Mat-2.108 Sovelletun Matematiikan erikoistyöt

The history of the central limit theorem

Max Mether
45748T
06.11.03

Table of Contents

1 Introduction

One of the most important theorems in statistical mathematics and probability theory is the Central Limit Theorem (CLT). It is used almost everywhere where statistical mathematics is applied. The usefulness of the theorem lies in its simple definition. The central limit theorem states that if some certain conditions are satisfied, then the distribution of the arithmetic mean of a number of independent random variables approaches a normal distribution as the number of variables approaches infinity. In other words, there is no need to know very much about the actual distribution of the variables, as long as there are enough instances of them their sum can be treated as normally distributed. The beauty of the theorem thus lies in its simplicity. As an example, we show the distribution of the sum of uniform distributions with 1, 2, 8, and 32 summands respectively in Figure 1.
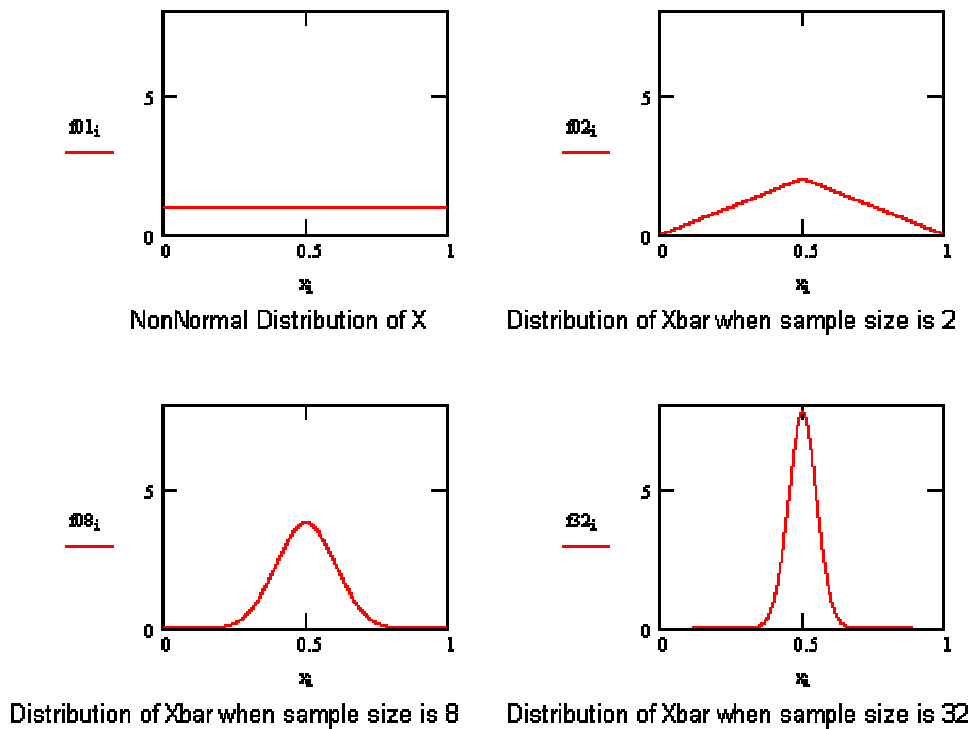

Figure 1: The distribution of the sum of uniform distributions, with 1, 2,8, and 32

In this study, we will take a look at the history of the central limit theorem, from its first simple forms through its evolution into its current format. We will discuss the early history of the theorem when probability theory was not yet considered part of rigorous mathematics. We will then follow the evolution of the theorem as more mathematicians play their part in its history to the final solution given by Lévy, Feller and Cramér in the 1930s. In this study, we will not supply complete proofs at any stage, but concentrate on the new things that each mathematician contributed.

2 Laplace's central limit theorem

The history of the central theorem starts with Laplace at the end of the 18th century. Sums of independent random variables had been studied for many different error distributions before 1810 when Laplace release his first paper about the CLT. However, they had mostly led to very complicated formulas. In Laplace's probabilistic work sums of random variables played an important role from the beginning. Already in one of his first papers of 1776, he was working on calculating the probability distribution of the sum of meteor inclination angles. He there faced the problem of the deviations between the arithmetic mean of the data (which were inflicted with observational errors) and the theoretical value. He assumed that all these angles were randomly distributed following a rectangular distribution between 0° and 90°. Due to the considerable amount of celestial bodies, he was not able to perform an exact calculation and he thus needed an approximation [Fis]. It was in the process of finding an approximation to this problem that Laplace eventually came to form the first versions of the CLT.

2.1 The introduction of characteristic functions

One important tool in Laplace's way of deducting the central limit theorem was the characteristic functions that Laplace introduced 1785. In this work, he started by regarding sums of errors. The error distribution he started his work with was a simple one; the error could take the values –1 and +1 with equal probabilities. The

sum of 2n errors could then get all the values –2n, -2n+2, ..., -2, 0, 2, ..., 2n-2, 2n with probabilities corresponding to the coefficients of the binomial $(1+1)^{2n}$. In his invention of the technique of the characteristic functions, he started with this simple binomial. Let us follow his work in this:

Laplce started with the binomial $(1+1)^{2n}$, let us denote the middle term of it by y_n. It is then clear that y_n will be equal to the term independent of e^{it} in the development of the binomial $(e^{it} + e^{-it})^{2n}$. Furthermore if we multiply this by dt and take the integral from t = 0 to t = π this integral will be equal to πy_n. Hence, we have:

$$y_n = \frac{1}{\pi} \int_0^\pi (e^{it} + e^{-it})^{2n} dt$$

Using the fact that $(e^{it} + e^{-it}) = 2\cos t$ we get an even simpler formula:

$$y_n = \frac{2^{2n}}{\pi} \int_0^\pi \cos^{2n} t \, dt$$

The beauty of it, as Laplace noted, lies in the fact that this can be extended further on, to find the middle term of the trinomial $(1+1+1)^{2n}$, the quadrinomial $(1+1+1+1)^{2n}$ etc. Each of these nomials now corresponds to having the corresponding amount of error terms (like $-1,0,1$ for the trinomial). If we generalize the calculations and allow the error to get values (–m, -m+1,..., -1, 0, 1,..., m-1, m) we get a corresponding m-nomial. By substituting the m-nomial 1:s with e^{it} : s we get $(e^{imt} + e^{i(m-1)t} + ... + 1 + ... + e^{i(-m+1)t} + e^{-imt})^{2n}$. The parenthesis can again be simplified by co sinus, and we end up with $\left(1 + 2\sum_{k=1}^{m} \cos tk\right)$. The

equal probability of each term is $(2m+1)^{-1}$ so we have to multiply the result by $(2m+1)^{-1}$.

The characteristic functions is thus defined in the following way:

$$\psi(t) = E(e^{itx})$$

For the first case we've studied (errors $-1,1$) we then had:

$$\psi(t) = \frac{1}{2}(e^{it} + e^{-it})$$

In the general case we had:

$$\psi(t) = \frac{1}{2m+1}(1 + 2\sum_{k=1}^{m} \cos tk)$$

We can now use the characteristic function to calculate the probabilities for $s_n = 0$ (s_n being the sum of the n possible errors) with the following formula:

$$P(s_n = 0) = \frac{1}{\pi}\int_0^\pi \psi^n(t)dt = \frac{1}{\pi(2m+1)^n}\int_0^\pi (1 + 2\sum_{k=1}^{m} \cos tk)^n dt$$

Laplace arrived to this formula his work and gave the following approximation to it:

$$P(s_n = 0) \approx \frac{\sqrt{3}}{\sqrt{2\pi nm(m+1)}}$$

2.2 The first versions of the CLT

The earlier result is the only result (regarding the CLT) that Laplace achieved in 1785. Considering how close he was to obtaining a general result it is quite amazing that he did not do it. Laplace calculated an approximation for $P(s_n = 0)$ but did not do it for any other values of s_n. Expanding the result to cover all values for s_n must surely have been within his grasp. Laplace did, however, not explore any other possible values for s_n than $s_n = 0$. Why he did not do it then, remains unknown. No one else took up his work either until Laplace himself returned to it in 1810, when he released and "proved" a generalization of his central limit theorem. In this paper of 1820, Laplace starts by proving the central limit theorem for some certain probability distributions. He then continues with arbitrary discrete and continuous distributions. We will here only look at his results for an arbitrary discrete distribution. Laplace's proof is the following [Fis]:

Assume that we have a discrete random variable x that gets the values $-m, -m+1, ..., m-1, m$ with the corresponding probabilities $p_{-m}, p_{-m+1}, ..., p_{m-1}, p_m$

Considering that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itx} e^{isx} dx = \delta_{t,s}$$

we get:

$$P(s_n = j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ijt} \psi^n(t) dt \qquad (1)$$

In this case the characteristic function is:

$$\psi(t) = \sum_{k=-m}^{m} p_k e^{ikt}$$

By inserting the characteristic function into (1) and by expanding the result for e^{ikt} we obtain:

$$P(s_n = j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ijt} \left(1 + it \sum_{k=-m}^{m} p_k k - \frac{t^2}{2} \sum_{k=-m}^{m} p_k k^2 + ... \right)^n dt \qquad (2)$$

Let's now take a closer look at the above parenthesis, particularly the natural logarithm of it:

$$\ln \psi^n = n \ln \left(1 + it \sum_{k=-m}^{m} p_k k - \frac{t^2}{2} \sum_{k=-m}^{m} p_k k^2 + ... \right)$$

$$= n \left(it\mu_x - \frac{t^2}{2} \sum_{k=-m}^{m} p_k k^2 + \frac{1}{2} t^2 \mu_x^2 + ... \right) \qquad (3)$$

Here μ_x is the expected value of x. We have used the Taylor series of the natural logarithm:

$$\ln(1 + x) = x - \frac{1}{2} x^2 + ...$$

The variance for x can be defined as $\sigma_x^2 = \sum_{k=-m}^{m} p_k k^2 - \mu_x^2$

Using that and (3) we can simplify the whole expression (2):

$$P(s_n = j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-ijt + itn\mu_x - \frac{1}{2} n\sigma_x^2 t^2 + ...)dt$$

By calculating $P(s_n - n\mu_x = s)$ instead we get:

$$P(s_n - n\mu_x = s) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(-ist - \frac{1}{2} n\sigma_x^2 t^2 + ...)dt$$

$$= \frac{1}{2\pi} \exp\left(-\frac{s^2}{2n\sigma_x^2} \right) \int_{-\pi}^{\pi} \exp\left[-\frac{1}{2} n\sigma_x^2 \left(t + \frac{is}{n\sigma_x^2} \right)^2 \right] dt$$

$$= \frac{1}{2\pi\sqrt{n\pi}} \exp\left(-\frac{s^2}{2n\sigma_x^2}\right) \int_{-\pi\sqrt{n}\sigma_x}^{\pi\sqrt{n}\sigma_x} \exp\left[-\frac{1}{2}\left(z + \frac{is}{\sqrt{n}\sigma_x}\right)^2\right] dt$$

Assuming that s is at the most of the order of $\sqrt{n}$ we get:

$$P(s_n - n\mu_x = s) = \frac{1}{\sqrt{2\pi n}\sigma_x} \exp\left(-\frac{s^2}{2n\sigma_x^2}\right)$$

This shows that $s_n - n\mu_x$ asymptotically approaches the normal distribution $N(0, n\sigma_x^2)$. Laplace also remarks that since m is part of the formula only through μ_x and σ_x, the result should be valid for a discrete distribution with infinite range, if μ_x and σ_x exist for the distribution.

The proof that Laplace provided for a continuous distribution follows the above proof for a discrete distribution quite closely. Laplace just replaces the continuous variable y with a discrete variable x such that $y = \frac{x}{a + a'}$. By regarding x, a, and a' as infinitely large integers, the integrals with y could be treated as sums. Laplace then follows the above proof, with some small modifications.

3 Poisson's contributions

Of all the contributions to the central limit theorem in the 19th century, the contributions by Siméon Denis Poisson would be the most influential on the contributions of later authors. Poisson published two articles (1824 and 1829) where he discussed the CLT. Poisson's idea was that all procedures in the physical world are governed by distinct mathematical laws. In this spirit, he tried to provide a more exact mathematical analysis to Laplace's theorem. Poisson's contributions to the CLT were twofold [Fis]:

i)   He provided a more rigorous proof for a continuous variable (thus sowing the seeds for the concept of random variables)

ii)   He discussed the validity of the central limit theorem, mainly by providing a few counterexamples.

In his paper, Poisson starts by providing a proof of the CLT for identically distributed variables. First for a sum of these and then for a linear combination of them He then generalizes his proof to the sum of random variables with different distributions. We will here take a look at the general results in the last part of his paper:

Let each random variable Y_i take values in the interval [a,b] with the continuous density $f_i(y) = F_i'(y)$, where $F_i(y) = P(Y_i \leq y)$. Let x_i, $\alpha \leq x_i \leq \beta$, ($\alpha\delta$ = a and $\beta\delta$ = b) be a discrete random variable, let δ be a small interval and set

$$p_{x_i} = f(x_i\delta)\delta \quad , \quad x_i = \alpha, \alpha+1,...,\beta$$

Writing the characteristic function for x_i we get:

$$\psi_{x_i}(t) = \sum_{x_i=\partial}^{\beta} p_x e^{itx_i} = \sum_{x_i\delta=\alpha\delta}^{\beta\delta} f_i(x_i\delta)e^{i(t/\delta)x_i\delta}\delta \qquad (4)$$

Setting $t = \theta\delta$ the right side of (4) tends to

$$\psi_{y_i}(\theta) = \int_a^b f(y)e^{i\theta y}dy$$

When we then want to find the probability distribution for the sum $S_n = Y_1 + Y_2 +...+ Y_n$, we get the following:

$$P\left(\delta\sum x_i = \delta s\right) = \frac{1}{2\pi}\int_a^b e^{-its}\prod_{i=1}^{n}\left(\psi_{x_i}(t)\right)dt$$

From this Poisson continued onwards and finally reached the following result:

$$P(c - \varepsilon \le S_n \le c + \varepsilon) = \frac{1}{\pi} \int_{-\infty}^{\infty} \left(\prod_{i=1}^{n} \int_a^b f_i(y) e^{i\theta y} dy \right) e^{-ic\theta} \frac{\sin(\varepsilon\theta)}{\theta} d\theta \qquad (5)$$

Poisson was unable to provide a rigorous proof for this general formula, but he did examine the validity of it in the special case of $n = 1$.

When $n = 1$ (5) writes to

$$P(c - \varepsilon \le Y_1 \le c + \varepsilon) = \frac{1}{\pi} \int_a^b \int_{-\infty}^{\infty} e^{i\theta(y-c)} \frac{\sin(\varepsilon\theta)}{\theta} d\theta f_1(y) dy \qquad (6)$$

With the help of the well known formula:

$$\frac{2}{\pi} \int_0^{\infty} \frac{\sin(kx)}{x} dx = \begin{cases} 1, & k > 0 \\ -1, & k < 0 \end{cases}$$

Poisson showed that

$$\int_{-\infty}^{\infty} e^{i\theta(y-c)} \frac{\sin(\varepsilon\theta)}{\theta} d\theta = \begin{cases} \pi, & y \in]c - \varepsilon, c + \varepsilon[\\ 0, & y \notin]c - \varepsilon, c + \varepsilon[\end{cases} \qquad (7)$$

With the help of (7) we then get the final result:

$$P(c - \varepsilon \le Y_1 \le c + \varepsilon) = \int_{c-\varepsilon}^{c+\varepsilon} f_1(y) dy$$

This concludes the proof.

Poisson's version of the CLT can be summarised in the following way: Let $Y_1,...,Y_n$ be independent random variables with density functions vanishing beyond the fixed interval [a,b]. If for the absolute values ρ_j of the characteristic functions of Y_j ($\psi_j(\theta) = \rho_j e^{i\varphi_j}$), there exists a function $r(\alpha)$ independent of j $0 \leq r(\alpha) \leq 1$ $\forall \alpha \neq 0$ such that $\rho_j \leq r(\alpha)$, then for arbitrary γ, γ':

$$P\left(\gamma \leq \frac{\sum(Y_i - EY_i)}{\sqrt{2\sum Var(Y_i)}} \leq \gamma'\right) \approx \frac{1}{\sqrt{\pi}} \int_\gamma^{\gamma'} e^{-u^2} du$$

The approximation gets better the larger n is and the difference between the left and right side becomes "infinitely small" with infinite n as Poisson explained [Fis].

As for Laplace, the main purpose of the CLT for Poisson was to be a tool in classical probability calculations, not so much to be a mathematical theorem in itself. Poisson did therefore not explicitly formulate any conditions for the central limit theorem to hold. It seems clear from his proofs and examples that he assumed the variances of the components of the sum to be bounded so that the variance of the sum would be of the order n. He does not say this explicitly though. He did, however, discuss a few counterexamples where the CLT does not hold.

One example of where the CLT does not hold are so called Cauchy-distributed variables where the probability density takes the following form:

$$f(x) = \frac{1}{\pi(1 + x^2)}$$

By inserting this into (5) we get

$$P(c - \varepsilon \le S_n \le c + \varepsilon) = \frac{1}{\pi} \arctan\left(\frac{2\varepsilon n}{n^2 + c^2 - \varepsilon^2}\right)$$

Hence, an approximate normal distribution cannot be obtained even with a large n.

4 Further development in the 19th century

Towards the end of the 19th century, mathematics was starting to change. The abstraction of mathematics was constantly taking place and the mathematical discussion was turning increasingly from computational mathematics to a more fundamental analysis, to "pure" mathematics. This had a big impact on probability theory as it had been considered more as "common sense" than a rigorous mathematical theory. Probability theory had been used where needed, but there had not been any need for rigorous proofs, as is the case with Laplace's first version of the CLT. For Laplace it had been enough that his theorem worked in practice and he didn't feel a need to turn it into mathematics since probability theory wasn't considered as such. This was about to change though. During this time, there were several attempts to give more "rigorous" proofs to the central limit theorem, the most important ones being by Bessel, Dirichlet, Cauchy, Ellis [Fis]. We will here look at a few of these attempts.

4.1 Dirichlet and Bessel

Dirichlet and Bessel mostly followed the footsteps of Laplace and Poisson in their proofs, but they introduced the so called "discontinuity factor" in their proofs. With the help of the discontinuity factor, they were able to prove Poisson's equation (5) for an arbitrary value of n, which Poisson himself had not been able to do.

Dirichlet's discontinuity factor states that:

$$\frac{2}{\pi}\int_0^\infty \frac{\sin(kx)}{x}\cos x\,dx = \begin{cases} 0, & |k| > 1 \\ 1, & |k| < 1 \end{cases}$$

With the use of this, both Dirichlet and Bessel we're able to prove Poisson's formula (5) for an arbitrary n. They followed different paths to achieve the proofs, but the principle remains the same. Bessel also pointed out that Poisson did in fact use the same discontinuity factor (although in a slightly different form) when he proved the formula (5) for n = 1.

Dirichlet also tried to estimate the error of the approximation. This attempt was not a very successful one and his technique differed heavily from modern techniques. It was, however, the first attempt to try to estimate the error of the approximation. Dirichlet had estimated the errors for other non-probabilistic approximations before and he showed that those same techniques could be applied to probability theory as well as they had been to "pure" mathematics.

4.2 Cauchy's version of the theorem

Cauchy was one of the first mathematicians to seriously consider probability theory as "pure" mathematics. He contributed in several different fields of mathematics and came up with a new way of proving the CLT. Cauchy's proof follows a different line compared to the previous proofs. He first found an upper bound to the difference between the exact value and the approximation and then specified conditions for this bound to tend to zero.

Cauchy gives his proof for independent identically distributed variables $y_1 \ldots y_n$ with a symmetric density f(y), finite support [-a, a], variance $\sigma^2 > 0$ and a characteristic function $\psi(\theta)$. He then considers the distribution of a weighted mean

$$z_n = \sum_{i=1}^{n} w_i y_i \qquad \sum_{i=1}^{n} w_i = 1$$

By following the lines of Poisson's proofs he gets:

$$P = P(-h < z_n < h) = \frac{2}{\pi} \int_0^\infty \psi(w_1\theta)...\psi(w_n\theta) \frac{\sin(h\theta)}{\theta} d\theta$$

If we assume that z_n is asymptotically normal with zero mean and the variance $\beta^2 = \sigma^2 \sum_{i=1}^{n} w_i^2$ (as it is according to the central limit theorem) then the corresponding probability can be written as

$$\Phi = \frac{2}{\pi} \int_0^\infty e^{-\beta^2\theta^2/2} \frac{\sin(h\theta)}{\theta} d\theta$$

Cauchy then tried to find an upper bound for $|P - \Phi|$ and specify conditions for this bound to tend to zero. In his proof, he had to divide the interval into three parts:

$$P - \Phi = (P - P_k) + (P_k - \Phi_k) + (\Phi_k - \Phi)$$

and study these parts separately. He thus got:

$$|\Phi_k - \Phi| = \left| \frac{2}{\pi} \int_k^\infty e^{-\beta^2\theta^2/2} \frac{\sin(h\theta)}{\theta} d\theta \right| < \frac{2}{\pi} \frac{1}{\beta^2 k^2} e^{-\beta^2 k^2/2}$$

$$|P_k - P| = \left| \frac{2}{\pi} \int_k^\infty e^{-\beta^2\theta^2\tilde{\gamma}/2} \frac{\sin(h\theta)}{\theta} d\theta \right| < \frac{2}{\pi} \frac{1}{\beta^2 k^2 \gamma} e^{-\beta^2 k^2 \gamma/2}$$

where $\gamma = \dfrac{1}{1 + \sigma^2 k^2 (\max(w_j))^2}$

For the last part, Cauchy stated that:

13

$$\left| P_k - \Phi_k \right| = \left| \frac{2}{\pi} \int\limits_0^k \left(e^{-\beta^2\theta^2\tilde{\gamma}/2} - e^{-\beta^2\theta^2/2} \right) \frac{\sin(h\theta)}{\theta} d\theta \right| < \frac{2\sqrt{3\alpha}}{\pi} \ln\left(\frac{kh}{\sqrt{3}} + \sqrt{1 + \frac{k^2h^2}{3}} \right)$$

where $\alpha = \max\left(e^{\frac{1}{8}\beta^2 k^4 a^2 (\max(w_j))^2}, 1 - e^{-\frac{\sigma^2\beta^2 k^4 (\max(w_1))^2}{4+2\sigma^2 k^2 (\max(w_j))^3}} \right)$

He finally concluded that if k is chosen so that $n^{1/2} < k < n^{3/4}$ then the three parts tend to zero for $n \to \infty$ and thus $P \to \Phi$ [Hal].

5 The second chapter in the history of the central limit theorem

Cauchy's proof finished what is called the first period of the central limit theorem (1810-1853). The proofs presented in this period were unsatisfactory in three respects [Hal]:

1. The theorem was not proved for distributions with infinite support.

2. There were no explicit conditions, in terms of the moments, under which the theorem would hold.

3. The rate of convergence for the theorem was not studied.

These problems were eventually solved by Russian mathematicians, between 1870 and 1910. Three probabilistic mathematicians are normally credited for this, namely Chebyshev, Markov and Liapounov; the so-called "St. Petersburg School". In the literature, there is normally no mention of others although it seems clear that there were other Russian mathematicians involved in working with the same problems, e.g. P.A. Nekrosov and I.V. Sleshinsky. There seems, however, to have been quite a lot of controversy between the "St. Petersburg School" and the others [Sen]. We will here focus on the achievements of the "St. Petersburg School" since these mathematicians are generally considered as having contributed the most to the central limit theorem.

5.1 Chebyshev's and Markov's proof of the central limit theorem

Chebyshev's paper in 1887 is generally considered the beginning of rigorous proofs for the central limit theorem. In his paper, Chebyshev states the following [Sen]:

Let z_1, z_2, z_3, ..., be independent random variables each described by probability densities. If

    i) $E(z_i) = 0 \quad \forall i$

    ii) $\left| E(z_i^k) \right| \leq C \quad \forall i, \ k \geq 2$

where C is a constant independent of i and k; then as $n \to \infty$

$$P(t < \frac{S_n}{B_n} < t') \to \frac{1}{\sqrt{2\pi}} \int_t^{t'} e^{-\frac{1}{2}x^2} dx$$

where $S_n = \sum_{i=1}^{n} z_i$ and $B_n^2 = \sum_{i=1}^{n} Var(z_i)$

In his proof, which is incomplete, Chebyshev used the "method of moments", which was developed earlier by him. It was later simplified and completed by Markov, who also completed Chebyshev's proof of the CLT.

In 1898, after Chebyshev's proof, Markov stated that: "a further condition needs to be added in order to make the theorem correct". He first proposed the following condition:

    iii) B_n^2 / n is uniformly bounded away from 0

Later Markov replaced the condition with:

iiia) $E(z_n^2)$ is bounded from 0 as $n \rightarrow \infty$

After Liapounov's proof of the CLT with characteristic functions in 1901, Markov worked hard to achieve the same level of generality with the method of moments. He finally succeeded with this in 1913 when he presented a paper that provided a rigorous proof of the CLT under Liapounov's condition by using the method of moments.

5.2 Liapounov's theorem

Liapounov was, as Markov was, a student of Chebyshev and thus a part of the "St Petersburg School". Liapounov wanted to introduce rigorous proofs to probability theory and he was successful in that. In his proof of the CLT, he did not follow Chebyshev and Markov in their use of the method of moments but he followed Laplace's idea and used characteristic functions instead. Liaupounov's proof, published in 1901, is considered the first "real" rigorous proof of the CLT. Let us take a look at the theorem as stated by Liapounov [Usp]:

Let x_1, x_2, x_3,..., x_n be independent random variables with the following properties:

$E(x_i) = 0 \ \forall i$

$E|x_i|^k \leq \infty \ \forall i, \ k \geq 2$

If there exists a $\delta > 0$ such that

$$\frac{\sum_{i=1}^{n} E|x_i|^{2+\delta}}{B_n^{2+\delta}} \xrightarrow{n\to\infty} 0$$

then

$$P\left(\frac{S_n}{B_n} < t\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}u^2} du$$

After this part, Liapounov's proof follows Laplace, by using characteristic functions. In his proof, however, he uses a fundamental lemma that is the key to the simplicity and rigorousness of his proof:

Let s_n be a variable depending on an integer n, with the mean 0 and variance 1. If its characteristic function

$$\psi_n(t) = E(e^{is_n t})$$

converges to $e^{-\frac{t^2}{2}}$ (the characteristic function of the normal distribution) uniformly in any given finite interval (-k,k) then the distribution function $F_n(t)$ of s_n tends uniformly to $\Phi(t)$ for all t.

Liapounov did not explicitly separate this fundamental lemma in his proof, but it is implicitly contained therein. Several other important contributors to the CLT, like Lindeberg and Lévy, used this lemma in their improvements of the CLT [Usp].

6 The last chapter in the history of the central limit theorem

The so-called 3rd and last chapter in the history of the central limit theorem 1920-1937, begins with Lindeberg's proof of it and Lévy's and Feller's proof that added necessary conditions to the theorem.

6.1 The Lindeberg condition

In 1922, Lindeberg published an elementary proof of the CLT. Lindeberg's proof is very simple and applies to Euclidean valued and even Hilbert valued random vectors as well as random variables. Lindeberg's proof was also the basis of both Feller's and Lévy's work on the same problem. Let's have a look at Lindeberg's theorem [Cam]:

Let the variables x_i be independent random variables with the expectations zero and variances σ_i^2. Let s_n be the standard deviation of the sum S, $s_n^2 = \sum \sigma_i^2$. If

$$\frac{1}{s_n^2} \sum_i E\left\{ x_i^2 I_{\left[x_i/s_n > \varepsilon \right]} \right\} \to 0$$

then

$$\frac{S}{s_n} \to N(0,1)$$

The above condition is generally called the Lindeberg condition [Gne]. In his proof, Lindeberg derived a limit for $\sup_x |F(x) - \Phi(x)|$ (where $\Phi(x)$ is the normal probability distribution) in terms of the third moment [Cam, GnK]. The same proof

can be carried out with characteristic functions as well, as it often is presented today. Lindeberg's method of proof did remain rather unused throughout the first half of the 20th century and it was not re-used before Trotter's paper in 1959. Trotter's paper did not make Lindeberg's method very popular either, but when limit theorems started to be created for infinite-dimensional spaces, the strength of Lindeberg's method was finally recognised. Lindeberg's method was easily transferable to infinite-dimensional spaces whereas the usage of characteristic functions was not. Today Lindeberg's condition is used in most cases where convergence to a normal distribution is considered with non-identically distributed variables [Pau].

The strength of Lindeberg's method lies mainly in two points:

1. It can be applied in a very general context

2. It takes the rate of convergence in the limit theorem under consideration.

It is appropriate to note that Lindeberg's method did not give the optimal order of the rate of convergence, even in the case of i.i.d. real-valued summands with finite third moment it gives the order n [Pau].

With Lindeberg's proof, there was a rigorous proof for the CLT that provided the *sufficient* conditions for the CLT. There were, however, no proofs for the *necessary* conditions of the CLT. As Poisson had shown already in 1824, the approximation to a normal distribution did not always hold for arbitrary independent variables. This lack was partly remedied by Lévy and Feller in 1935 and 1937.

6.2 Feller's theorem

Feller's paper of 1935 gives the necessary and sufficient conditions for the CLT, but the result is somewhat restricted. Feller considered an infinite sequence x_i of independent random variables. He then gave conditions for a_n and c_n such that

$$\frac{1}{a_n}\left(S_n - c_n\right)$$

where $S_n = \sum_{i=1}^{n} x_i$ tends to N(0,1).

Feller, however, restricted his treatment to the case where the a_n, if they exist, must be such that for each k the variables x_k/a_n tend to zero in probability. Feller's conditions for a_n and c_n are valid under this restriction, so Feller did in a sense give the final solutions to the CLT, but since he posed this restriction and because he only treated normed sums, it can not be considered the final solution to the CLT. Feller's theorem is often called the Lindeberg-Feller theorem as it uses the Lindeberg condition [Rao].

6.3 Lévy's theorem

Lévy proved Lindeberg's condition in 1925, using characteristic functions. He did, however, consider Lindeberg's proof to be simpler and superior to his own [Cam]. He published several papers related to the central limit theorem between 1925 and 1930, mostly using characteristic functions in his proofs. After 1930 he, however, avoided using characteristic functions, and in his 1935 paper, he does not use characteristic functions at all. Lévy's 1935 paper was presented only a few months after Feller's and despite the fact that the papers treated the same problem, both Feller and Lévy deny having had any previous contact. In his paper from 1935, Lévy proved several things related to the central limit theorem [Cam]:

i)   He gave necessary and sufficient conditions for the convergence of normed sums of independent and identically distributed random variables to a normal distribution

ii) Lévy also gave the sufficient and necessary conditions for the general case of independent summands

iii)   He also tried to give the necessary and sufficient conditions for dependent variables, martingales.

There were, however, some problems with Lévy's proofs. Lévy's proof of the necessary conditions for the martingale case was not quite satisfactory and did not stand a test of rigorousness. His proofs of the necessary and sufficient conditions for the general case of independent variables were correct, but his proofs relied on a hypothetical lemma, that had not yet been proved (as did Fellers proof). The lemma is the following:

"If the sum $S = X + Y$ of two independent random variables (X and Y) has a normal distribution then so do X and Y."

Lévy's whole proof relies on this lemma and was thus not quite satisfactory at the time being presented. The following year (1936), Cramér proved the lemma (as a theorem) and the matter was settled. With help of the same theorem the use of normed sums could be shown to be valid and both Lévy's and Feller's theorems where thus generally applicable. Both Feller and Lévy returned and refined their proofs in 1937 after they could use Cramérs result. The CLT was thus proved with both necessary and sufficient conditions [Cam] and the final chapter in the history of the CLT could be closed.

7 Conclusions

In this study, we have closely followed the history of the CLT. We started by discussing Laplace and his need for approximating the distribution of the sums of errors. We followed Laplace's reasoning quite closely, in how he first discovered the CLT. We then followed how various contributors contributed to the CLT, how

Poisson gave examples of distributions that cannot be approximated by the CLT and how he provided a more rigorous proof of the continuous case. We then saw how Dirichlet and Bessel eventually proved Poisson's theorem and we saw Cauchy define the first upper bound for the difference between the distribution of the sum and the normal distribution. After this first chapter, the Russian mathematicians provided the first rigorous proofs of the CLT. First Chebyshev and Markov with the method of moments. After their proof, Liapounov proposed the Liapounov condition and used characteristic functions again in his proof. After the Russians, it was Lindeberg who gave had the final word regarding the sufficient conditions by proposing the Lindeberg condition. The case was not closed, however, as Lindeberg had not provided necessary conditions for the CLT. In the end, Feller and Lévy provided the necessary conditions for the CLT that were eventually rigorously proved by Cramér. We have thus arrived more or less, to where the CLT stands today. The Lindeberg condition has been improved somewhat [Cram] and the CLT has been given sufficient and necessary conditions for various dependent variables as well, but the basic principles of Lindeberg, Lévy and Feller still remains the ones used to date.

8 References

[Cam] Le Cam, L. (1986), The Central Limit Theorem around 1935, Statistical Science 1(1): 78-96

[Fis] Fischer, Hans   The Central Limit Theorem from Laplace to Cauchy: Changes in Stochastic Objectives and in Analytical Methods, Translation from German to English of the first 2 chapters from the book: Die verschiedenen Formen und Funktionen des zentralen Grenzwertsatzes in der Entwicklung von der klassischen zur modernen Wahrscheinlichkeitsrechnung, 2000, found at http://mathsrv.ku-eichstaett.de/MGF/homes/didmath/seite/fischer.html

[Gne] Gnedenko, B.V. (1962). The Theory of Probability. Mir Publishers, Moscow

[GnK] Gnedenko, B.V. and Kolmogorov, A.N. (1954). Limit Distributions for Sums of Independent Random Variables. Addison-Wesley Publishing Company, Inc.

[Hal] Hald, Anders (1998). A History of Mathematical Statistics from 1750 to1930. John Wiley & Sons, Inc.

[Pau] Paulauskas, V (1999). J.W. Lindeberg and the Central Limit Theorem, from Statistics, registries and science: experiences from Finland (ed. Juha Alho). Statistics Finland.

[Rao] Rao, C. Radhakrishna (1965). Linear Statistical Interference and Its Applications. John Wiley & Sons, Inc.

[Sen] Seneta, E. (1984). The central limit problem and linear least squares in prerevolutionary Russia: The background, Mathematical Scientist 9: 37-77

[Usp] Uspensky, J.V. (1937). Introduction to Mathematical Probability. McGraw-Hill, New York.