual human individuals were variants from that ideal that arose as sort of copying errors. Nature was aiming for a certain model, but missed the mark in a random, and I might add, evenly distributed, way that clustered around that ideal model. To Quetelet, the *average* value for any human statistic, from chest girth to age at death to tendency to engage in criminal activity, was the best estimate of Nature's ideal for humanity. All of Quetelet's concern was in establishing what that best estimate was. It was the center of the distribution that interested Quetelet. He had no interest in the outliers, and in fact regarded them as anomalies to be discarded, since they were Nature's mistakes. (Quetelet, *Treatise*, p. 8.)

To summarize what I have been sketching here, the "degree of belief" school of probability interpretation, which was dominant on the European continent in the 18th and 19th centuries, was allied with a thoroughly deterministic view of Nature, where there really was no such thing as chance. Instead, the view was that the world was completely determined down to the last and most inconsequential event, and what probability represented was a measure of our incomplete and imperfect ability as human beings to figure it out. But central to this thinking was the notion that there was a rational plan, an order to the universe, and our best guide to that order was what happened most often. The least important events were those which happened rarely and unexpectedly, which we call outliers.

Clearly the degree-of-belief school carries a lot of philosophical baggage about cause and effect and implied order in the universe. It fit neatly with the ideological philosophical viewpoint characteristic of the Continent.

This interpretation did not go down well in Britain, where empiricism was the favored viewpoint. Thinkers such as Robert Leslie Ellis, John Venn, and John Stuart Mill attacked the foundation of these interpretations, pointing to the inherent circular reasoning that went into such simplifying devices as the principle of insufficient reason that assigns a probability precisely at the point where no information is available. Ellis and Venn in the middle of the 19th century argued that the only justification we have for assigning a probability for the occurrence of a future event is that we have considerable data from previous trials that are similar in all important ways. To take the coin toss, for example, they argue that the only justification we have for saying that the odds of a coin landing either heads or tails is 50:50 is because we have lots of experience with tossing coins and the actual data collected has turned out to support the assertion that in the long run the number of heads will closely approximate the number of tails. (E.g., Ellis, p. 4)

The *frequency interpretation*, as this view of probability is called, had the advantage that it sidestepped the morass of circular reasoning and commitment to a particular view of how the world was organized and argued that inferences about matters on which we have incomplete knowledge should not get too far removed from the data on which they are based. As a result, it is the frequency school of thinking that paved the way for the broadening of statistics to a discipline that makes inferences based upon pure correlation without having to commit to a certain view of cause and effect, or of a grand plan in Nature. But it should be remembered that the frequency interpretation is most convincing when a significant body of data supports the assertion that a future event is likely to happen with a specified probability.

What if the event in question has hardly ever happened at all? What is its probability then? What would an assertion of the probability of the likelihood of failure of a lakeside levee in New Orleans be based upon? Or the likelihood of the Earth being struck by a stray asteroid that will provoke a nuclear winter that will make life intolerable. Or the likelihood that collapses of ill-secured mortgages will cascade and bring down the world's financial system. All of these possibilities are deemed to be highly unlikely. Nevertheless, probabilities have been assigned to them and important decisions about how we organize society have been based upon those assigned numerical probabilities.

What are these numbers based upon? When was it decided that we could come up with measures of the likelihood of such "improbable" events, and how could that have been justified in the face of the opposition of the frequentist interpreters? I suggest that this most likely happened during the period when the frequency interpretation had its fewest watchguards to keep the hocus-pocus out of probability inferences and at the time when statistical inference was just beginning to be extended beyond the most obvious uses where there would have been a fair bit of data, such as setting life insurance premiums. That would place it after the period when Robert Leslie Ellis, George Boole, and John Stuart Mill were writing, and also after when John Venn was actively looking at questions of probability. That would put it after 1897, when Venn turned away from logic and probability. It would also likely be before mathematical statistics became established as a distinct academic discipline with its own paradigm of normal science in which, I contend, we will find accepted procedures to calculate the likelihood of highly improbable events. Stephen Stigler has ventured the view that mathematical statistics reached that position in 1933. (Stigler, ch. 8, pp. 157-172).

So if I am right, the mathematical formulation for the calculation of the likelihood of highly improbable events slipped into statistics without too much objection sometime in that 35-40 year interval and has led us to a certain unwarranted overconfidence that we have taken appropriate precautions to protect ourselves from unforeseen futures. Over the next six months, I shall be searching the literature trying to find evidence of such formulations.

List of Works Cited

Ellis, R. L. "On the Foundations of the Theory of Probabilities," *Transactions of the Cambridge Philosophical Society*, 8, pt. 1:(1844) 1-6. Read 14 February 1842.

Hacking, Ian. *The Taming of Chance.* Cambridge: Cambridge University Press, 1990.

Mill, John Stuart. *A System of Logic, Rratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, ed. J.M. Robson. *Collected Works of John Stuart Mill*, vols 7-8. Toronto: University of Toronto Press, 1974 [1843].

Quetelet, Adolphe. *Sur L'Homme et le Dévelopment de ses Facultés.* 1835. Translated as *A Treatise on Man and the Development of his Faculties.* Edinburgh: William and Robert Chambers, 1842.

Stigler, Stephen M. *Statistics on the Table: The History of Statistical Concepts and Methods.* Cambridge, MA: Harvard University Press, 1999.

Taleb, Nassim Nicholas. *The Black Swan: The Impact of the Highly Improbable.* New York: Random House, 2007. Paperback edition, Penguin, 2008. Citations above are from the paperback edition.

Venn, John. The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to its Application to Moral and Social Science. London and Cambridge: Macmillan and Co, 1866.

Zabell, S. L. *Symmetry and Its Discontents: Essays on the History of Inductive Probability*.New York: Cambridge University Press, 2005.