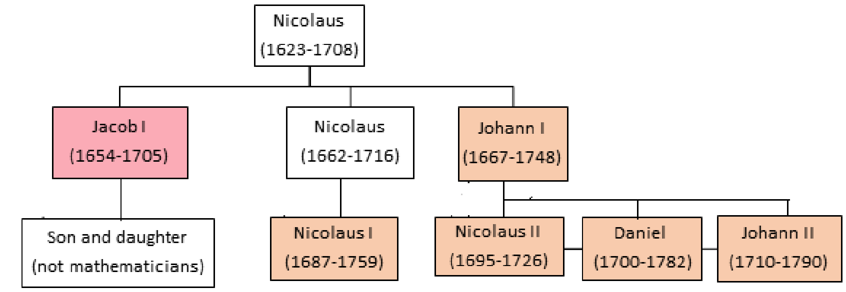
History of Probability (Part 3) - Jacob Bernoulli (1654-1705) – Law of Large Numbers



The Bernoulli family is the most prolific family of western mathematics. In a few generations in the 17th and 18th centuries more than ten members produced significant work in mathematics and physics. They are all descended from a Swiss family of successful spice traders. It is easy to get them mixed up (many have the same first name.) Here are three rows from the family tree. The six men whose boxes are shaded were mathematicians.

For contributions to probability and statistics, **Jacob Bernoulli** (1654-1705) deserves exceptional attention. Jacob gave the first proof of what is now called the (weak) **Law of Large Numbers.** He was trying to broaden the theory and application of probability from dealing solely with games of chance to "civil, moral, and economic" problems.

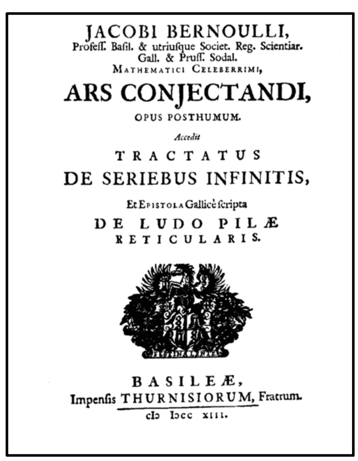
Jacob Bernoulli

Jacob and his younger brother, Johann, studied mathematics against the wishes of their father. It was a time of exciting new mathematics that encouraged both cooperation and competition. Many scientists (including Isaac Newton and Gottfried Leibniz) were publishing papers and books and exchanging letters. In a classic case of sibling rivalry, Jacob and Johann, themselves, were rivals in mathematics their whole lives. Jacob, the older one, seems to have been threatened by his brilliant little brother, and was fairly nasty to him in their later years.

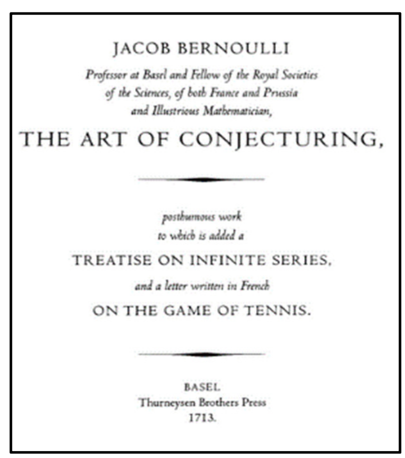
Here is a piece of a letter (1703) from Jacob Bernoulli to Leibniz that reveals the tension between the brothers. Jacob seems worried that Johann, who had told Leibniz about Jacob's law of large numbers, may not have given him full credit.

> I would very much like to know, dear sir, from whom you have it that a theory of estimating probabilities has been cultivated by me. It is true that several years ago I took great pleasure in this sort of speculations, so that I could hardly have thought any more about them. I had the desire to write a treatise on this matter. But I often put it aside for whole years because my natural laziness compounded by my illnesses made me most reluctant to get to writing. I often wished for a secretary who could easily understand my ideas and put them down on paper. Nevertheless, I have completed most of the book, but there is lacking the most important part, in which I teach how the principles of the art of conjecturing are applied to civil. moral. and economic matters. [This I would do after] having solved finally a very singular problem, of very great difficulty and utility. I sent this solution already twelve years ago to my brother, even if he, having been asked about the same subject by the Marquis de l'Hopital, may have hid the truth, playing down my work in his own interest."
> (April, 1703)

Jacob died in 1705, leaving many unpublished papers. Eight years later, in 1713, his family published Jacob's notes for the book he mentioned in the letter to Leibniz, ***Ars Conjectandi*** (*The art of conjecturing*). The book fundamentally changed the way mathematicians approached probability. Its crowning achievement was the detailed *proof* of Jacob's law of large numbers.



JACOBI BERNOULLI,
Profess. Basil. & utriusque Societ. Reg. Scientiar. Gall. & Pruss. Sodal.
MATHEMATICI CELEBERRIMI,
ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè scripta
DE LUDO PILÆ
RETICULARIS.
BASILEÆ,
Impensis THURNISIORUM, Fratrum.
clɔ Iɔcc xiii.

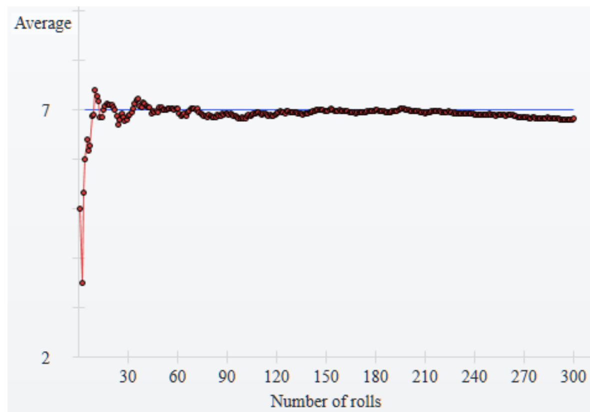


JACOB BERNOULLI
Professor at Basel and Fellow of the Royal Societies of the Sciences, of both France and Prussia and Illustrious Mathematician,
THE ART OF CONJECTURING,
posthumous work
to which is added a
TREATISE ON INFINITE SERIES,
and a letter written in French
ON THE GAME OF TENNIS.
BASEL
Thurneysen Brothers Press
1713.

What is Bernoulli's Law of Large Numbers?

Jacob's theorem is what we informally call the *Law of Averages*. It says that if you take more and more observations of some random outcome (such as the sum of two dice), then the *mean* of all your observations will *eventually* approach the theoretical Expected Value and from that point stay very close to it. The label, "Law of Large Numbers," first appeared in 1837, more than 100 years after Bernoulli proved it, in a book by the French mathematician, Simeon Denis Poisson, who proved a more general version of the theorem.

> Things of every kind of nature are subject to a universal law which one may well call the **Law of Large Numbers**. It consists in that if one observes large numbers of events of the same nature depending on causes which are constant and causes which vary irregularly, . . . , one finds that the proportions of occurrence are almost constant . . .[[In the original French, Poisson wrote *Loi des grands nombres*.]



Here's typical graph, based on a computer simulation of 300 rolls of two dice. Recall, the expected value for the sum of two dice is 7. The graph should be approaching height 7. This image makes it clear that "eventually" may mean a really long time.

In this simulation the mean was 6.82 after 300 rolls. It was off from the expected value by 0.18. True, it was no longer wildly away from 7 like it was at the beginning, but suppose you need to be "sure" (Bernoulli said "morally certain") that your experimental mean was off from the true expected value by no more than 0.01. *The law of large numbers gives you a minimum value for the number of rolls, n, to accomplish this.*

Bernoulli's version of the law focused on *proportions of success and failure*, especially on the ratio of success to failures. For example, the theoretical probability of a die coming up 1 is 1 out of 6. His law says that if you throw a die *enough times* the ratio of times it comes up 1 to the times it doesn't, will, in fact, approach 1 to 5.

On the face of it, the theorem just says that more and more evidence will continue to get you closer to the truth. That's obvious. Bernoulli says as much right up front. The new contribution was his procedure for determining the number of observations needed to get to a prescribed degree of closeness.

Here are two translations of Bernoulli's Latin words on the "obviousness" of the basic claim. Note how translations can affect your sense of the emotion in an author's work.

"For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one's goal" Quoted in Stigler, *The History of Statistics*	"…even the most foolish person, by some instinct of nature, alone and with no previous instruction (which is truly astonishing), has discovered that the more observations of this sort are made, the less danger there will be of error." Sylla translation, *The Art of Conjecturing*

Using the law to investigate potential causes

Jacob's book made an important distinction between "*a priori*" and "*a posteriori*" probabilities. These Latin labels mean "from theory" and "after observation." For example, if you know that an urn contains 100 identical white balls and 300 black, then the *a priori* probability of pulling white at random from the urn is 1/4. In contrast, if you don't know the fraction of white in the urn, then you can only estimate it by looking at *a posteriori* evidence you collect after doing an experiment.

Jacob wanted to show how to use *a posteriori* ratios of success to failure to estimate the unknown *a priori* ratio. This would let him broaden the scope of application to issues of importance in society, to situations in which it was impossible to know *a priori* the relevant probabilities.

The following quotation is from *Ars Conjectandi* just before the proof. It's one thing, Jacob says, to figure out questions about dice and balls in urns ---

> But what mortal, I ask, may determine, for example, the number of diseases, as if they were just as many cases, which may invade at any age the innumerable parts of the human body and which imply our death? And who can determine how much more easily one disease may kill than another - the plague compared to dropsy, dropsy compared to fever? Who, then, can form conjectures on the future state of life and death on this basis? Likewise, who will count the innumerable cases of the changes to which the air is subject every day and on this basis conjecture its future constitution after a month, not to say after a year? Again, who has a sufficient perspective on the nature of the human mind or on the wonderful structure of the body so that they would dare to determine the cases in which this or that player may win or lose in games that depend in whole or in part on the shrewdness or the agility of the players? In these and similar situations, since they may depend on causes that are entirely hidden and that would forever mock our diligence by an innumerable variety of combinations, it would clearly be mad to want to learn anything in this way.

But, don't give up!

> Nevertheless, another way is open to us by which we may obtain what is sought. **What cannot be ascertained *a priori*, may at least be found out *a posteriori* from the results many times observed in similar situations, since it should be presumed that something can happen or not happen in the future in as many uses as it was observed to happen or not to happen in similar circumstances in the past.**

In Jacob's correspondence with Leibniz in 1703 and 1704 he did lay out, without proof, what his theorem was supposed to accomplish. [Note: "double ratio" means 2 to 1.]

> I will give you an example. I posit that some number of black and white tokens is hidden in an urn and that the number of white tokens is double the number of black ones. But that you do not know the ratio of the two types but want to determine it by experiments. You take one token after another out of the urn (replacing each one after you take it so that the number of tokens in the urn is net reduced), and you observe whether each one you select is light or dark.
>
> Now I say that, taking two ratios as near as you like to a double ratio, one larger and one smaller. say 201:100 and 199:100, I will scientifically determine the number of observations necessary for it to be ten or a hundred or a thousand times more probable to you that the ratio of the number of times you select a white token to the number of times you select a black token will fall within rather than outside the limits 201:100 and 199:100 of a double ratio-until finally you can be morally certain that the ratio to be found by experiment will approach the true double ratio as closely as you like.

To sum up, his procedure allows you to set two criteria for an estimate.

1. State *how close* you want your observed ratio be to the true ratio. In the example he gave Leibniz, he wants to be sure that the observed ratio is between 201 to 100 and 199 to 100, that is, between 1.99 and 2.01. In other words, set the maximum error of your observation-based estimate of the true ratio to be less than .01.

2. Chose an odds ratio that will satisfy you. Do you want the odds to be ten to one, say, or a thousand to one that the observed ratio is inside the band from 1.99 to 2.01 rather than outside?

After you specify these criteria his theorem will tell you the required number of trials. You can see in his reasoning the origins of our ideas about confidence intervals.

Ars Conjectandi was published about ten years after this exchange of letters. The whole book is a textbook on probability, and the last part contains this theorem. After an opening section where Bernoulli clarifies terms and discusses what he wants to do, he points out how hard and how important this is.

> This, therefore, is the problem that I have proposed to publish in this place, after I have already concealed it for twenty years. Both its novelty and its great utility combined with its equally great difficulty can add to the weight and value of all the other chapters of this theory.

Bernoulli's proof is based on a model of r white and s black "tokens" in an urn. Today we might call the two possibilities represented by r and s "success" and "failure." He called them *fertile* and *sterile*. He uses t to represent the sum of r and s.

Here is what he wants to establish:

> Let the number of fertile cases and the number of sterile cases have exactly or approximately the ratio r/s, and let the number of fertile cases to all the cases be in the ratio r/ (r + s) or r/t, which ratio is bounded by the limits (r + 1)/t and (r- 1)/t. **It is to be shown that so many experiments can be taken that it becomes any given number of times (say c times) more likely that the number of fertile observations will fall between these bounds than outside them**, that is, that the ratio of the number of fertile to the number of all the observations will have a ratio that is neither more than (r + 1)/t nor less than (r - 1)/t.

His approach involves many statements describing inequalities and computations based on the logarithms of ratios of terms in the expansion of $(r+s)^{nt}$. [This approach is very similar to the way Abraham DeMoive derived the formula for the normal curve.] He uses c to represent the degree of certainty, essentially the odds that his estimate is correct. His solution is,

$$\textbf{number of experiments} = mt + \frac{mrt - rt}{s+1} \quad \textbf{, where } m \geq \frac{\log c(r-1)}{\log(s+1) - \log s}$$

After he completes the proof he demonstrates it with a specific numerical example where the unknown ratio of r to s is 3 to 2. He wants to know how many trials it will take so that he can be "certain" that his observed ratio is no more than .02 off from the true ratio. He notes he will get better precision if he thinks of this ratio as $r = 30$ to $s = 20$. (That way his desired error will then be $\frac{1}{30+20} = \frac{1}{50}$). He picks $c = 1000$ as his index for moral certainty. Using these values for r, s, and c, his formula reveals that he will need *25,550 trials!* This was so huge as to be impractical in real applications.

After showing the 25,550 result only one paragraph remains in the book, and here Bernoulli notes that he can keep getting more and more certain by increasing c, closing in on a perfect estimate of the unknown *a priori* ratio. Then he gets philosophical, wondering what it might mean if you *could* have an infinite number of observations.

> Whence at last this remarkable result is seen to follow, that if the observations of all events were continued for the whole of eternity (with the probability finally transformed into perfect certainty) then everything in the world would be observed to happen in fixed ratios and with a constant law of alternation. Thus in even the most accidental and fortuitous we would be bound to acknowledge a certain quasi-necessity and, so to speak, fatality. I do not know whether or not Plato already wished to assert this result in his dogma of the universal return of things to their former positions, in which he predicted that after the unrolling of innumerable centuries everything would return to its original state.

Exercises

1. Confirm that Bernoulli's calculation of 25,550 is correct.

2. See this applet that demonstrates the law of large numbers for the proportion of successes. http://digitalfirst.bfwpub.com/stats_applet/stats_applet_10_prob.html.

   Use it to fill in this table, just to get a rough sense of pattern.

Trial	1	2	3	4	5		50		100
Successes (Yes or No)									
Proportion of successes so far									

Sources

Bernoulli, Jacob, *The Art of conjecturing*, translation by Edith Dudley Sylla, Johns Hopkins University Press, 2006

Letter from Jacob Bernoulli to Leibniz, translation by Edith Dudley Sylla, *Perspectives on Science*, (1998), https://muse.jhu.edu/article/28021#b5

Katz, Victor J. *A History of Mathematics*, Addison Wesley, 1998 (2nd edition)

Seneta, Eugene. A Tricentenary history of the Law of Large Numbers, *Bernoulli Journal*, 2013, 19(4)

Stigler, Stephan, *The History of Statistics*, Harvard University Press, 1986