The Tricentenary of the Weak Law of Large Numbers.

Eugene Seneta

presented by Peter Taylor

July 8, 2013

Jacob and Nicolaus Bernoulli

Jacob Bernoulli (1654–1705)

In 1687 Jacob Bernoulli (1654–1705) became Professor of Mathematics at the University of Basel, and remained in this position until his death.

Jacob and Nicolaus Bernoulli

- The title of Jacob Bernoulli's work *Ars Conjectandi (The Art of Conjecturing)* was an emulation of the *Ars Cogitandi (The Art of Thinking)*, of Blaise Pascal. Pascal's writings were a major influence on Bernoulli's creation.
- Jacob Bernoulli was steeped in Calvinism. He was thus a firm believer in predestination, as opposed to free will, and hence in determinism in respect of “random" phenomena. This coloured his view on the origins of *statistical regularity* in nature, and led to its mathematical formalization, as *Jacob Bernoulli's Theorem*, the first version of the *Law of Large Numbers*.
- Jacob Bernoulli's *Ars Conjectandi* remained unfinished in its final part, the *Pars Quarta*, the part which contains the Theorem, at the time of his death.

Jacob and Nicolaus Bernoulli

- Nicolaus Bernoulli (1687-1759) was Jacob's nephew. With Pierre Rémond de Montmort (1678-1719) and Abraham De Moivre (1667-1754), he was the leading figure in "the great leap forward in stochastics", the period from 1708 to the first edition of De Moivre's *Doctrine of Chances* in 1718.
- In early 1713, Nicolaus helped Montmort prepare the second edition of his book *Essay d'analyse sur les jeux d'hasard*, and returned to Basel in April, 1713, in time to write a preface to *Ars Conjectandi* which appeared in August 1713, a few months before Montmort's book, whose tricentenary we also celebrate.

Jacob and Nicolaus Bernoulli

- In his preface to *Ars Conjectandi* in 1713, Nicolaus says of the fourth part that Jacob intended to apply what he had written in the earlier parts to civic, moral and economic questions, but due to prolonged illness and untimely death, Jacob left it incomplete. Describing himself as too young and inexperienced to complete it, Nicolaus decided to let the *Ars Conjectandi* be published in the form in which its author left it.

Jacob Bernoulli's Theorem

In modern notation Bernoulli showed that, for fixed p, any given small positive number ϵ, and any given large positive number c,

$$P(|\frac{X}{n} - p| > \epsilon) < \frac{1}{c+1}$$

for $n \geq n_0(\epsilon, c)$.

- Here X is the number of successes in n binomial trials relating to sampling with replacement from a collection of $r + s$ items, of which r were “fertile" and s “sterile", so that $p = r/(r + s)$.

Jacob Bernoulli's Theorem

- Bernouilli's conclusion was that $n_0(\epsilon, c)$ could be taken as the integer greater than or equal to:

$$(r+s)\max\left\{\frac{\log c(s-1)}{\log(r+1)-\log r}\left(1+\frac{s}{r+1}\right)-\frac{s}{r+1},\right.$$
$$\left.\frac{\log c(r-1)}{\log(s+1)-\log s}\left(1+\frac{r}{s+1}\right)-\frac{r}{s+1}\right\}.$$

- Jacob Bernoulli's concluding numerical example takes $r = 30$ and $s = 20$, so $p = 3/5$, and $\epsilon = 1/50$. With $c = 1000$, he derived the (no doubt disappointing) result $n_0(\epsilon, c) = 25,550$. A small step for Jacob Bernoulli, but a very large step for stochastics.

De Moivre

- De Moivre (1730) distinguished clearly between the approach of Jacob Bernoulli in 1713 in finding an n sufficiently large *for specified precision*, and of Nicolaus Bernoulli *of assessing precision* for fixed n for the “futurum probabilitate", alluding to the fact that the work was for a general, and to be estimated, p, on which their *bounds* depended.
- In the English translation of his 1733 paper, De Moivre (1738) praised the work of the Bernoullis on the summing of several terms of the binomial term $(a + b)^n$ when n is large, but says

. . . yet some things were further required; for what they have done is not so much an Approximation as the determining of very wide limits, within which they demonstrated that the sum of the terms was contained.

De Moivre

- De Moivre's (1733) motivation was *to approximate sums of individual binomial probabilities when n is large*, and the probability of success in a single trial is p, that is when $X \sim B(n, p)$. His initial focus was on the symmetric case $p = 1/2$.
- De Moivre's results provide a strikingly simple, good, and easy-to-apply approximation to binomial sums, in terms of an integral of the normal density curve. His (1733) theorem may be stated as follows in modern terms. For any $s > 0$ and $0 < p = 1 - q < 1$, the sum of the binomial terms

$$\sum \binom{n}{x} p^x q^{n-x}$$

over the range $|x - np| \leq s\sqrt{npq}$, approaches as $n \to \infty$, the limit

$$\frac{1}{\sqrt{2\pi}} \int_{-s}^{s} e^{-z^2/2} dz.$$

De Moivre

- The focus of De Moivre's application of his result, the limit aspect of Jacob Bernoulli's Theorem, also revolves conceptually around the mathematical formalization of statistical regularity, the empirical phenomenon that De Moivre attributed to

  ... that Order which naturally results from ORIGINAL DESIGN.
- De Moivre's (1733) result already contained an approximate answer, via the normal distribution to *estimating precision* of the relative frequency X/n as an estimate of an *unknown* p, for *given* n; or of *determining* n for *given* precision (the inverse problem), in *frequentist fashion*, using the inequality $p(1-p) \leq 1/4$.

Laplace, the Inversion Problem and the Centenary

- In a paper of 1774, the young Pierre Simon de Laplace (1749-1827) saw that *Bayes' Theorem* provides a means to solution of Jacob Bernoulli's *inversion problem.*
- Laplace considered binomial trials with success probability x in each trial, assuming x has uniform prior distribution on $(0, 1)$, and calculated the posterior distribution of the success probability random variable Θ after observing p successes and q failures. Its density is:

$$\frac{\theta^p(1-\theta)^q}{\int_0^1 \theta^p(1-\theta)^q d\theta} = \frac{(p+q+1)!}{p!q!}\theta^p(1-\theta)^q$$

and Laplace proved that for any given $w > 0, \delta > 0$

$$P(|\Theta - \frac{p}{p+q}| < w) > 1 - \delta$$

for large p, q.

Laplace, the Inversion Problem and the Centenary

- This is a *Bayesian analogue of Jacob Bernoulli's Theorem*, the beginning of Bayesian estimation theory of success probability of binomial trials and of Bayesian-type LLN and Central Limit theorems. Early in the paper Laplace took the *mean*

$$\frac{p+1}{p+q+1}$$

 of the posterior distribution as his total predictive probability on the basis of observing p and q, and this is what we now call the Bayes estimator.

Laplace, the Inversion Problem and the Centenary

- The first (1812) and the second (1814) edition of Laplace's *Théorie analytique des probabilités* span the centenary year of Bernoulli's Theorem. The (1814) edition is an outstanding epoch in the development of probability theory.
- Laplace's (1814), Chapitre III, is frequentist in approach, contains De Moivre's Theorem, and in fact adds a continuity correction term (p. 277):

$$P(|X - np| \leq t\sqrt{npq}) \approx \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-u^2/2} du + \frac{e^{-t^2/2}}{\sqrt{2\pi npq}}.$$

  Laplace remarked that this is an approximation to $O(n^{-1})$, provided that np is an integer.

Laplace, the Inversion Problem and the Centenary

- On p.282 Laplace inverted this expression to give an interval for p centred on $\hat{p} = X/n$, but the ends of the interval still depend on the unknown p, which Laplace replaces by $\hat{p}$, since n is large. This gives an interval of random length, in fact a confidence interval in modern terminology, for p.
- In Laplace's (1814) *Notice historique sur le Calcul des Probabilités*, both Bernoullis, Montmort, De Moivre and Stirling receive due credit. In particular a paragraph refers to De Moivre's Theorem, in both its contexts, that is as facilitating a proof of Jacob Bernoulli's Theorem; and as:

  . . . an elegant and simple expression that the difference between these two ratios will be contained within the given limits.

Laplace, the Inversion Problem and the Centenary

- Subsequently to Laplace (1814), while the name and statement of Jacob Bernoulli's Theorem persist, it figures in essence as a frequentist corollary to De Moivre's Theorem; or in its Bayesian version, following the Bayesian (predictive) analogue of De Moivre's Theorem, originating in Laplace (1814), Chapitre VI.
- Finally, Laplace (1814) considered sums of independent integer-valued but not necessarily identically distributed random variables, using their generating functions, and obtained a Central Limit Theorem. The idea of inhomogeneous sums and averages leads directly into subsequent French (Poisson) and Russian (Chebyshev) directions.

Poisson's Law

- The major work in probability of Siméon Denis Poisson (1781-1840) was his book of 1837 *Recherches sur la probabilité.* It is largely a treatise in the tradition of, and a sequel to, that of his great predecessor Laplace's (1814) *Théorie analytique* in its emphasis on the large sample behaviour of averages.
- The term *Loi des grands nombres [Law of Large Numbers]* appears for the first time in the history of probability on p. 7 of Poisson (1837), within the statement

> *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 . . .*

Poisson's Law

- The LLN which is now called Poisson's Law of Large Numbers, has probability of success in the ith trial *fixed*, at $p_i, i = 1, 2, \ldots, n$. Poisson showed that

  $$P(|\frac{X}{n} - \bar{p}(n)| > \epsilon) < Q$$

  for sufficiently large n, using Laplace's Central Limit Theorem for sums of non-identically distributed random variables. The special case where $p_i = p, i = 1, 2, \ldots$ gives Jacob Bernoulli's Theorem, so Poisson's LLN is a genuine generalization.
- Inasmuch as $\bar{p}(n)$ itself need not even converge as $n \to \infty$, Poisson's LLN displays as a primary aspect *loss of variability* of proportions X/n as $n \to \infty$, rather than a tendency to *stability*, which Jacob Bernoulli's Theorem established under the restriction $p_i = p$.

Chebyshev's Thesis

- The 1845 thesis of Pafnutiy Lvovich Chebyshev (1821-1894) at Moscow University was entitled *An Essay in Elementary Analysis of the Theory of Probabilities.*
- Much of the thesis was in fact devoted to producing tables (correct to seven decimal places) by summation of what are in effect tail probabilities of the standard normal distribution.
- Laplace's (1814) Chapitre VI, on predictive probability, starting with uniform prior on $(0, 1)$ was adapted by Chebyshev to his "discrete" circumstances. Chebyshev's examples were also motivated by Laplace (1814).
- Jacob Bernoulli's Theorem was mentioned at the end of Chebyshev's (1845) thesis, where he proceeded to obtain an approximation to the binomial probability using *bounds* for $x!$ in place of Stirling's approximation.

Chebyshev's Thesis

- Such careful *bounding arguments* (rather than approximate asymptotic expressions) are characteristic of Chebyshev's work, and of the Russian probabilistic tradition which came after him. This is very much in the spirit of the bounds in Jacob Bernoulli's Theorem.
- Poisson's (1837) *Recherches sur la probabilité* came to Chebyshev's attention after the publication of Chebyshev (1845). In his Section 1 Chebyshev (1846) says of Poisson's LLN:

  All the same, no matter how ingenious the method utilized by the splendid geometer, it does not provide bounds on the error in this approximate analysis, and, in consequence of this lack of degree of error, the derivation lacks appropriate rigour.

Chebyshev's Thesis

- For the inhomogeneous case, Chebyshev (1846) repeated his bounds for homogeneous Bernoulli trials which he dealt with in Chebyshev (1845).
- His final result, where, as usual, X stands for the number of successes in n trials, p_i is the probability of success in the ith trial, and $p = \frac{\sum_{i=1}^{n} p_i}{n}$.

$$P(|\frac{X}{n} - p| \geq z) \leq Q$$

if

$$n \geq \max\left\{\left(\frac{\log[Q\frac{z}{1-p}\sqrt{\frac{1-p-z}{p+z}}]}{\log H}\right), \left(\frac{\log[Q\frac{z}{p}\sqrt{\frac{p-z}{1-p+z}}]}{\log H_1}\right)\right\}$$

where

$$H = \left(\frac{p}{p+z}\right)^{p+z}\left(\frac{1-p}{1-p-z}\right)^{1-p-z}, \; H_1 = \left(\frac{p}{p-z}\right)^{p-z}\left(\frac{1-p}{1-p+z}\right)^{1-p+z}.$$

Chebyshev's Thesis

Structurally, these are very similar to Jacob Bernoulli's expressions in his Theorem, so it is relevant to compare what they give in his numerical example when $z = 1/50$, $p = 30/50$, $Q = 1/1001$.

The answer is $n \geq 12241.293$. Compare this with Bernoulli's answer of $25,550$.

The Bienaymé-Chebyshev Inequality

- Irenée Jules Bienaymé (1796-1878) was influenced by the demographic content of Laplace's *Théorie analytique*. He became a fervent devotee of Laplace's work in all its statistical manifestations.
- Bienaymé thought that Poisson's law did not exist as a separate entity from Jacob Bernoulli's Theorem. He did not understand that in Poisson's Law a *fixed* probability of success, p_i is associated with the i-th trial. This misunderstanding led him to develop various generalizations of Jacob Bernoulli's sampling scheme, and so Jacob Bernoulli's theorem.

The Bienaymé-Chebyshev Inequality

- In (1853) Bienaymé showed mathematically that for the sample mean $\bar{X}$ of independently and identically distributed random variables whose mean is μ and variance is σ^2, so $E\bar{X} = \mu,\ Var\bar{X} = \sigma^2/n$, then for any $t > 0$,

$$Pr((\bar{X} - \mu)^2 \geq t^2\sigma^2) \leq 1/(t^2 n)\,.$$

- The proof which Bienaymé used is the simple one that we use in the classroom today. When $EX^2 < \infty$ and $\mu = EX$, for any $\epsilon > 0$,

$$Pr(|X - \mu| \geq \epsilon) \leq (VarX)/\epsilon^2.$$

This is commonly referred to in probability theory as Chebyshev's Inequality, and less commonly as the Bienaymé-Chebyshev Inequality.

The Bienaymé-Chebyshev Inequality

- If the $X_i, i = 1, 2, \ldots$ are independent, but not necessarily identically, distributed, and $S_n = X_1 + X_2 + \cdots + X_n$, we similarly obtain

$$Pr(|S_n - ES_n| \geq \epsilon) \leq (\Sigma_{i=1}^{n} VarX_i)/\epsilon^2.$$

This inequality was obtained by Chebyshev (1867) for discrete random variables and published simultaneously in French and Russian. Bienaymé (1853) was reprinted immediately preceding the French version in Liouville's journal.

The Bienaymé-Chebyshev Inequality

In 1874 Chebyshev wrote

> *The simple and rigorous demonstration of Bernoulli's law to be found in my note entitled: Des valeurs moyennes, is only one of the results easily deduced from the method of M. Bienaymé , which led him, himself, to demonstrate a theorem on probabilities, from which Bernoulli's law follows immediately . . .*

The Bienaymé-Chebyshev Inequality

- Actually, not only *the limit theorem aspect of Jacob Bernoulli's Theorem* is covered by the Bienaymé-Chebyshev Inequality, but also *the inversion aspect*, by using $p(1 - p) \leq 1/4$ to allow *for unspecified* p. The result is exact, but for Jacob Bernoulli's example the conclusion is weak.

Sample Size in Jacob Bernoulli's Example

- The normal approximation to the binomial in the manner of De Moivre can be used to determine n for specified precision *if p is known*. For Bernoulli's example where $r = 30$, $s = 20$, $p = 3/5$, $c = 1000$, and $\epsilon = 1/50$ the result is $n_0(\epsilon, c) \geq 6498$.
- To effect "approximate" inversion *if we do not know the value of p,* to get the specified accuracy of the *estimate* of p presuming that n would still be large, we could use De Moivre's Theorem and the "worst case" bound $p(1-p) \leq 1/4$, to obtain

$$n \geq \frac{z_0^2}{4\epsilon^2} = 0.25(3.290527)^2(50)^2 = 6767.23 \geq 6767$$

where $P(|Z| \leq z_0) = 0.999001$. The now commonly used substitution of the estimate $\hat{p}$ from a preliminary performance of the binomial experiment in place of p in $p(1-p)$ would improve the inversion result.

Sample Size in Jacob Bernoulli's Example

- In the tradition of Chebyshev, Markov (1899) had developed a method using continued fractions to obtain tight *bounds* for binomial probabilities when p is known and n is also prespecified. In looking for smallest n for given p and given precision, he began with an approximate n ($n = 6498$ for Jacob Bernoulli's example) and then examined bounds on precision for n in the vicinity. For this example he decided n was at most 6520.
- Recall that if $p \neq 1/2$ one problem with the normal approximation to the bionomial is that the asymmetry about the mean is not reflected. Thus,

$$\frac{c}{c+1} < P(|\frac{X}{n} - p| \leq \epsilon) = P(X \leq np + n\epsilon) - P(X < np - n\epsilon)$$

involves binomial tails of differing probability size.

Sample Size in Jacob Bernoulli's Example

For this classical example when $p = 0.6$, we seek the smallest n to satisfy

$$0.9990009999 = \frac{1000}{1001} < P(X \leq 0.62n) - P(X < 0.58n)$$

where $X \sim B(n, 0.6)$.

Using **R**, $n = 6491$ on the right hand side gives 0.9990126, while $n = 6490$ gives 0.9989679, so the minimal n which will do is 6491.

The Bicentenary in St. Petersburg

In a letter from Markov to Chuprov, 15 January, 1913, Markov wrote

> *Firstly, do you know: the year 1913 is the two hundredth anniversary of the law of large numbers (Ars Conjectandi, 1713), and don't you think that this anniversary should be commemorated in some way or other? Personally I propose to put out a new edition of my book, substantially expanded.*

The Bicentenary in St. Petersburg

Then in a letter to Chuprov, (31 January, 1913), Markov wrote

> *. . . Besides you and me, it was proposed to bring in Professor A.V. Vasiliev . . . Then it was proposed to translate only the fourth chapter of Ars Conjectandi; the translation will be done by the mathematician Ya.V. Uspensky, who knows the Latin language well, and it should appear in 1913. All of this should be scheduled for 1913 and a portrait of J. Bernoulli will be attached to all the publications.*

The Bicentenary in St. Petersburg

- The respective topics presented were: Vasiliev: Some questions of the theory of probabilities up to the theorem of Bernoulli; Markov: The Law of Large Numbers considered as a collection of mathematical theorems; Chuprov: The Law of Large Numbers in contemporary science.
- The early part of Markov's talk contrasted Jacob Bernoulli's *exact* results with the *approximate procedures* of De Moivre and Laplace, which use the limit normal integral structure to determine probabilities. Markov mentions Laplace's second degree correction, and also comments on the *proof* of Jacob Bernoulli's Theorem in its limit aspect by way of the DeMoivre-Laplace "second limit theorem".

The Bicentenary in St. Petersburg

- Markov went on to discuss Poisson's LLN as an approximate procedure " . . . not bounding the error in an appropriate way", and continues with Chebyshev's (1846) proof in Crelle's journal. He then summarizes the Bienaymé - Chebyshev interaction in regard to the Inequality and its application; and the evolution of the method of moments.

The Bicentenary in St. Petersburg

Markov concluded his talk as follows, in a story which has become familiar.

> *. . . I return to Jacob Bernoulli. His biographers recall that, following the example of Archimedes he requested that on his tombstone the logarithmic spiral be inscribed with the epitaph “Eadem mutato resurgo". . . . It also expresses Bernoulli’s hope for resurrection and eternal life. . . . More than two hundred years have passed since Bernoulli’s death but he lives and will live in his theorem.*

Markov (1913) and Markov's Theorems

Andrei A. Markov (1856–1922)

The Bicentenary edition

- The translation from Latin into Russian by J.V. Uspensky was published in 1913, edited, and with a Foreword, by Markov.
- To celebrate the bicentenary, Markov published in 1913 the 3rd substantially expanded edition of his celebrated monograph *Ischislenie Veroiatnostei [Calculus of Probabilities]*. The title page is headed

K 200 lietnemu iubileiu zakona bol'shkh chisel.
[To the 200th-year jubilee of the law of large numbers.]

with the title *Ischislenie Veroiatnostei* below it.

The Bicentenary edition

- For the portrait of Jacob Benoulli following the title page, Markov expressed his gratitude to the chief librarian of Basel University, Dr. Carl Christoph Bernoulli.
- In this 3rd Bicentenary edition, Chapter III (pp. 51-112), is titled *The Law of Large Numbers.*
- Of specific interest to us is what has come to be known as *Markov's Inequality*: for a non-negative random variable U and positive number u

$$P(U \geq u) \leq \frac{E(U)}{u}$$

which occurs as a Lemma on p. 61-63. It is then used to prove the Bienaymé-Chebyshev Inequality, on pp. 63-65, in what has become the standard modern manner, inherent already in Bienaymé's (1853) proof.

Markov's Theorems

- Section 16 (of Chapter III) is entitled *The Possibility of Further Extensions.* On p. 76 Markov asserted that

$$\frac{Var(S_n)}{n^2} \to 0 \text{ as } n \to \infty$$

  is sufficient for the WLLN to hold, for arbitrary summands $\{X_1, X_2, \ldots\}$.
- Thus *the assumption of independence is dropped, although the assumption of finite individual variances is retained.* In the Russian literature, for example in Bernstein's (1927) textbook, this is called *Markov's Theorem.* We shall call it Markov's Theorem 1.

Markov's Theorems

- Amongst the innovations in this 3rd edition was an advanced version of the WLLN which came to be known also as *Markov's Theorem*, and which we shall call Markov's Theorem 2:

$$\frac{S_n}{n} - E\left(\frac{S_n}{n}\right) \xrightarrow{p} 0$$

  where $S_n = \sum_{i=1}^{n} X_i$ and the $\{X_i, i = 1, 2, \ldots\}$ are independent and satisfy $E(|X_i|^{1+\delta}) < C < \infty$ for some constants $\delta > 0$ and C. The case $\delta = 1$ came to be known in Russian-language literature as *Chebyshev's Theorem.*
- Markov's Theorem 2 thus *dispenses with the need for finite variance of summands X_i, but retains their independence.*

Markov's Theorems

- Markov's publications of 1914 strongly reflect his background reading activity in preparation for the Bicentenary. In particular, in a paper entitled *O zadache Yakova Bernoulli [On the problem of Jacob Bernoulli]*, in place of what Markov calls the approximate formula of De Moivre,

$$\frac{1}{\sqrt{\pi}} \int_z^\infty e^{-z^2} dz \text{ for } P(X > np + z\sqrt{2npq})$$

he derived the expression

$$\frac{1}{\sqrt{\pi}} \int_z^\infty e^{-z^2} dz + \frac{(1-2z^2)(p-q)e^{-z^2}}{6\sqrt{2npq\pi}}$$

which Markov calls Chebyshev's formula. This paper of Markov's clearly motivated Uspensky (1937) in his English-language monograph to ultimately resolve the issue.

Bernstein's monograph (1927)

- Markov died in 1922 well after the Bolshevik seizure of power, and it was through the 4th (1924, posthumous) edition of *Ischislenie Veroiatnestei* that his results were publicized and extended, in the first instance in the Soviet Union due to the monograph S.N. Bernstein (1927).
- The third part of Bernstein's book was titled *The Law of Large Numbers* and consisted of three chapters: Chapter 1: *Chebyshev's inequality and its consequences.* Chapter 2: *Refinement of Chebyshev's Inequaliity.* and Chapter 3: *Extension of the Law of Large Numbers to dependent quantitities.* Chapter 3 began with Markov's Theorem 1. Markov's Theorem 2 was mentioned, and a proof was included in the second edition, Bernstein (1934).

Bernstein's monograph (1927)

- Bernstein (1924) returned to the problem of accuracy of the normal approximation to the binomial via bounds. He showed that there exists an α $(|\alpha| \leq 1)$ such that $P = \Sigma_x \binom{n}{x} p^x q^{n-x}$ summed over x satisfying $|x - np - \frac{t^2}{6}(q-p)| < t\sqrt{npq} + \alpha$ is

$$\frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-u^2/2} du + 2\theta e^{-(2npq)^{1/3}}$$

where $|\theta| < 1$ for any n, t, provided that

$$npq \geq \max(t^2/16, 365).$$

The tool used, perhaps for the first time ever, was what came to be known as *Bernstein's Inequality*.

Bernstein's monograph (1927)

- Bernstein's Inequality reads

$$P(V > v) \leq e^{-v\epsilon} E(e^{V\epsilon})$$

for any $\epsilon > 0$, which follows from Markov's Inequality $P(U > u) \leq E(U)/u$.

If $E(e^{V\epsilon}) < \infty$, the bound is particularly effective for a non-negative random variable V such as the binomial, since the bound may be tightened by manipulating ϵ.

Uspensky's monograph (1937)

- The entire issue of normal approximation to the binomial was resolved into an ultimate exact form by Uspensky (1937) who showed that P taken over the usual range $t_1\sqrt{npq} \leq x - np \leq t_2\sqrt{npq}$ for any real numbers $t_1 < t_2$, can be expressed as

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-u^2/2} du &+ \frac{(1/2-\theta_1)e^{-t_1^2/2} + (1/2-\theta_2)e^{-t_2^2/2}}{\sqrt{2\pi npq}} \\
&+ \frac{(q-p)\{(1-t_2^2)e^{-t_2^2/2} - (1-t_1^2)e^{-t_1^2/2}\}}{6\sqrt{2\pi npq}} \\
&+ \Omega,
\end{aligned}
$$

where θ_1 and θ_2 have explicit expressions and $|\Omega|$ is suitably bounded.

Uspensky's monograph (1937)

- The symmetric case follows by putting $t_2 = -t_1 = t$ so the "Chebyshev" term vanishes. When both np and $t\sqrt{npq}$ are integers, $\theta_1 = \theta_2 = 0$, reducing the correction term to Laplace's $e^{-t^2/2}/\sqrt{2\pi npq}$. But in any case, bounds which are within $O(n^{-1})$ of the true value are thus available.
- Uspensky's (1937) book carried Markov's theory to the English-speaking countries. Uspensky (1937) cited Markov (1924) and Bernstein (1927) in his two-chapter discussion of the LLN. Markov's Theorem 2 was stated and proved.
- The ideas in the proof of Markov's Theorem 2 were used to prove the now famous "Khinchin's Theorem", an ultimate form of the WLLN.

Uspensky's monograph (1937)

- For independent identically distributed (iid), Khinchin (Khintchine (1929)) showed that the existence of a finite mean, $\mu = EX_i$, is sufficient for the Weak Law of Large Numbers. Finally, Uspensky (1937), proved the Strong Law of Large Numbers (SLLN) for the setting of Bernoulli's Theorem, and called this strengthening "Cantelli's Theorem".

Bernstein's monograph (1934)

- Bernstein (1934), in his third part has an additional Chapter 4: *Statistical probabilities, average values and the coefficient of dispersion*. It begins with a *precise Bayesian inversion* of Jacob Bernoulli's Theorem, proved under a certain condition on the prior distribution of the number of "successes", X, in n trials. The methodology uses Markov's Inequality applied to $P\big((\Theta - \frac{X}{n})^4 > w^4 \big| \Theta\big)$ and, in the classical case of a uniform prior distribution over $(0, 1)$ of the success probability Θ, gives for any $w > 0$

$$P\big(|\Theta - \frac{X}{n}| < w \big| X = m\big) > 1 - \frac{3(n_0 + 1)}{16nw^4 n_0}$$

  for $n > n_0$ and $m = 0, 1, \ldots, n$. This apparently little-known result can be seen to be a precise version of Laplace's theorem.

Bernstein's monograph (1934)

- Bernstein (1934) also had four new appendices. The fourth of these is titled *A Theorem Inverse to Laplace's Theorem*. This is the Bayesian inverse of De Moivre's Theorem, with an arbitrary prior density, and convergence to the standard normal integral as $m, n \to \infty$ provided that m/n behaves appropriately. A version of this theorem is now called the *Bernstein-von Mises Theorem*.

Necessary and Sufficient Conditions.

- The expression

$$\frac{S_n}{n} - E\left(\frac{S_n}{n}\right) \xrightarrow{p} 0$$

 is the classical form of what is now called the WLLN. We have confined ourselves to sufficient conditions for this result to hold, where $S_n = \sum_{i=1}^{n} X_i$ and the $\{X_i, i = 1, 2, \ldots\}$ are independent and not necessarily identically distributed.

- In particular, in the tradition of Jacob Bernoulli's Theorem as limit theorem, we have focused on the case of "Bernoulli" summands where $P(X_i = 1) = p_i = 1 - P(X_i = 0)$.

Necessary and Sufficient Conditions.

- From the 1920s attention had turned to *necessary and sufficient* conditions for the WLLN for *independent* summands. Kolmogorov in 1928 obtained the first such condition for “triangular arrays", and there were generalizations by Feller in 1937 and Gnedenko in 1944.

Necessary and Sufficient Conditions.

- In another paper (Khintchine (1936)) on the WLLN in Cantelli's journal, *Giorn. Ist. Ital. Attuari*, Khintchine turned his attention to necessary and sufficient conditions for the existence of a sequence $\{d_n\}$ of positive numbers such that

$$\frac{S_n}{d_n} \xrightarrow{p} 1 \text{ as } n \to \infty$$

where the (iid) summands X_i are *non-negative*.

Necessary and Sufficient Conditions.

- Two new features in the consideration of limit theory for iid summands make their appearance in Khinchin's several papers in Cantelli's journal: a focus on the *asymptotic structure of the tails of the distribution function*, and the expression of this structure in terms of what was later realized to be *regularly varying functions*.

  Putting $F(x) = P(X_i \leq x)$ and $\nu(x) = \int_0^x (1 - F(u))du$, Khinchin's necessary and sufficient condition for the WLLN is $\frac{x(1-F(x))}{\nu(x)} \to 0$ as $x \to \infty$. This is equivalent to $\nu(x)$ being a slowly varying function at infinity. In this event, d_n can be taken as the unique solution of $n\nu(d_n) = d_n$.

Necessary and Sufficient Conditions.

- Khinchin's Theorem itself was generalized by Feller (see for example Feller (1966) Section VII.7) in the spirit of Khintchine (1936) for iid, but not necessarily nonnegative, summands.
- Petrov's (1995) book gives necessary and sufficient conditions for the existence of a sequence of constants $\{b_n\}$ such that $S_n/a_n - b_n \to 0$ for any given sequence of positive constants $\{a_n\}$ such that $a_n \to \infty$, where the independent summands X_i are not necessarily identically distributed.
- There is a little-known necessary and sufficient condition for the WLLN, due to Gnedenko, for arbitrarily dependent not necessarily identically distributed random variables (Gnedenko's (1963) textbook).

Conclusion.

There is much more to say, and more is said in this year's special issue of the appropriately named journal, *Bernoulli.*

This is a good time and place to stop.

The precise reference is :

Seneta, E. (2013) A Tricentenary history of the Law of Large Numbers. Bernoulli 19(4), 1088-1121.