

STUDY ON THE SPEED OF CONVERGENCE OF THE CENTRAL LIMIT THEOREM BY A SIMULATION PROGRAM

JUNG-CHUNG TSAY

Collage of Engineering, National Chiao Tung University

(Received October 17, 1972)

Abstract—Although sufficient mathematical tools are available to prove the correctness of the central limit theorem¹, it is only proved in an infinite sense, while the speed of convergence of a sum (S_n) of n independent identically distributed random variables to a normal distributed random variable is not easy to find theoretically². This paper describes a simulation program which may be used to simulate the distribution of S_n for any given n . Then the speed of convergence of the central limit theorem can be studied.

I. INTRODUCTION

For $n=1, 2, \dots$ let X_n be identically distributed as the random variable X , with finite mean $E[X]$ and standard deviation $\sigma[X]$. Let the sequence $\{X_n\}$ be independent, and let Z_n be defined by

$$Z_n = \frac{S_n - E[S_n]}{\sigma[S_n]}, \text{ where } S_n = X_1 + X_2 + \dots + X_n$$

and has a characteristic function $\phi_{z_n}(u)$. Then the central limit theorem states¹

$$\lim_{n \rightarrow \infty} \phi_{z_n}(u) = e^{-\frac{u^2}{2}} \text{ or } \lim_{n \rightarrow \infty} P[Z_n \leq z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{y^2}{2}} dy = \lim_{n \rightarrow \infty} \int_{-\infty}^z f_{z_n}(z) dz$$

In other word, the distribution of the sum of a number of independent identically distributed random variables with finite means and variance, normalized to have mean zero and variance 1, converge to $N(0, 1)$. Although this theorem has been proven, the speed of convergence is not wellknown. This speed of convergence will be studied by a simulation program. To simulate the behavior of the random number X , a pseudo-random number generator RND (1) that generates numbers almost uniformly in the range $[0, 1]$ is used. This uniform distributed numbers may be modified to the actual distribution of X . To give a more flexible simulation program, the modification is done in a subroutine, so that different types of distribution may be simulated on this simulation program by changing the subroutine. Many of the parameters; for examples, n , $E[X]$, $\sigma[X]$ may also be easily changed. The simulation program is written in BASIC language which can be run on the WANG 3300 computer or other computers which have a BASIC compiler.

II. THE SIMULATION PROGRAM

By theoretical point of view, the distribution of Z_n can be known from the repeat application of the formula for the probability density function of $Y=R_1+R_2$ (where R_1 and R_2 are independent random variables)¹:

$$f_{R_1+R_2}(y) = \int_{-\infty}^{\infty} dr f_{R_1}(r) f_{R_2}(y-r)$$

However, the integration on the right-hand side of the above formula is not easy to find even for a simple uniform distribution. This difficulty can be seen from the example given below. Let X_1, X_2, X_3, X_4 be independent random variables, uniformly distributed on the unit interval. Then from the above given formula, the probability density function of Z_2 and Z_4 are⁵:

$$f_{z_2}(z) = \begin{cases} (1 - |z/\sqrt{6}|) / \sqrt{6}, & |z| \leq \sqrt{6} \\ 0, & \text{otherwise} \end{cases}$$

$$f_{z_4}(z) = \begin{cases} (2/3 - 4|z/(2\sqrt{3})|^2 + 4|z/(2\sqrt{3})|^3) / \sqrt{3}, & |z| \leq \sqrt{3} \\ (2 - |z\sqrt{3}|^3) / (6\sqrt{3}) & \sqrt{3} \leq |z| \leq 2\sqrt{3} \\ 0 & \text{otherwise} \end{cases}$$

These consideration illustrates that the formula for $f_{z_n}(z)$ is not easy to evaluate theoretically. Especially, when n becomes large. This fact makes the simulation

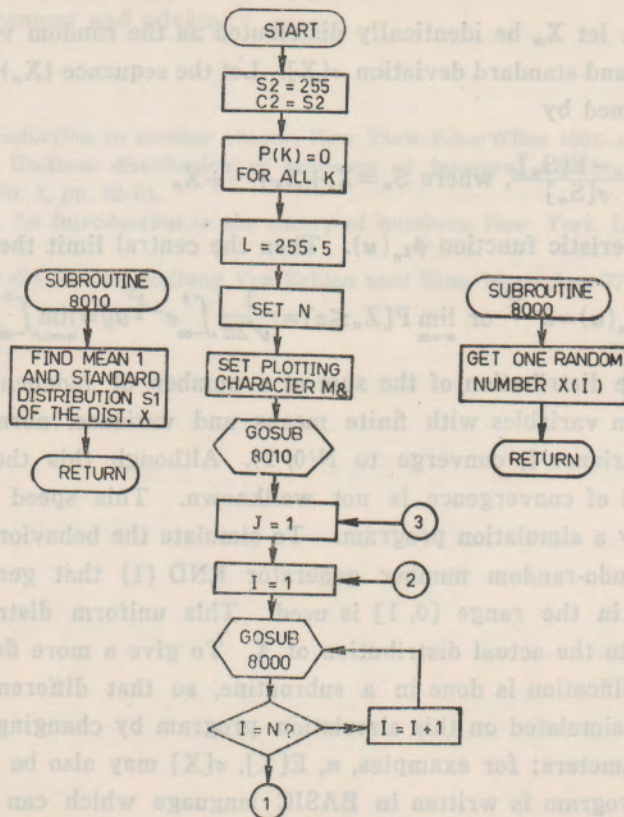


Fig. 1

program described latter more attractive than a theoretical approach. To compare with the distribution given by the simulation program as described latter, the actual probability density function for $f_{z_1}(z)$, $f_{z_2}(z)$, $f_{z_4}(z)$ and $N(0, 1)$ or $f_{z_{\infty}}(z)$ are plotted in Fig. 1, where Z_n is normalized $S_n = X_1 + X_2 + \dots + X_n$ and X_i 's are independent random variables, uniform distributed on the unit interval $[0, 1]$.

The operating principle of the simulation program is better understood by the flowchart of the simulation program as shown in Fig. 2. Various symbols used in the flowchart are:

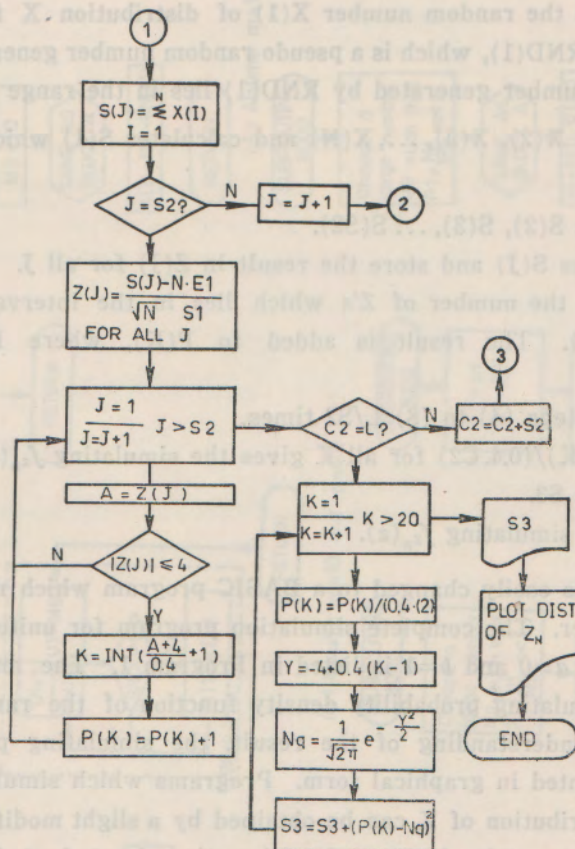


Fig. 2. Flowchart of the Simulation Program

- S2, a constant
- C2, index for controlling a loop in step of S2.
- P(K), simulating probability density function of Z_n .
- L, number of samples of S_n generated.
- N, number of samples taken from X to form S_n .
- M\$, a character used for plotting the simulating $f_{z_n}(z)$.
- E1, mean value of X.
- S1, standard deviation of X.
- J, subscript of S_n , range from 1 to S2.
- I, subscript of X, range from 1 to N.

- X(I), the random variable X_i which has the distribution of X.
- S(J), the random variable which is the sum of N random X's.
- S3, sum of the square of the difference between N (0, 1) and the simulating $f_{z_n}(z)$ at some discrete points.

Main functions of the flowchart are described in the following steps:

- (1) S2=225, C2=S2, L=255*5.
- (2) Sets N and M\$.
- (3) Finds the mean and standard deviation of the random number X.
- (4) Generates the random number X(1) of distribution X from the built in function RND(1), which is a pseudo random number generator. The pseudo random number generated by RND(1) lies in the range [0.1].
- (5) Generates X(2), X(3), ..., X(N) and calculates S(1) which is equal to $\sum_{i=1}^N X(i)$.
- (6) Generates S(2), S(3), ... S(S2).
- (7) Normalizes S(J) and store the result in Z(J) for all J.
- (8) Computes the number of Z's which lies in the interval $[-4+0.4(K-1), -4+0.4K)$. The result is added to P(K), where K varies from 1 to 20.
- (9) Repeats steps (4) to (8) L/S2 times.
- (10) $P(K)=P(K)/(0.4.C2)$ for all K gives the simulating $f_{z_n}(z)$.
- (11) Computes S3.
- (12) Plots the simulating $f_{z_n}(z)$.

The flowchart is easily changed to a BASIC program which may be run on the Wang 3300 computer. The complete simulation program for uniform distribution of X with parameters $a=0$ and $b=1$ is listed in Program 1. The main output of the program is the simulating probability density function of the random variable Z_n . To given a quick understanding of the result, the simulating probability density function is represented in graphical form. Programs which simulate other types of distribution of distribution of X can be obtained by a slight modification of Program 1. This modification can be down at the subroutines located at line numbers 8000 and 8010 of Program 1. The subroutine 8010 return the values S1 and E1, while the subroutine 8000 return a random number X(I) which has the given distribution of X. Subroutines that corresponding to some frequently encountered probability laws are shown in Fig. 3, these probability laws and their characteristic values are listed in Table 1.^{1,5}

The simulating result of the probability density function $f_{z_n}(z)$ for some of the probability laws listed in Table 1 is shown on Fig. 4. As can be seen from Fig. 4, for various types of distribution of X, there is a tendency for the simulating result approaching the normal distribution N(0, 1). This fact, of course, come from the result of central limit theorem. The value of the variable S3 for the exponential and absolute distribution is listed in Table 2.

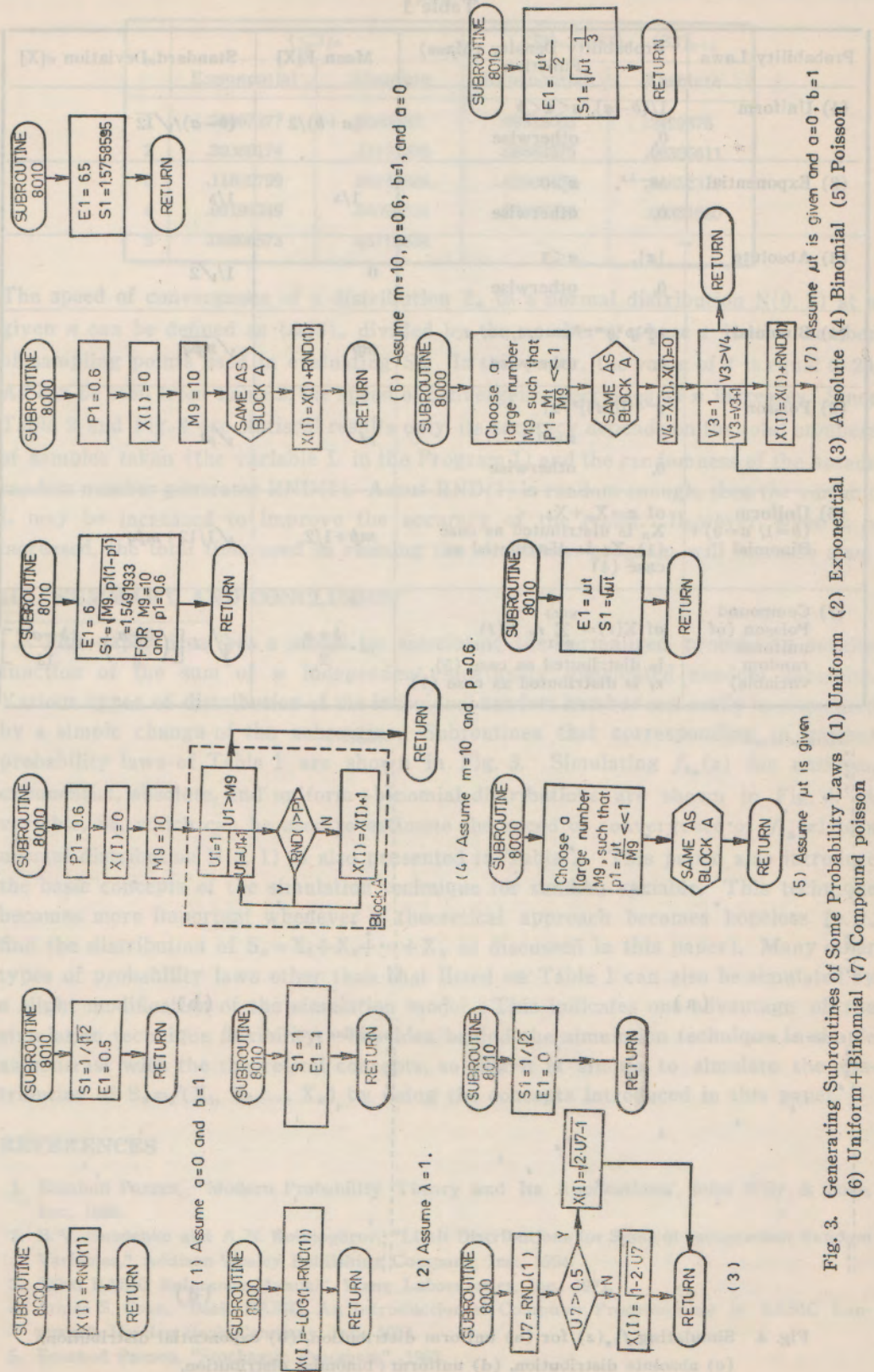


Fig. 3. Generating Subroutines of Some Probability Laws (1) Uniform (2) Exponential (3) Absolute Binomial (4) Uniform + Binomial (5) Compound Poisson (6) Uniform (7) Poisson

Table 1

| Probability Laws | Probability Density (Mass) Function | Mean E[X] | Standard Deviation $\sigma[X]$ |
|---|--|-----------------------------|---|
| (1) Uniform | $1/(b-a), a < x < b$ 0, otherwise | $(a+b)/2$ | $(b-a)/\sqrt{12}$ |
| (2) Exponential | $\lambda e^{-\lambda x}, x > 0$ 0, otherwise | $1/\lambda$ | $1/\lambda$ |
| (3) Absolute | $ x , x \leq 1$ 0, otherwise | 0 | $1/\sqrt{2}$ |
| (4) Binomial | $\binom{n}{x} p^x q^{n-x}, x=0, 1, \dots, n$ 0, otherwise | np | \sqrt{npq} |
| (5) Poisson | $e^{-\mu t} [(\mu t)^x / x!]$ $x=0, 1, \dots$ 0, otherwise | μt | $\sqrt{\mu t}$ |
| (6) Uniform (of $b=1, a=0$) + Binomial | of $x=X_a+X_b$ X_a is distributed as case (1), X_b is distributed as case (4) | $mp+1/2$ | $\sqrt{1/12+mpq}$ |
| (7) Compound Poisson (of uniform random variable) | of $X(t) = \sum_{n=1}^{N(t)} r_n N(t)$ is distributed as case (5), r_i is distributed as case (1) | $\mu t \cdot \frac{b+a}{2}$ | $\sqrt{\mu t \cdot \left[\left(\frac{b+a}{2} \right)^2 + \frac{(b-a)^2}{12} \right]}$ |

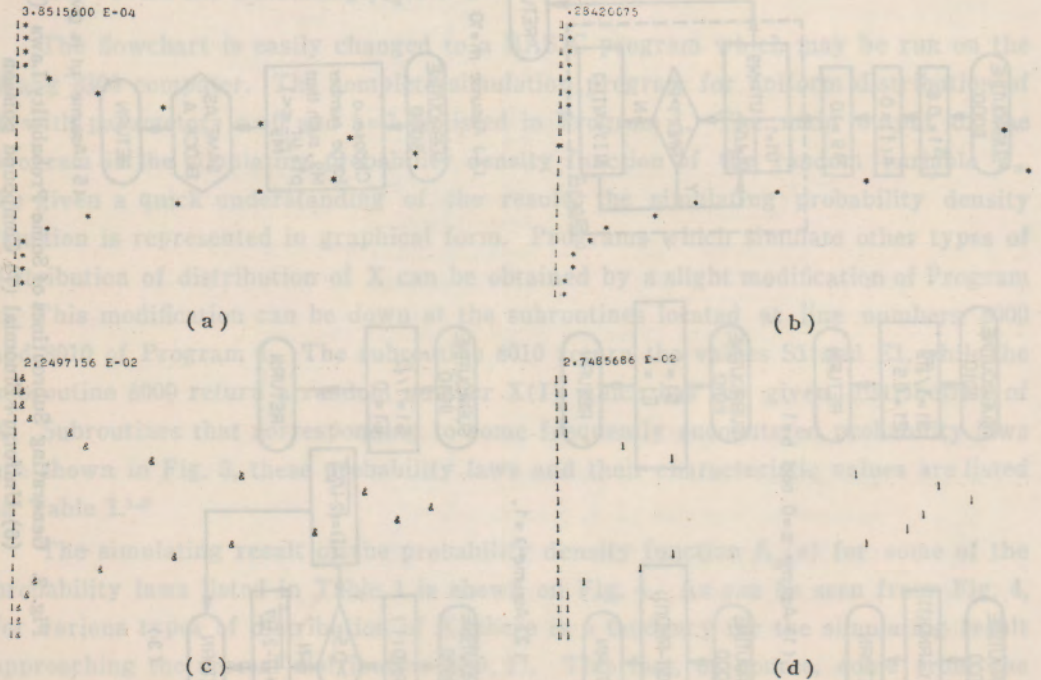


Fig. 4. Simulating $f_{z_n}(z)$ for (a) uniform distribution, (b) exponential distribution, (c) absolute distribution, (d) uniform + binomial distribution,

Table 2

| n | $(\leq 3)_n$ | | $(\Delta S3)_n = (S3)_n - (S3)_{n+1}$ | |
|---|--------------|-----------|---------------------------------------|-----------|
| | Exponential | Absolute | Exponential | Absolute |
| 1 | .30007177 | .3580281 | .09618003 | .2499375 |
| 2 | .20389174 | .11103435 | .08583375 | .06356611 |
| 3 | .11805799 | .04746824 | .02606450 | .00652310 |
| 4 | .09199349 | .04094514 | .00592776 | .00381680 |
| 5 | .08606573 | .03712834 | — | — |

The speed of convergence of a distribution Z_n to a normal distribution $N(0, 1)$ at a given n can be defined as $(\Delta S3)_n$ divided by the number t , where t is the number of sampling points used in evaluating $S3$. In this paper, the value of t is equal to 20. As can be seen from Table 2, the speed of convergence decreases as n increases. Since Table 2 and Fig. 4 are statistical results only, its accuracy depends on the total numbers of samples taken (the variable L in the Program 1) and the randomness of the pseudo random number generator $RND(1)$. Assume $RND(1)$ is random enough, then the variable L may be increased to improve the accuracy of the result. However, when L is increased, the total time used in running the simulation program will increase also.

III. SUMMARY AND CONCLUSION

This paper describes a model for simulating the normalized probability density function of the sum of n independent identically distributed random variables. Various types of distribution of the individual random number can easily be simulated by a simple change of the subroutine. Subroutines that corresponding to various probability laws of Table 1 are shown in Fig. 3. Simulating $f_{z_n}(z)$ for uniform, exponential, absolute, and uniform+binomial distributions are shown in Fig. 4. A variable, $S3$, which can be used to estimate the speed of convergence of $f_{z_n}(z)$ to a normal distribution $N(0, 1)$ is also presented in Table 2. This paper also introduce the basic concepts of the simulation technique for random variable. This technique becomes more important whenever a theoretical approach becomes hopeless (e.g., find the distribution of $S_n = X_1 + X_2 + \dots + X_n$ as discussed in this paper). Many other types of probability laws other than that listed on Table 1 can also be simulated by a slight modification of the simulation model. This indicates one advantage of the simulation technique flexibility. The idea behind the simulation technique is simple as contrast with the theoretical concepts, so that it is simple to simulate the distribution of $S_n = f(X_1, X_2, \dots, X_n)$ by using the concepts introduced in this paper.

REFERENCES

1. Emanufi Parzen, "Modern Probability Theory and Its Applications", John Wiley & Sons, Inc., 1960.
2. B. V. Gnedenko and A. N. Kolmogorov, "Limit Distributions for Sums of Independent Random Variables," Addison-Wesley Publishing Company, Inc., 1954.
3. "3300 BASIC Reference Manual", Wang Laboratories, Inc., 1971.
4. James S. Coan, "Basic BASIC, An Introduction to Computer Programming in BASIC Language", Hayden Book Company, Inc., 1971.
5. Emanufi Parzen, "Stochastic Processes", 1967.