THE LAW OF LARGE NUMBERS and Part IV

N. H. BINGHAM

Tercentenary Celebration of Jacob Bernoulli's "Ars Conjectandi"

RSS, 26 April 2013

Speakers: A. W. F. Edwards, Ivo Schneider, Eberhard Knobloch, N. H. Bingham, David Bellhouse

1. Before AC

Mathematics, and games of chance, go back to antiquity. Extant probability theory does not. The idea of a "Law of Averages" is folklore, known to the man/woman in the street – but from how far back?

Religion (and/or superstition) play a role: the hand of God/the gods was seen in what we regard as chance outcomes. See e.g.

F. N. DAVID, Games, gods and gambling: The origins and history of probability and statistical ideas from the earliest times to the Newtonian era. Hafner, 1962.

First books on probability:

Girolamo CARDANO (1501-1576), *Liber de ludo aleae*, 1663 (posth., written c. 1563); Christiaan HUYGENS (1629-1695), *De rati-ociniis in ludo aleae*, 1657;

Pierre Raymond de MONTMORT (1678-1719), *Essai d'analyse sur les Jeux de Hazards*, 1708. Todhunter, p.78: "In 1708 he published his work on Chances, where with the courage of Columbus he revealed a new world to mathematicians."

2. Bernoulli, AC and LLN

The first immortal book on probability theory was also posthumous:

James BERNOULLI (1654-1705), *Ars conjectandi*, 1713.

Part I: reprint of Huygens, with commentary; Part II: combinatorics (incl. Bernoulli numbers);

Part III: 24 problems;

Part IV: applications to civics, morals and economics. *Bernoulli's Theorem* (Todhunter, s.123-126, p.71-73): in a sequence of independent Bernoulli trials (outcomes head = 1 with prob. p, tail = 0 with prob. q := 1 - p), for S_n the number of heads in n trials,

 $S_n/n \to p \quad (n \to \infty) \quad \text{in probability},$ i.e. for all $\epsilon > 0$,

 $P(|S_n/n-p| > \epsilon) \to 0 \quad (n \to \infty).$

This is (in modern language) the *Weak Law* of Large Numbers (WLLN) for Bernoulli trials, and begins a mathematical journey lasting 220 years.

3. De Moivre, DC and CLT

Abraham de MOIVRE (1667-1754), *The Doctrine of Chances (DC)*, 1718, 2nd ed. 1738, 3rd ed. 1756, posth. In Bernoulli's Theorem, $S_n/n \rightarrow p$: how fast?

In 1733, de Moivre derived the error curve:

$$\phi(x) := e^{-\frac{1}{2}x^2} / \sqrt{2\pi},$$

thus laying the foundations for modern Statistics. This was in a 7-page note in Latin, *Approximatio* ..., circulated but not published. An English translation was included in the 1738 and 1756 editions of DC. De Moivre showed:

 $P((S_n/n-p)\sqrt{n/pq} \in [a,b]) \rightarrow \int_a^b \phi(x)dx \quad (n \to \infty).$

This is the *de Moivre-Laplace* theorem, the special case for Bernoulli trials of the Central Limit Theorem (CLT) – known to the physicist in the street as the "Law of Errors". Since $P(S_n = k) = {n \choose k} p^k q^{n-k}$, this involves *Stirling's* formula, $n! \sim \sqrt{2\pi}e^{-n}n^{n+\frac{1}{2}}$ ($n \to \infty$) (James STIRLING (1692-1770); *Methodus Differentialis*, 1730).

4. Laplace, TAP and CLT

Pierre Simon de LAPLACE (1749-1827), *Théorie Analytiques des Probabilités (TAP)*, 1812, 2nd ed. 1814, 3rd ed. 1820.

Laplace was called the French Newton. In addition to TAP, he is remembered for his *Mécanique Céleste (MC)*, 5 volumes, 1799-1825. In TAP, Laplace extended de Moivre's theorem (hence the name de Moivre-Laplace CLT), and treated the *method of least squares*. Least squares was introduced by Legendre, and studied by Gauss; there was a priority dispute between Legendre and Gauss on this. The whole area of the normal (or Gaussian) distribution(s), CLT and least squares was studied by Gauss and Laplace (the two greatest mathematicians to that date who had worked on probability), leading to the *Gauss-Laplace synthesis* in TAP.

This ends the first Heroic Period of probability and statistics.

5. Heroic Periods and the Long Pause

There follows the Long Pause (Quetelet, Bienaymé, Lexis, ...).

The second Heroic Period relates more to Statistics, through the work of the English School (Galton, K. Pearson, Edgeworth, ...). See e.g. S. M. STIGLER, *The history of statistics: The measurement of uncertainty before 1900.* Harvard UP, 1986;

NHB, Heroic Periods and the Long Pause, *Mathématiques et Sciences humaines – Mathematics and Social Sciences* **176**, 2006, 31-42. Complementing the normal (Gaussian) distribution and LLN is the Poisson distribution and the *Law of Small numbers*:

 $P(X = k) = e^{-\lambda} \lambda^k / k!$ (k = 0, 1, 2, ...)

S. D. POISSON (1781-1840) in 1837); L. von BORTKIEWICZ (1868-1931), *Das Gesetz der kleinen Zahlen*, 1898

(deaths by kicks from a horse in the Prussian cavalry; accidents, insurance, actuarial mathematics).

6. The Russian School

Probability theory during this Long Pause (in Statistics) sees the beginning of the great Russian school:

P. L. CHEBYSHEV (TCHEBYCHEV, etc.) (1821-94), Professor at St. Petersburg U. *Chebyshev's Inequality* (*J. Math. Pures Appl.* 88 (1867), 177-184):

 $P(|X - E[X]| \ge \epsilon) \le var X/\epsilon^2.$

Hence, for Bernoulli trials,

 $P(|S_n/n-p| \ge \epsilon) \le pq/(n\epsilon^2).$

Chebyshev also introduced the *method of moments* (1890/91, *Acta Math.*). His most distinguished pupil was A. A. MARKOV (1856-1922), for whom Markov chains and processes are named.

Markov addressed the St. Petersburg Academy in 1913 on the *bicentennial* of LLN.

A. M. LYAPUNOV (1857-1918), in two papers of 1900 and 1901, introduced *characteristic functions*, and proved the CLT under weaker conditions (not requiring all moments).

7. The 20th C and Measure Theory

Henri LEBESGUE (1875-1941): Intégrale, longueur, aire, Annali di Math. 7 (1902), 231-259. Lebesgue introduced Measure Theory, the mathematics of length, area and volume, and gave his new integral, the Lebesgue integral. It turns out that this was also the mathematics of probability, with the integral as expectation. Emile BOREL (1871-1956); Borel's Normal Number Theorem of 1909. Almost all numbers are normal to all bases simultaneously (each digit in their decimal expansions has frequency 1/10; similarly for dyadic expansions, etc. Note. 1. We have no explicit example! 2. Borel's proof was not fully rigorous. Felix HAUSDORFF (1868-1942), Grundzüge der Mengenlehre, 1914 (2nd ed. 1927). The first edition his book (which introduces General Topology) contains a rigorous proof of the Strong Law of Large Numbers for Bernoulli trials: writing "a.s." (almost surely) for "with probability 1",

 $S_n/n \to p \quad (n \to \infty) \quad a.s.$

8. The 20th C and the Russian School

Alexandr Yakovlevich KHINCHIN (1894-1959). Khinchin was a Russian analyst and probabilist (and one of the two teachers of Kolmogorov, below).

In 1924, he proved the *Law of the Iterated Logarithm (LIL)*, which "splits the difference between LLN and CLT", and completes this trilogy of classical limit theorems.

In 1925, Khinchin and Kolmogorov published a paper on limit theorems (Kolmogorov's first work in probability). Khinchin's work in the 1920s included a complete treatment of WLLN by modern methods.

Andrei Nikolaevich KOLMOGOROV (1903-87) – *the* great name in Probability Theory.

Grundbegriffe der Wahrscheinlichkeitsrechnung, 1933.

This book successfully harnesses Measure Theory to the service of Probability Theory, following earlier work in this direction by Maurice

FRÉCHET (1878-1973) and others.

Why did this take so long? See e.g.

NHB, Measure into probability: from Lebesgue to Kolmogorov, Studies in the history of probability and statistics **XLVI**. *Biometrika* **87** (2000), 145-156.

The *Grundbegriffe* did two other vital things. First (Daniell-Kolmogorov theorem) it showed how a suitably consistent family of 'finitedimensional distributions' give one 'infinitedimensional distribution' (stochastic process). Secondly, it gave the definitive form of the SLLN, *Kolmogorov's SLLN*: if X, X_1, X_2, \ldots are independent and identically distributed (ii) random variables, the following are equivalent:

$$\frac{1}{n}\sum_{1}^{n}X_{i} \to c \quad (n \to \infty) \quad a.s.;$$

 $E[|X|] < \infty \quad \& \quad E[X] = c.$

This is the definitive mathematical version of the "Law of Averages" with which we began. This year is its 80th anniversary, as well as AC's 300th.

9. Ergodic theorems

Ergodic theory has its roots in Statistical Physics in the 19th C, in particular Statistical Mechanics and Thermodynamics, the creation of three men: the Scot James Clerk MAXWELL (1831-1879), the Austrian Ludwig BOLTZMANN (1844-1906) and the American Josiah Willard GIBBS (1839-1903). In particular, the distribution of velocity of particles in a gas is given by a Gaussian, the Maxwell-Boltzmann distribution. Lebesgue's creation of Measure Theory led to advances in Statistical Mechanics, and the study of measure-preserving transformations in phase space (high-dimensional: 6^N , with N particles: position and velocity are both 3-vectors). In 1931 George D. BIRKHOFF (1884-1944) proved his ergodic theorem: an analogue of the SLLN for measure-preserving transformations (for dynamical systems only - extended to the general case by Khinchin in 1932 – the Birkhoff-Khinchin ergodic theorem). The direct statement extends Kolmogorov's SLLN of 1933, but there is no converse.

10. Martingales

Martingales model fair games. If \mathcal{F}_t represents the information available at time t (so these increase with t), $X = (X_t)$ is a martingale (mg) if $E[|X_t|] < \infty$ for all t, X_t is \mathcal{F}_t -measurable (known at time t), and for s < t

 $E[X_t | \mathcal{F}_s] = X_s.$

These were developed by Paul LÉVY (1886-1971) and J. L. DOOB (1910-2004) in the 40s and 50s. By Doob's *martingale convergence theorem*, an L_1 -bounded mg is a.s. convergent: if $\sup_n E[|X_n|] < \infty$, then for some X_{∞} , $X_n \to X_{\infty}$ as $n \to \infty$. This generalises (the direct half of) Kolmogorov's SLLN.

Some mgs – the uniformly integrable (UI) ones – are convergent a.s. and in mean (in L_1):

 $X_n = E[X_{\infty}|\mathcal{F}_n], \quad X_n \to X_{\infty}$ a.s. and in L_1 .

These are the mgs used in math. finance. The quick way to prove Kolmogorov's SLLN by mg methods is to use the *reverse mg convergence theorem*.

11. Higher dimensions

All this holds when we work on the real line \mathbb{R} (or the complex plane \mathbb{C}). Similarly for d dimensions, in \mathbb{R}^d or \mathbb{C}^d . All this holds also for *Hilbert space* (David HILBERT (1862-1943)), which we can think of as *Euclidean space of infinitely many dimensions* (there is an inner product; Pythagoras' theorem holds; we can use projections as in Euclidean space; so we can think geometrically, and draw diagrams). The SLLN holds also in *Banach space* (Stefan BANACH (1892-1945), in 1932) – Robert FORTET (1912-98) in 1954.

Some more detailed results hold in *some but not all* Banach spaces B, and single out some interesting geometrical property of B. E. g.: (i) the Martingale Convergence Theorem and the Radon-Nikodym Theorem (of Measure Theory) are essentially the same theorem: one holds in a Banach space B iff the other does. (ii) the *Marcinkiewicz-Zygmund LLN* (SLLN with *p*th moments, 0) holds in <math>B iff Bhas *type* p.

NHB, 26.4.13.