

# Bernoulli's Law of Large Numbers

Erwin Bolthausen\* and Mario V. Wüthrich†

## Abstract

This year we celebrate the 300 years anniversary of Jakob Bernoulli's path breaking work *Ars conjectandi* which has appeared in 1713, eight years after his death. In Part IV of his masterpiece Bernoulli proves the law of large numbers which is one of the fundamental theorems in probability theory, statistics and actuarial science. We review and comment his original proof.

## 1 Introduction

In a correspondence Jakob Bernoulli writes to Gottfried Wilhelm Leibniz in October 1703 [5]: “*Obwohl aber seltsamerweise durch einen sonderbaren Naturinstinkt auch jeder Dümme ohne irgend eine vorherige Unterweisung weiss, dass je mehr Beobachtungen gemacht werden, umso weniger die Gefahr besteht, dass man das Ziel verfehlt, ist es doch ganz und gar nicht Sache einer Laienuntersuchung, dieses genau und geometrisch zu beweisen*”, saying that anyone would guess that the



Jakob Bernoulli

more observations we have the less we can miss the target, however, a rigorous analysis and proof of this conjecture is not trivial at all. This extract refers to the law of large numbers. Furthermore, Bernoulli expresses that such thoughts are not new, but he is proud of being the first one who has given a rigorous mathematical proof to the statement of the law of large numbers. Bernoulli's results are the foundations of the estimation and prediction theory that allows to apply probability theory well beyond combinatorics. The law of large numbers is derived in Part IV of his centennial work *Ars conjectandi* which has appeared in 1713, eight years after his death (published by his nephew Nicolaus Bernoulli), see [1].

\*University of Zurich, Institut für Mathematik, 8057 Zurich

†ETH Zurich, RiskLab, Department of Mathematics, 8092 Zurich

Jakob Bernoulli (1655-1705) was one of the many prominent Swiss mathematicians of the Bernoulli family. Following his father's wish, he studied philosophy and theology and he received a *lic. theol.* degree in 1676 from the University of Basel. He discovered mathematics almost auto-didactically and between 1676 and 1682 he served as a private teacher. During this period he also traveled through France, Germany, The Netherlands and the UK where he got into contact with many mathematicians and their work. Back in Basel in 1682 he started to teach private lectures in physics at university, and in 1687 he was appointed as professor in mathematics at the University of Basel. During this time he started to apply Leibniz' infinitesimal calculus and he started to examine the law of large numbers.

The law of large numbers states that (independent) repetitions of an experiment average over long time horizons in an arithmetic mean which is obviously not generated randomly but is a well-specified deterministic value. This exactly reflects the intuition that a random experiment averages if it is repeated sufficiently often. For instance, if we toss a coin very often we expect about as many heads as tails, which means that we expect about 50% (deterministic value) of each possible outcomes.

The law of large numbers formulated in modern mathematical language reads as follows: assume that  $X_1, X_2, \dots$  is a sequence of uncorrelated and identically distributed random variables having finite mean  $\mu = \mathbb{E}[X_1]$ . Define  $S_N = \sum_{i=1}^N X_i$ . For every  $\varepsilon > 0$  we have

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \left| \frac{S_N}{N} - \mu \right| \geq \varepsilon \right) = 0. \quad (1)$$

Bernoulli has proved (1) for i.i.d. Bernoulli random variables  $X_1, X_2, \dots$  taking only values in  $\{0, 1\}$  with probability  $p = \mathbb{P}(X_1 = 1)$ . We call the latter a *Bernoulli experiment* and in this case we have  $\mu = p$ . We will discuss Bernoulli's proof below, the general formulation (1) is taken from Khinchin [4]. In introductory courses on probability theory and statistics one often proves (1) under the more restrictive assumption of  $X_1$  having finite variance. In this latter case the proof easily follows from Chebychev's inequality.

Today, Bernoulli's law of large numbers (1) is also known as the *weak law of large numbers*. The *strong law of large numbers* says that

$$\mathbb{P} \left( \lim_{N \rightarrow \infty} \frac{S_N}{N} = \mu \right) = 1. \quad (2)$$

However, the strong law of large numbers requires that an infinite sequence of random variables is well-defined on the underlying probability space. The existence of these objects however has only been proved in the 20th century. In contrary, Bernoulli's law of large numbers only requires probabilities of finite sequences. For instance, for the Bernoulli experiment they are described by the binomial distribution given by

$$\mathbb{P}(S_N = k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad \text{for } k = 0, \dots, N.$$

This allows a direct evaluation of the limit in (1).

## 2 Bernoulli's proof

Bernoulli proves the result in a slightly more restricted situation. He considers an urn containing  $r$  red and  $b$  black balls. The probability of choosing a red ball is given by  $p = r/(r + b)$ . Bernoulli chooses  $\varepsilon = 1/(r + b)$  and then he investigates  $N = n(r + b)$  drawings with replacement as  $n \rightarrow \infty$ . For instance, the choices  $r = 3$  and  $b = 2$  provide  $p = 0.6$  and  $\varepsilon = 0.2$  in this set-up. For the same experiment he can also choose multiples of 3 and 2 which provides results for smaller  $\varepsilon$ 's. His restriction leads to simpler calculations, however, we believe that Bernoulli has been aware of the fact that this restriction on  $\varepsilon$  is irrelevant for the deeper philosophy of the proof.

For simplicity, we only give Bernoulli's proof in the symmetric Bernoulli case  $p = 0.5$  considering  $N$  drawings with replacement, however, with no further restriction on  $\varepsilon \in (0, 1/2)$ . Thus, denote by  $S_N$  the total number of successes in  $N$  i.i.d. Bernoulli experiments having success probability  $p = 0.5$ . For  $\varepsilon \in (0, 1/2)$  we examine, see also (1),

$$\mathbb{P}\left(\left|\frac{S_N}{N} - \frac{1}{2}\right| \geq \varepsilon\right) = 2 \mathbb{P}\left(S_N \geq \left(\frac{1}{2} + \varepsilon\right)N\right), \quad (3)$$

where we have used symmetry. We introduce the following notation for  $k = 0, \dots, N$

$$b_N(k) = \mathbb{P}(S_N = k) = \binom{N}{k} 2^{-N}.$$

By considering the following quotients for  $k < N$ , Bernoulli obtains

$$\frac{b_N(k)}{b_N(k+1)} = \frac{N!}{k!(N-k)!} \frac{(k+1)!(N-k-1)!}{N!} = \frac{k+1}{N-k}, \quad (4)$$

from which he concludes that  $\max_k b_N(k) = b_N(\lceil N/2 \rceil)$ . For simplicity we choose  $N$  to be even, then the latter statement tells us that the maximum of  $b_N(k)$  is taken in the middle of  $\{0, \dots, N\}$ . Observe that from de Moivre's central limit theorem (1733) we know that  $b_N(N/2)$  behaves asymptotically as  $\sqrt{2}/\sqrt{\pi N}$ , but this result has not been known to Bernoulli. Also Stirling's formula has only been discovered later in 1730. For  $j \geq 0$ , Bernoulli furthermore looks at the quotients

$$\frac{b_N(N/2 + j)}{b_N(N/2 + \lceil N\varepsilon \rceil + j)} = \frac{N/2 + j + 1}{N/2 - j} \frac{N/2 + j + 2}{N/2 - j - 1} \cdots \frac{N/2 + j + \lceil N\varepsilon \rceil}{N/2 - j - \lceil N\varepsilon \rceil + 1}. \quad (5)$$

This representation implies for  $j = 0$

$$\lim_{N \rightarrow \infty} \frac{b_N(N/2)}{b_N(N/2 + \lceil N\varepsilon \rceil)} = \lim_{N \rightarrow \infty} \frac{N/2 + 1}{N/2} \frac{N/2 + 2}{N/2 - 1} \cdots \frac{N/2 + \lceil N\varepsilon \rceil}{N/2 - \lceil N\varepsilon \rceil + 1} = \infty;$$

indeed the products under the limit are successively multiplied with additional terms that converge to a constant bigger than 1 as  $N \rightarrow \infty$ . Moreover, the quotients on the right-hand side of (5) are monotonically increasing in  $j$ . This implies *uniform convergence in  $j$*  to  $\infty$ , which in turn implies

$$\lim_{N \rightarrow \infty} \frac{\sum_{j=0}^{\lceil N\varepsilon \rceil - 1} b_N(N/2 + j)}{\sum_{j=0}^{J_{\varepsilon,1}(N)} b_N(N/2 + \lceil N\varepsilon \rceil + j)} \geq \lim_{N \rightarrow \infty} \frac{\sum_{j=0}^{J_{\varepsilon,1}(N)} b_N(N/2 + j)}{\sum_{j=0}^{J_{\varepsilon,1}(N)} b_N(N/2 + \lceil N\varepsilon \rceil + j)} = \infty, \quad (6)$$

where we have set  $J_{\varepsilon,k}(N) = \min\{\lceil N\varepsilon \rceil - 1, N/2 - k\lceil N\varepsilon \rceil\}$  for  $k \in \mathbb{N}$ . Using monotonicity once more we obtain similarly to above

$$\lim_{N \rightarrow \infty} \frac{\sum_{j=0}^{\lceil N\varepsilon \rceil - 1} b_N(N/2 + j)}{\sum_{j=0}^{J_{\varepsilon,k}(N)} b_N(N/2 + k\lceil N\varepsilon \rceil + j)} = \infty, \quad \text{for } k \geq 1. \quad (7)$$

Choose  $1 \leq k \leq N/(2\lceil N\varepsilon \rceil)$  fixed, then the denominator in (7) provides

$$\sum_{j=0}^{J_{\varepsilon,k}(N)} b_N(N/2 + k\lceil N\varepsilon \rceil + j) = \mathbb{P}\left(N/2 + k\lceil N\varepsilon \rceil \leq S_N \leq \min\{N/2 + (k+1)\lceil N\varepsilon \rceil - 1, N\}\right).$$

This implies that for large  $N$  we need at most  $\lceil 1/(2\varepsilon) \rceil$  such events to cover the entire event  $\{S_N \geq N/2 + \lceil N\varepsilon \rceil\}$ . This and (7) provide

$$\lim_{N \rightarrow \infty} \frac{\mathbb{P}(N/2 \leq S_N < N/2 + \lceil N\varepsilon \rceil)}{\mathbb{P}(S_N \geq N/2 + \lceil N\varepsilon \rceil)} = \lim_{N \rightarrow \infty} \frac{\sum_{j=0}^{\lceil N\varepsilon \rceil - 1} b_N(N/2 + j)}{\sum_{j=\lceil N\varepsilon \rceil}^{N/2} b_N(N/2 + j)} = \infty.$$

This and symmetry immediately prove the statement

$$\lim_{N \rightarrow \infty} \mathbb{P}(S_N \geq N/2 + \lceil N\varepsilon \rceil) = 0,$$

which in view of (3) proves Bernoulli's law of large numbers for the symmetric Bernoulli experiment.

The argument of Bernoulli to obtain (6) is interesting because at that time the notion of uniform convergence has not been known. He says that those who are not familiar with considerations at infinity may argue that

$$\lim_{N \rightarrow \infty} \frac{b_N(N/2 + j)}{b_N(N/2 + \lceil N\varepsilon \rceil + j)} = \infty, \quad \text{for every } j,$$

does not imply (6). To dispel this criticism he argues by calculations for finite  $N$ 's and, in fact, he provides an explicit exponential bound for the rate of convergence which today is known as a "large deviation principle". Specifically, Bernoulli derives

$$\mathbb{P}\left(\left|\frac{S_N}{N} - \frac{1}{2}\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon} \exp\left\{-\left[\frac{\varepsilon + 2\varepsilon^2}{2(1 + \varepsilon)} \log(1 + 2\varepsilon)\right] N\right\}. \quad (8)$$

He obtains this bound by evaluating the terms (4) in a clever way; for details see Bolthausen [3]. Though the bound (8) is not optimal, it nevertheless is remarkable. The optimal bound is given by, see Bolthausen [3],

$$2 \exp \left\{ - \left[ \left( \frac{1}{2} + \varepsilon \right) \log(1 + 2\varepsilon) + \left( \frac{1}{2} - \varepsilon \right) \log(1 - 2\varepsilon) \right] N \right\}. \quad (9)$$

The calculations leading to (8) are somewhat hidden in Bernoulli’s manuscript, and at the end he calculates an explicit example for which he receives the result that he needs about 25’000 Bernoulli simulations to obtain sufficiently small deviations from the mean (the correct answer in his example using the optimal bound (9) would have been 6’520). Stigler [7] mentions in his book that perhaps Bernoulli was a little disappointed by that bad rate of convergence and for that reason did not publish these results during his lifetime.

### 3 Importance of Bernoulli’s result

The reason for the delayed publication of *Ars conjectandi* is found in the correspondence to Leibniz already mentioned above. Bernoulli writes that due to his bad health situation it is difficult to complete the work and that the main part of his book is already written. But he then continues by stressing that the most important part, where he demonstrates how to apply the theory of uncertainty to society, ethics and economy, is still missing. Bernoulli then briefly explains that he has started to think about the question how unknown probabilities can be approximated by samples, i.e. how a priori unknowns can be determined by a posteriori observables of a large number of similar experiments. That is, he has started to think about a scientific analysis of questions of public interest using rigorous statistical methods. Of course, this opens the whole field of statistics and raises questions that are still heavily debated by statisticians today. It seems that Leibniz has not been too enthusiastic about Bernoulli’s intention and he argued that the urn model is much too simple to answer real world questions.

Many of the ideas which Jakob Bernoulli no longer was able to complete during his lifetime were taken up by his nephew Nicolaus Bernoulli in his thesis 1709 [2]. It is clear that Nicolaus was familiar with the thoughts of his uncle and he freely used whole passages from the *Ars conjectandi*. He writes that he is “the more eager to present this material as he sees that many very important questions on legal issues can be decided with the help of the art of conjecturing (i.e. probability theory)”. A particularly important topic he addresses are life annuities and their correct valuation. For that he relies on mortality tables by Huygens, Hudde, de Witt and Halley, and he made a clear distinction between the expected and the median lifetime. He realizes that he cannot base the value of a life annuity on the expected survival time, but that one has to take the expected value under the distribution of the remaining life span, because “the price does not grow

proportional with time”, as he writes. Of course, the law of large numbers is crucial for the applicability of these probabilistic computations. For detailed comments on the thesis of Nicolaus Bernoulli and its relation with the *Ars conjectandi*, see [6]

From an actuarial point of view, Bernoulli’s law of large numbers is considered to be the cornerstone and explains why and how insurance works. The main argument being that a pooling of similar uncorrelated risks  $X_1, X_2, \dots$  in an insurance portfolio  $S_N$  provides an equal balance within the portfolio that makes the outcome the more “predictable” the larger the portfolio size  $N$  is. Basically what this says is that for sufficiently large  $N$  there is only a small probability that the total claim  $S_N$  exceeds the threshold  $(\mu + \varepsilon)N$ , see (1), and thus, the bigger the portfolio the smaller the required security margin  $\varepsilon$  (per risk  $X_i$ ) so that the total claim  $S_N$  remains below  $(\mu + \varepsilon)N$  with sufficiently high probability. This is the foundation of the functioning of insurance; it is the aim of every insurance company to build sufficiently large and sufficiently homogeneous portfolios which makes the claim “predictable” up to a small shortfall probability.

## References

- [1] Bernoulli, J. *Wahrscheinlichkeitsrechnung (Ars conjectandi, 1713)*. Ostwalds Klassiker der exakten Wissenschaften, W. Engelmann, Leipzig, 1899.
- [2] Bernoulli, N. *Usu Artis conjectandi in Jure*. Doctoral Thesis, Basel 1709. In: Die Werke von Jakob Bernoulli, Vol. 3, Birkhäuser Basel 1975, 287-326.
- [3] Bolthausen, E. *Bernoullis Gesetz der Grossen Zahlen*. Elemente der Mathematik 65 (2010), 134-143.
- [4] Khinchin, A. *Sur la loi des grands nombres*. Comptes rendus de l’Académie des Sciences 189 (1929), 477-479.
- [5] Kohli, K. *Aus dem Briefwechsel zwischen Leibniz und Jakob Bernoulli*. In: Die Werke von Jakob Bernoulli. Vol. 3, Birkhäuser Basel 1975, 509-513.
- [6] Kohli, K. *Kommentar zur Dissertation von Niklaus Bernoulli*. In: Die Werke von Jakob Bernoulli. Vol. 3, Birkhäuser Basel 1975, 541-557.
- [7] Stigler, S.M. *The History of Statistics: The Measurement of Uncertainty before 1900*. The Belknap Press of the Harvard University Press, Cambridge, Massachusetts 1986.