THE STORY OF THE CENTRAL LIMIT THEOREM

Loh Wei Yin

The central limit theorem (CLT) occupies a place of honour in the theory of probability, due to its age, its invaluable contribution to the theory of probability and its applications. Like all other limit theorems, it essentially says that all large-scale random phenomena in their collective action produce strict regularity. The limit law in the CLT is the well-known Normal distribution from which is derived many of the techniques in statistics, particularly the so-called "large sample theory".

Because the CLT is so very basic, it has attracted the attention of numerous workers. The earliest work on the subject is perhaps the theorem of Bernoulli (1713) which is really a special case of the Law of Large Numbers. De Moivre (1730) and Laplace (1812) later proved the first version of the CLT. This was generalized by Poisson to constitute the last of the main achievements before the time of Chebyshev.

The theorems mentioned above deal with a sequence of independent events $\xi_1, \xi_2, \xi_3, \ldots$, with their respective probabilities denoted by $p_n = P(\xi_n)$. The number of actually occurring events among the first n events $\xi_1, \ldots, \xi_n$ is denoted by the random variable Z_n. The above-mentioned results can now be stated as follows. (The first two theorems have $p_n = p$ for all n, and $0 < p < 1$.)

1. Bernoulli's Theorem. For every $\varepsilon > 0$,

$$P\left(\left|\frac{Z_n}{n} - p\right| > \varepsilon\right) \rightarrow 0 \qquad \text{as } n \rightarrow \infty$$

2. Laplace's Theorem

$$P(z_1 < \frac{Z_n - np}{\sqrt{np(1-p)}} < z_2) \to \Phi(z_2) - \Phi(z_1)$$

as $n \to \infty$ uniformly with respect to z_1 and z_2.

We have used the notation

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

which is the standard Normal distribution function.

3. CLT in Poisson's Form

Let $A_n = p_1 + \ldots + p_n$, $B_n^2 = p_1(1-p_1) + \ldots + p_n(1-p_n)$.

Then

$$P(z_1 < \frac{Z_n - A_n}{B_n} < z_2) \to \Phi(z_2) - \Phi(z_1)$$

as $n \to \infty$ uniformly with respect to z_1 and z_2.

If we introduce the indicator random variable

$$I_\xi = \begin{cases} 1 & \text{if } \xi \text{ occurs} \\ 0 & \text{if } \xi \text{ does not occur,} \end{cases}$$

Z_n can be written as

$$Z_n = I_{\xi_1} + I_{\xi_2} + \ldots + I_{\xi_n}.$$

Thus the above three theorems are in fact special cases of limit theorems concerning sums of independent random variables.

The rigorous proof of the more general CLT for sums of arbitrarily distributed independent random variables was made possible by the creation in the second half of the nineteenth century of powerful methods due to Chebyshev, whose work signalled the dawn of a new development in the entire theory of probability.

Chebyshev considered a sequence of independent random variables $X_1, X_2, \ldots, X_n, \ldots$ with finite means and variances, denoted respectively by $a_n = EX_n$, $b_n^2 = E(X_n - a_n)^2$. Let $S_n = X_1 + \ldots + X_n$, $A_n = a_1 + \ldots + a_n$, and $B_n^2 = b_1^2 + \ldots + b_n^2$. Chebyshev studied and solved the following problem.

Problem. What additional conditions ensure the validity of the CLT:

$$P\left(\frac{S_n - A_n}{B_n} < z\right) \to \Phi(z)$$

for every real z as $n \to \infty$?

To solve this problem, Chebyshev created the method of moments. His proof, in a paper in 1890, was based on a lemma which was proved only later by Markov (1899). Soon afterwards, Lyapunov (1900, 1901) solved the same problem under considerably more general conditions using another method, although Markov later showed that the method of moments is also capable of obtaining Lyapunov's theorem. However, it turned out that Lyapunov's method was simpler and more powerful in its application to the whole class of limit theorems concerning sums of independent variables. This is the method of characteristic functions using Fourier analytic techniques. It is so powerful that to date no other method can yield better results for the case of independent random variables.

The condition Lyapunov used to solve Chebyshev's problem was

$$\lim_{n\to\infty} C_n/B_n^{2+\delta} = 0 \ ,$$

where $C_n = c_1 + \ldots + c_n$, $c_k = E|X_k - a_k|^{2+\delta}$ for some $\delta>0$.

An even weaker condition is the famous Lindeberg condition that for every $\delta>0$.

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x|\geq\varepsilon B_n} x^2 dF_k(x+a_k) = 0,$$

where F_k is the distribution of X_k. Subsequently Feller (1937) showed that the Lindeberg condition is not only sufficient but also necessary for the limit law to be normal, provided an appropriate uniform asymptotic negligibility of the X_i/B_n is assumed.

In practical applications the CLT is used essentially as an approximate formula for sufficiently large values of n. In order that this use is justified, the formula must contain an estimate of the error involved. One way of doing this is to consider the various asymptotic expansions for the distribution

$$F_n(x) = P(\frac{S_n - A_n}{B_n} < x).$$

In his 1890 paper Chebyshev indicated without proof the following expansion for the difference $F_n(x) - \Phi(x)$, when the random variables are identically distributed:

$$F_n(x) - \Phi(x) \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \left(\frac{Q_1(x)}{n^{\frac{1}{2}}} + \frac{Q_2(x)}{n} + \ldots + \frac{Q_j(x)}{n^{j/2}} + \ldots \right)$$

where the $Q_j(x)$ are polynomials. The most definitive results in this direction are due to Cramer. Edgeworth (1965) studied in detail the expansion in a slightly different form.

When the random variables are identically distributed and possess finite third moments, Berry (1941) and Esseen (1945) independently proved the celebrated result

$$|F_n(x) - \Phi(x)| \leq \frac{K}{\sqrt{n}} \cdot \frac{\beta}{\sigma^{3/2}},$$

where $\beta = E|X_1 - EX_1|^3$, $\sigma^2 = EX_1^2 - (EX_1)^2$ and K is a constant. Later results have generalized this to the case of non-identically distributed summands with the best bound achieved by Esseen (1969) in terms of truncated third moments.

A natural question generated by Lyapunov's CLT is whether the condition that the random variables be independent can be generalized. It was forty-seven years later before Hoeffding and Robbins (1948) proved a CLT for an m-dependent sequence of random variables. (The concept of m-dependence essentially requires that given the sequence $X_1, X_2, \ldots, X_n, \ldots$, it is m-dependent if $(X_1, X_2, \ldots, X_r)$ is independent of $(X_s, X_{s+1}, \ldots)$ for $s-r>m$. In this terminology an independent sequence is 0-dependent.) Later Diananda (1955) and Orey (1958) improved on this result by assuming only Lindeberg's condition and the boundedness of the sum of the individual variances.

Almost at the same time, Rosenblatt (1956) proved a CLT for a "strong mixing" sequence. This condition requires only that the dependence between X_n and X_{n+k} diminishes as k increases. Thus m-dependence is included as a special case. Rosenblatt's results were subsequently improved by Philipp (1969a, 1969b) who not only relaxed the former's conditions but also obtained bounds for the error in the normal approximation. Soon after, in the Sixth Berkeley Symposium. Dvoretzky (1972) presented very general results for dependent random variables. For the particular case of strong mixing, he went beyond Philipp's (1969a) theorem by dropping the condition that the variables be uniformly bounded. In this connection, the author and Chen [22] have added a refinement to one of Dvoretzky's theorems. Recently too, McLeish (1974) has made improvements on Dvoretzky's paper.

At the same symposium in which Dvoretzky presented his results, an equally interesting paper was given by Stein (1972). This paper is concerned with bounding the error in the normal approximation for dependent random variable. Its significance lies not so much in its improvement of known results, which it did manage handsomely, but rather in its introduction of a new method vastly different from the established Fourier techniques. The method, which makes no use of characteristic functions, essentially depends upon an identity and a perturbation technique.

The interest created by Stein's paper was almost immediate. Chen (1972) used it to give an elementary proof of the CLT for independent random variables while Erickson (1974) obtained an L_1 bound for the error for m-dependent sums. The latter has since been generalized by the author and Chen [21] to ϕ-mixing sequences. Chen [9] has meanwhile employed a variation of Stein's method to obtain necessary and sufficient conditions for the dependent central limit problem where the limit law need not be normal. In the case that the limit is normal, the author and Chen [22] have improved on the existing results for strong mixing sequences. Although Stein's method appears to be more easily applicable to dependent random variables, the classical Fourier method is still superior for independent variables. This is because to date no one has been able to apply Stein's method to yield the classical Berry-Esseen theorem.

In yet another direction of generalization, Markov was among the first to prove a multidimensional CLT, where the sequence of random variables is now a sequence of independent random vectors. The limit law then becomes the multidimensional Gaussian distribution. Apart from the extra work of dealing with matrices, the proof of the multidimensional CLT appears to be a simple extension of the one-dimensional case.

The corresponding problem of bounding the error in the multi-dimensional CLT is more interesting. Among the first to look for estimates was Rao (1961). He was closely followed by a host of others, mainly Russians, like Bikjalis (1966), von Bahr (1967), Bhattacharya (1968), Sadikova (1968) Sazanov (1968), Bergstrom (1969), Paulanskas (1970) and Rotar (1970). With the exception of the last two, all the authors mentioned above considered only independent and identically distributed random vectors. The last two dropped the assumption of identical distributions. When third moments exist, an order of $n^{-\frac{1}{2}}$ is obtained, which is equivalent to the Berry-Esseen rate. However, this is only possible for the class of convex Borel sets. In fact, Bikjalis (1966) has shown that for arbitrary Borel sets, additional conditions had to be assumed.

This is therefore the present situation regarding developments in the study of the CLT. There are still many nagging questions left to be asnwered, particularly in bounding the error in the normal approximation. By considering coin-tossing, it is seen that the rate given in the Berry-Esseen theorem is achieved and hence further work on this may only be found in reducing the absolute constant. A more challenging problem is to obtain a proper generalization to dependent variables. So far, all estimates, with the exception of that of Stein (1972), do not reduce to the Berry-Esseen rate. Stein (1972) obtained the correct order of $n^{-\frac{1}{2}}$ for a sequence of stationary n-dependent random variables with eighth moments. The others manage at best an order of $n^{-\frac{1}{4}}$ (see e.g. Philipp (1969b), Erickson (1974), Loh and Chen [21]) for more general types of dependence. Another problem awaiting future research is to get bounds for the corresponding multidimensional case for dependent random vectors. There does not appear to have been any work on this problem yet.

Work on the CLT has generated much interest in related problems like the Poisson approximation and the Central Limit Problem. With the latter are associated some of the great pioneers in probability like Levy, Khitchine and more recently, Kolmogorov. To retrace their work would require another essay as long as the present.

It is perhaps justified to add that no other topic in the theory of probability has attracted so much attention for so long as the CLT. For two hundred and fifty years since its birth, the CLT has held man's fascination and will continue to do so for many years to come.

References

[1] von Bahr, B. (1967). On the central limit theorem in R^k. *Ark. Mat.* 7, 5, 61-69.

[2] Bergstrom, H. (1969). On the central limit theorem in R^k. The remainder term for special Borel sets. *Z. Wahrscheinlichkeitstheorie* 14, 113-126.

[3] Bernoulli, J. (1713). *Conjestandi*. Basle.

[4] Berry, A.C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* 49, 122-136.

[5] Bhattacharya, R.N. (1968). Berry-Esseen bounds for the multi-dimensional central limit theorem. *Bull. Amer. Math. Soc.* 74, 285-287.

[6] Bikjalis, A. (1966). The remainder terms in multi-dimensional limit theorems. *Soviet Math. Dokl.* 7, 705-707.

[7] Chebyshev, P.L. (1890). Sur deux theorems relatifs awe probabilites. *Acta Math.* 14, 305-315.

[8] Chen, Louis H.Y. (1970). An elementary proof of the central limit theorem. *Bull. Singapore Math. Soc.* 1-12.

[9] Chen, Louis H.Y. A new approach to dependent central limit problems. To appear.

[10] Diananda, P.H. (1955). The central limit theorem for m-dependent variables. *Proc. Camb. Phil. Soc.* 51, 92-95.

[11] Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist.* Prob. 2, 513-535.

[12] Edgeworth, F. Y. (1905). The law of error. *Proc. Camb. Phil. Soc.* 20, 36-65.

[13] Erickson, R.V. (1974). L_1 bounds for asymptotic normality of m-dependent sums using Stein's technique *Ann. Probability.* 2, 532-529.

[14] Esseen, C.G. (1945). A mathematical study of the Laplace-Gaussian law. *Acta Math.* 77, 1-125.

[15] Esseen, C.G. (1969). On the remainder term in the central limit theorem. *Ark. Mat.* 8, 7-15.

[16] Feller, W. (1937). Ueber den zentralen Grenzwerksatz der Wahrscheinlichkeitsrechnung. *Math. Z.* 42, 301-312.

[17] Gnedenko, B.V. and Kohmogorov, A,N. *Limit distributions for sums of independent random variables.* Addison-Wesley. 1954.

[18] Hoeffding, W. and Robbins, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.* 15, 773-780.

[19] Laplace, R. (1812). *Theorie analytique des probabilites.* Paris.

[20] Lindeberg, Y.W. (1922). Eine nene Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* 15, 211-225.

[21] Loh, Wei-Yin and Chen, Louis H.Y. The L_1 normal approximation for ϕ-mixing random variables. To appear.

[22] Loh, Wei-Yin and Chen, Louis H.Y. On the asymtotic normality for sums of strong mixing random variables. To appear.

[23] Lyapunov, A.M. (1900). Sur une proposition de la theorie des probabilites. *Bull. de l'Acad. Imp. des Sci. de Mt. Petersbourg.* 13, 359-386.

[24] Lyapunov, A.M. (1901). Nouvelle forme du theoreme sur la limite de probabilites. *Mem. Acad. Sc. St. Petersbourg* 12, 1-24.

[25] Markov, A.A.(1899). The law of large numbers and the method of least squares. *Izv. Fiz. - Mat. Obshchestva Kazan. Univ.* 8, 110-128.

[26] McLeish, D.L. (1974). Dependent Central Limit Theorems and Invariance Principles. *Ann Probability* 2, 620-628.

[27] de Moivre, A. (1730). *Miscellanea Analytica Supplementum.* London.

[28] Orey, S.A. (1958). Central limit theorem for m-dependent random variables. *Duke Math. J.* 25, 543-546.

[29] Paulauskas, V. (1970). The multi-dimensional central limit theorem. *Litovsk. Mat. Sb.* 10, 783-789.

[30] Philipp, W. (1969a). The central limit problem for mixing sequences of random variables. Z. *Wahrscheinlichkeitstheorie*. 12, 155-171.

[31] Philipp, W. (1969b). The remainder in the central limit theorem for mixing stochastic processes. *Ann. Math. Statist.* 40, 601-609.

[32] Rao, R. Ranga (1961). On the central limit theorem in R_k. *Bull. Amer. Math. Soc.* 67, 359-361.

[33] Rosenblatt, M. (1956). A central limit theorem and a mixing condition. *Proc. nat. Acad. Sci. USA*. 42, 412-413.

[34] Rotav, V.I. (1970). The rate of convergence in the multi-dimensional central limit theorem. *Theor. Prob. Appl.* 15, 354-356.

[35]. Sadikova, S.M. (1968). On the multi-dimensional central limit theorem. *Theor. Prob. Appl.* 13, 164-170.

[36] Sazanov, V.V. (1968). On the multi-dimensional central limit theorem. *Sankhya*, Ser. A, 30 pg.2.

[37] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* 2, 583-602.