

The Proof and Illustration of the Central Limit Theorem by Brownian Numerical Experiments in Real Time within the Java Applet

Monika Gall, Ryszard Kutner, and Wojciech Wesela

Institute of Experimental Physics, Department of Physics, Warsaw University,
Smyczkova 5/7, PL-02678 Warsaw, Poland

Abstract. We developed the interactive Java applet¹ which made it possible to prove and illustrate by Brownian numerical experiments the Central Limit Theorem (CLT); the theorem which yet constitutes the basis for the Gaussian stochastic processes and the reference case for the non-Gaussian ones. Our approach emphasizes the contrast between theoretical complexity and simplicity provided by our probabilistic simulations. We argue that the present approach should be the preliminary stage for the advanced educational process before the analytical stage is developed. We stress that the Gaussian probability distribution function (PDF) is as stable as distinguished, e.g., from the delta-Dirac, Poisson and t-Student ones which were also considered for comparison. The latter distribution was chosen so as to have all the moments higher than the second order diverging so as to verify the validity of CLT. As our experiments were performed in real time we were able to visualize the convergence of the processes (versus the size of the statistical ensemble of Brownian experiments) both to the variance of the sum of independent identically distributed random variables (which linearly increases with the number of the latter) and to the Gaussian PDF (which is the asymptotic distribution of this sum independently of the single-step PDF used here). We hope that our experimental approach will inspire students to undertake their own studies, e.g., to consider the non-Gaussian processes where CLT is violated which is the modern trend in statistical physics and its applications.

1 Motivation

It is commonly known that the limit theorems constitute one of the main trends in the theory of stochastic processes and their applications. Among them the Central Limit Theorem (CLT) plays a crucial role in broad applications covering many fields ranging from statistical physics across stochastic simulations to signal processing [1].

¹ This applet is available under the internet address: http://primus.okwf.fuw.edu.pl/erka/DIDACT/CTG_BROWN/.

Roughly speaking, the CLT is one of the main statistical properties responsible for the main question, namely: what is the probability distribution function (PDF) of the sum of an unlimited number of (independent identically distributed) random variables? This problem was intensively and extensively studied in the last century by a huge number of researchers; these studies provided many handbooks, monographs, review papers and scientific works in mathematics and in natural, social and economical sciences. However, it is really surprising that up to now no Java applets were published as a modern educational tool directly verifying CLT and supporting, e.g., the distant learning of the above mentioned sciences.

In this paper we study the (unbiased) Brownian motion or random walk (hence, throughout this article we shall use the physical notation where random variables can be treated as displacements or increments of the Brownian particle). By using our Brownian numerical experiments we verify (with good approximation) the weakest version of CLT which should be advantageous both for the random walk theories and for various applications of CLT [2].

2 CLT and Possibilities of the Advanced Java Applet

There are many mathematical formulations of various versions (weaker or stronger) of the Central Limit Theorem. In the present work we consider the one, often used in natural, social and economical sciences.

General description. The user of the applet can choose a concrete distribution of single displacements of the walker from among the typical ones: (i) the delta-Dirac, (ii) the Gaussian, (iii) the Poisson and (iv) the Student's distribution offered by the option *Distribution of single displacements* (see e.g., Fig.1).

The Student's PDF was defined by three degrees of freedom so as to have all moments higher than the second order diverging (opposite to the first three distributions having all moments finite). The latter case is particularly interesting as the speed of convergence to the limit Gaussian distribution cannot be covered in this case neither by the well-known Chebyshev-Gnedenko-Kolmogorov expansion nor by the Berry-Esséen theorems 1 and 2 [3,4]. Thus the decisive role of the finite variance is emphasized and the validity of CLT is verified. Other PDFs (for example, so popular as the uniform, log-normal, and Gamma ones) can be also attached to the applet by the user himself as the Java is an object, platform independent programming and the source code of our applet is available (our applet is working under the Java 2 version 1.4.1.1 or higher).

The fundamental step of each experiment is to draw the length of a single, independent displacement of the walker from the chosen PDF common for all steps and all experiments. This is the initial assumption of the CLT. For example, in this work we present the results for Gauss and Poisson PDFs. Since in all our cases time is homogeneous, the most general version of the CLT, which assumes various distributions for different steps of the walker, cannot be considered here.

The asimuth of the displacement is always drawn from the uniform probability distribution function as we consider unbiased (isotropic) random walks (i.e. the unbiased version of the CLT). Hence, as the number of steps n is again introduced by the user (by applying the option *Number of particle steps*, see e.g., Fig.1), the characteristic zig-zag trajectory can be designed in each experiment together with the total displacement $\mathbf{R}(n)$, marked by the interval which connects the beginning (empty circle) and the end (full circle) of the trajectory.

In this way we illustrate the Brownian motion i.e. an incessant and irregular motion of the particle immersed in an ambient medium (which is invisible here), e.g. suspended in a fluid. Each single step of the Brownian particle (defined by consequent turning points) is characterized by the abruptness, hence the velocity of the particle is a quantity practically unmeasurable so a less definite description was used. The key point of the theory of Brownian motion was given by Smoluchowski who found that sufficiently large fluctuations, although relatively rare, are still common enough to explain the phenomenon [2].

Detailed description. The total displacement $\mathbf{R}(n)$ is composed of n single, elementary displacements (increments) \mathbf{r}_j , $j = 1 \dots, n$,

$$\mathbf{R}(n) = \sum_{j=1}^n \mathbf{r}_j; \quad (1)$$

of course, the CLT involves a fundamental constraint

$$\sigma^2 = \langle (\mathbf{r}_j)^2 \rangle_{pdf} = \langle \mathbf{r}^2 \rangle_{pdf} < \infty, \quad j = 1, 2, 3, \dots, \quad (2)$$

obeyed for any step j and for each distribution mentioned above (here $\langle \dots \rangle_{pdf}$ means the average over the single-step PDF).

The basic, global aim of the applet is to construct a statistical ensemble (series) of the mentioned above independent experiments within which the particle performs a two-dimensional random walk. For example, in the frame of the left window in all the enclosed figures the zig-zag trajectories of the walker are shown. Note that by using the zoom, the trajectory can be enlarged 2 or 4 times to allow observation of its structure (e.g., some elements of self-similarity in the stochastic sense). The number of such experiments as well as the number of particle steps common for all experiments is declared by the user at the beginning. Usually, the applet is able to conclude series of experiments (consisting of hundred thousands) sufficient for a good statistics within a few minutes while their visualization is performed (during the simulation) as often as it is set by the user (by using the option *Snapshot picture after each ... experiments*, cf. Fig.1). Moreover, the option *Delay* (again see Fig.1) makes it possible to execute a single experiment sufficiently slowly to observe, e.g., elementary displacements of the Brownian particle. By setting this statistical ensemble of experiments the required averages (just over the statistical ensemble) and required statistics can be easily constructed (cf. the final stages presented in Figs.1 - 3).

In fact, the applet calculates two time-dependent quantities: (i) the mean-square displacement (MSD) of the walker $\langle [\mathbf{R}(t)]^2 \rangle_{se}$ (here $\langle \dots \rangle_{se}$ is the L -

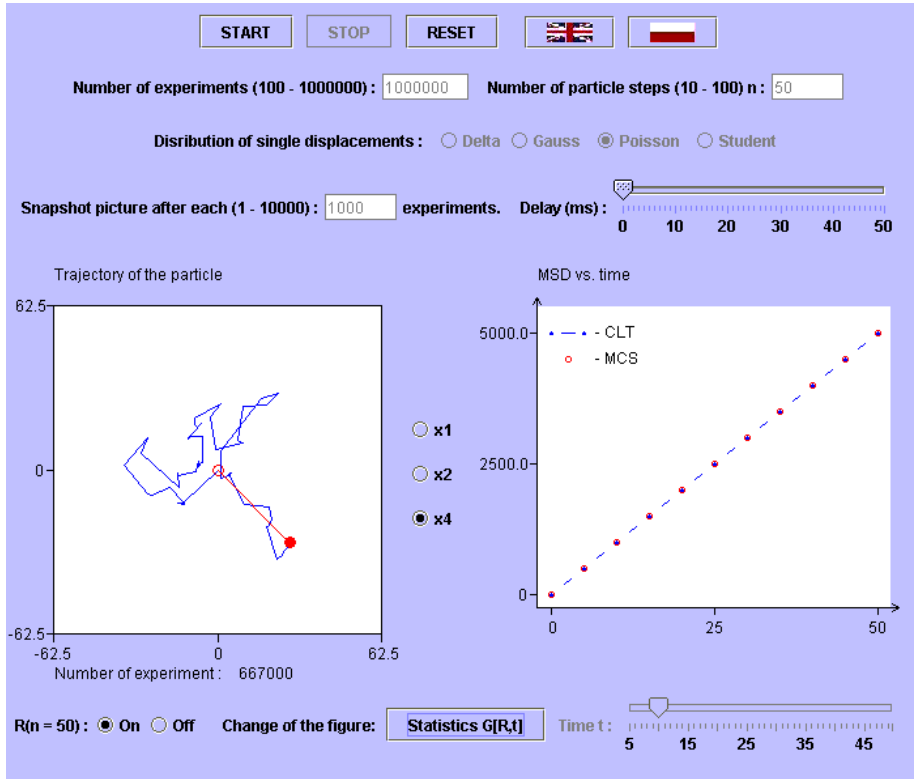


Fig. 1. Snapshot of the screen's picture of a single experiment of the Brownian motion together with the MSD vs. time of the Brownian particle almost for the final stage (after 667000 experiments).

dependent average over the statistical ensemble where L is the size of this ensemble, cf. Sec.3) and (ii) the statistics (in the form of the histogram) $G(\mathbf{R}; t)$ of the total displacement passed within time t^2 . These two quantities are the main ones considered by the CLT.

The right window of the applet is working in real time in two modes, which are selected by the user: the first, which visualizes the preparation of the MSD vs. time t and the second one, which analogously shows the statistics at time t which is (almost) smoothly fixed by using the applet's slider. In both cases the applet shows the movie, which is the more stable the greater is the size of the statistical ensemble (L); this is a reminiscence of the Law of Large Numbers (LLN) [5].

² Here time $t = \tau n$, where τ is the time needed by the walker to perform a single step, common for all steps (in each experiment we put $\tau = 1$); this defines, the so called 'discrete-time' random walks.

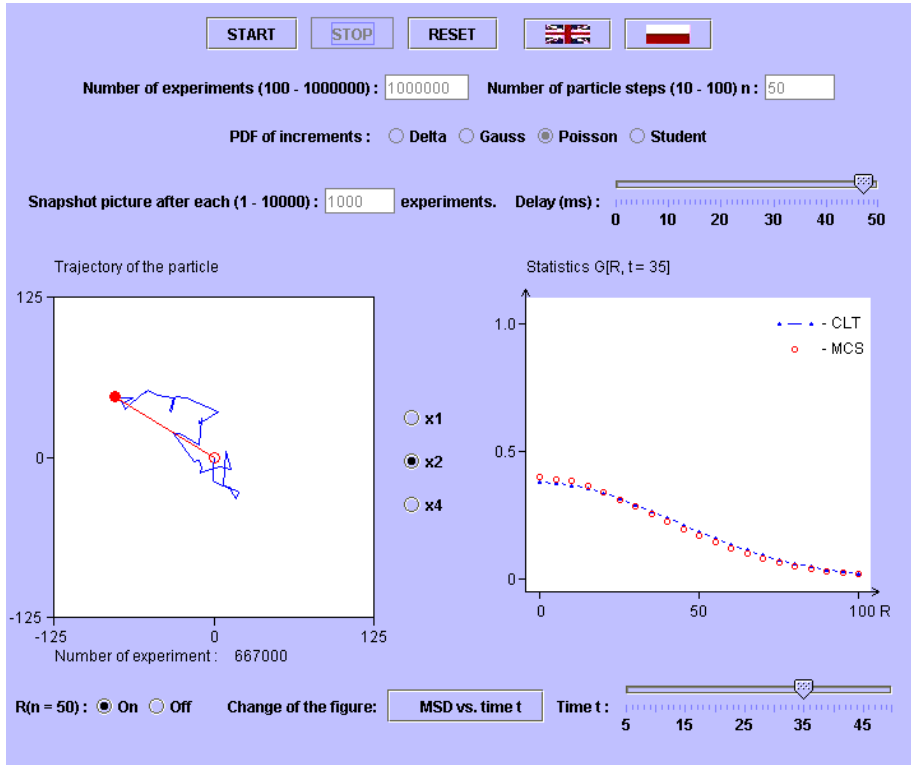


Fig. 2. Snapshot of screen's picture of a single experiment of the Brownian motion together with the statistics $G(R, t)$ of the Brownian particle displacements for long time ($t = 35$) almost the final stage (after 667000 experiments).

By using our applet we exploit conditions sufficient for the Central Limit Theorem to occur.

3 The Algorithm and Results

The applet constructs a statistical ensemble of independent (similar) experiments within which the walker performs two-dimensional random walks; all necessary details of these walks are recorded.

Preparation of the MSD. By using the statistical ensemble, the mean-square displacement is easily calculated

$$\langle [\mathbf{R}(m)]^2 \rangle_{se} = \frac{1}{L} \sum_{l=1}^L [\mathbf{R}^l(m)]^2, \quad m = 1, 2, 3, \dots, n, \quad (3)$$

where m is the current number of the walker steps common for all experiments, which increases up to its maximal value n set before the applet starts, and

analogously L is the current number of experiments which increases up to its maximal value set by the user at the beginning. The MSD is currently presented in the right window of the applet within the MSD mode. As the parameters L and m increase up to their maximal values we can observe convergence of the MSD to its stable limit which is just a reminiscence of the LLN.

Preparation of statistics. As the real space is isotropic we can use the following formula to simplify the relation which constitutes the basis for the histogram and to remarkably improve the statistics

$$\begin{aligned} \frac{L(R, R + \Delta R; t)}{L} &= \int_0^{2\pi} d\phi \int_R^{R+\Delta R} dR' R' \cdot PDF(R'; t) \\ &\approx 2\pi \Delta R \cdot \left(R + \frac{\Delta R}{2}\right) \cdot PDF(R_\xi; t), \end{aligned} \quad (4)$$

where quantity $L(R, R + \Delta R; t)$ is the number of experiments (from the statistical ensemble) in which we found the particle within the ring of inner radius R and thickness ΔR after t steps. Hence, $L(R, R + \Delta R; t)/L$ estimates the (related) probability of finding the walker within the ring of inner radius R and thickness ΔR at time t ; here $R = |\mathbf{R}|$ is the length of the position vector \mathbf{R} and the thickness $\Delta R \ll \sigma$. The radius R_ξ obeys inequalities $R \leq R_\xi \leq R + \Delta R$ and was chosen so as the experimental statistics fulfilled the normalization condition. It is seen from relation (4) that we calculate the searched one-sided probability distribution function, $PDF(R(\approx R_\xi); t)$, as a function of $R(\geq 0)$ for various times t ; this is visualized in the right window by using the applet slider given by option *Time t*.

Results. The final (in practice) situation after 667000 experiments (for the Poisson PDF of single-step increments) is shown in Figs.1 and 2. In these figures the final stage of convergence is observed; a good agreement between experimental results and theoretical predictions is well seen, which proves experimentally the Central Limit Theorem. The discrepancy can be observed only for the short time range ($t \approx 10$). Although the data shown in these figures concern the Poisson PDF for a single walker step, the results for other PDFs are very similar; they differ only by the speed of convergence to the asymptotic MSD and the Gaussian PDF. As for the Gaussian PDF of single-step increments the convergence is quickest (cf. Fig.3 where time $t = 5$ was sufficient) a comment should be made.

There is a characteristic difference between the Gaussian and other PDFs, namely, the Gaussian one is a stable distribution in the sense that PDF of $\mathbf{R}(n)$ is again a Gaussian probability distribution function for any $n \geq 1$. This effect can be observed by using the applet where the results (for the Gaussian PDF of single-step increments) can be shown, for example, for $n = 5, 10$ and 40 .

For any case, the final stage of statistics requires a much larger size of the statistical ensemble than that required by the time-dependent mean-square displacement; this can be easily observed within the applet where after about $L \approx 10000$ the MSD vs. time well agrees with the theoretical prediction but not the statistics which still exhibits too large scattering of the data points. However, the speed of convergence to the Gaussian PDF for increasing L is the greatest for the case of the Gaussian PDF of single-step increments.

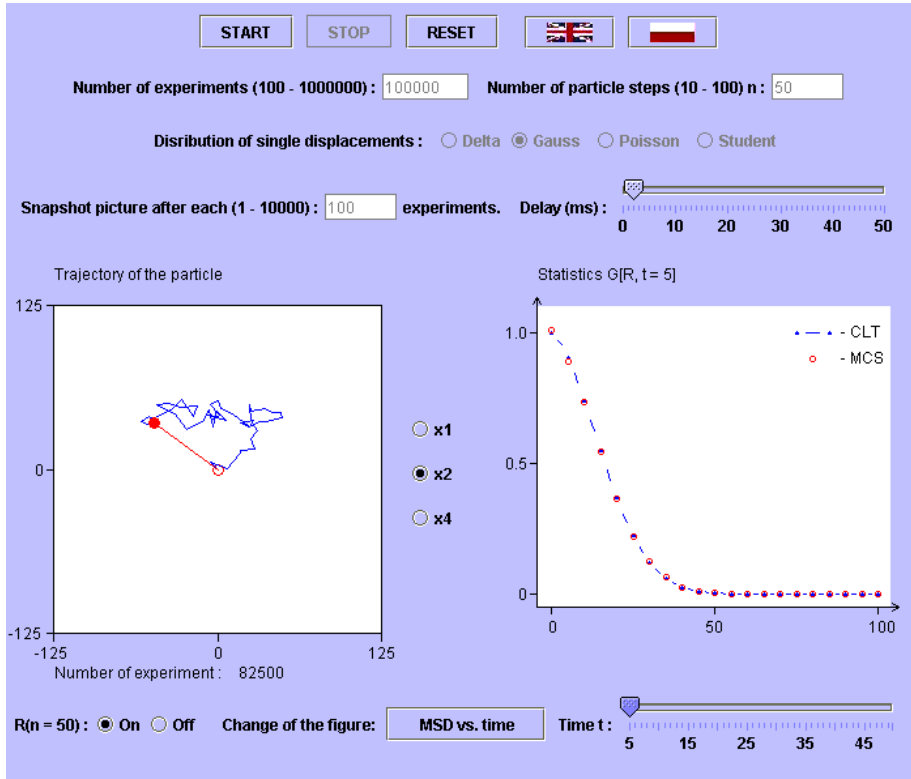


Fig. 3. Snapshot of the screen's picture of a single experiment of the Brownian motion together with the statistics $G(R, t = 5)$ of the Brownian particle displacements for the final stage (after 82500 experiments).

4 Concluding Remarks

It is remarkable that so rich numerical results were obtained by using a so simple probabilistic algorithm [6]. Our algorithm obeys two constraints fundamental for our considerations. The first one

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n \frac{\langle [\mathbf{r}_j]^2 \rangle_{pdf}}{j^2} = \zeta(2) \cdot \langle \mathbf{r}^2 \rangle_{pdf} < \infty, \quad (5)$$

(where $\zeta(\dots)$ is the Riemann ζ -function) is sufficient for the strong Kolmogorov Law of Large Numbers to occur as we took into account only the PDFs having a finite variance of single-step displacements; the variance is identical for all successive steps (as a consequence of time homogeneity).

The second constraint

$$\lim_{n \rightarrow \infty} \frac{\langle [\mathbf{R}(n)]^2 \rangle_{se}}{n^2} = \lim_{n \rightarrow \infty} \frac{\langle \mathbf{r}^2 \rangle_{pdf}}{n} = 0 \quad (6)$$

is sufficient for both the Central Limit Theorem and the Law of Large Numbers to occur, where the former quantitatively and the latter qualitatively were verified by our numerical experiments.

Our main result clearly illustrates the concept of the **basin of attraction** in the functional space of probability distribution functions. To describe this concept we focus our attention on delta-Dirac, Gaussian, Poisson and Student distributions which converge to the Gaussian PDF, which is the attractor of all these distributions. Their convergence is practically quite rapid as need only dozen single steps of the walker. Note that only the Gaussian PDF remains the Gaussian along the trajectory in the functional space, which illustrates the concept of the **stable distribution** (i.e., its shape is preserved all the time); in this sense other distributions considered by us are unstable. Though the Gaussian attractor is, maybe, the most important one in functional space of PDFs, other (stable and unstable) attractors also exist and play an increasing role in probabilistic theories and their applications [7,8,9].

References

1. Sornette D.: Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools. Springer-Verlag, Berlin 2000
2. Mazo R.M.: Brownian Motion. Fluctuations, Dynamics and Applications. Clarendon Press, Oxford 2002
3. Mantegna R.N., Stanley H.E.: An Introduction to Econophysics. Correlations and Complexity in Finance. Cambridge Univ. Press, Cambridge 2000
4. Bouchaud J.-P., Potters M.: Theory of Financial Risks. From Statistical Physics to Risk Management. Cambridge Univ. Press, Cambridge 2001
5. Feller W.: An introduction to probability theory and its applications. Vol.I. J. Wiley & Sons, New York 1961
6. Landau D.P., Binder K.: Monte Carlo Simulations in Statistical Physics. Cambridge Univ. Press, Cambridge 2000
7. Bouchaud J.-P., Georges A., Anomalous Diffusion in Disordered Media: Statistical Mechanisms, Models and Physical Applications. Phys. Rep., Vol.195 (1990) 127-293
8. Shlesinger M.F., Zaslavsky G.M., Frisch U., (Eds.): Lévy Flights and Related Topics in Physics. LNP, Vol.450. Springer-Verlag, Berlin 1995
9. Pękalski A., Kutner R., Sznajd-Weron K., (Eds.): Anomalous Diffusion. From Basics to Applications. LNP Vol.519. Springer-Verlag, Berlin 1999