Z. Wahrscheinlichkeitstheorie verw. Gebiete 32, 111–131 (1975)

An Approximation of Partial Sums of Independent RV'-s, and the Sample DF. I

J. Komlós, P. Major and G. Tusnády

Summary. Let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum of i.i.d.r.v.-s, $EX_1 = 0$, $EX_1^2 = 1$, and let $T_n = Y_1 + Y_2 + \cdots + Y_n$ be the sum of independent standard normal variables. Strassen proved in [14] that if X_1 has a finite fourth moment, then there are appropriate versions of S_n and T_n (which, of course, are far from being independent) such that $|S_n - T_n| = O(n^{\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log\log n)^{\frac{1}{4}})$ with probability one. A theorem of Bártfai [1] indicates that even if X_1 has a finite moment generating function, the best possible bound for any version of S_n, T_n is $O(\log n)$. In this paper we introduce a new construction for the pair S_n, T_n, and prove that if X_1 has a finite moment generating function, and satisfies condition i) or ii) of Theorem 1, then $|S_n - T_n| = O(\log n)$ with probability one for the constructed S_n, T_n. Our method will be applicable for the approximation of sample DF., too.

1. Introduction

Let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum of i.i.d.r.v.-s, with distribution function $F(x)$, such that $EX_1 = 0$, $EX_1^2 = 1$, and let $W(t)$ be a Wiener process. The Skorohod's embedding scheme (cf. [13]) provides a sequence of i.i.d.r.v.-s τ_j such that

$$P(S_n < x) = P\left(W\left(\sum_{j=1}^{n} \tau_j\right) < x\right), \qquad n = 1, 2, \ldots,$$

$E\tau_1 = 1$, and $E\tau_1^2 < \infty$ if $EX_1^4 < \infty$. Define now S_n as $W(\sum_{i=1}^n \tau_i)$, and let $T_n = W(n)$, then

$$|S_n - T_n| = \left|W\left(\sum_{j=1}^{n} \tau_j\right) - W(n)\right| \approx n^{\frac{1}{4}},$$

because $\sum_{j=1}^n \tau_j - n \approx n^{\frac{1}{2}}$.

This is the construction used by Strassen in [14] for proving that

$$|S_n - T_n| = O\left(n^{\frac{1}{4}}(\log n)^{\frac{1}{2}}(\log\log n)^{\frac{1}{4}}\right)$$

with probability one. He asked whether this construction was the best possible for approximating the sums of i.i.d.r.v.-s with normal ones.

This question was answered in the negative by Csörgő and Révész, who proved in [5] that there is a version of S_n, T_n such that

$$|S_n - T_n| = o(n^{\frac{1}{6}+\varepsilon})$$

with probability one if $EX_1^3 = 0$, $EX_1^8 < \infty$, the Cramer condition holds for $F(x)$ and there exists an interval $\infty \leqq A < B \leqq \infty$ such that $F(x)$ is strictly increasing in

$[A, B]$. The proof of their theorem indicates that there is a version of S_n, T_n such that

$$|S_n - T_n| = o\left(n^{\frac{1}{2r}+\varepsilon}\right)$$

with probability one if the first r moments of X_1 are equal to those of a standard normal distribution, and the first $3r-1$ moments of X_1 are finite. The question, however, remained unsolved, what happens if F is nice enough, say X_1 is bounded, but the moments of X_1 are not specialized.

Another result related to these questions is connected with the so-called "stochastic geyser problem". Assume that for some reason only the sample

$$S_n + r_n \qquad (n = 1, 2, \dots)$$

is available, where r_n is an arbitrary sequence of random variables such that $|r_n| \leqq R_n$ with some constant R_n. A theorem of Bártfai [1] states that if $R_n = o(\log n)$ and X_1 has a finite moment generating function, then the sequence $\{S_n + r_n\}_{n=1}^{\infty}$ determines the distribution function $F(x)$ with probability one. This indicates that the best possible bound for any version of S_n, T_n is $O(\log n)$. In this paper we prove the following:

Theorem 1. *Let F be a distribution function for which*

$$\int_{-\infty}^{\infty} x F(dx) = 0, \qquad \int_{-\infty}^{\infty} x^2 F(dx) = 1,$$

and there is a $t_0 > 0$ such that

$$R(t) = \int_{-\infty}^{\infty} e^{tx} F(dx) < \infty, \qquad \textit{for } |t| < t_0.$$

Furthermore one of the following two conditions holds:

i) $\int_{-\infty}^{\infty} |R(t+iu)|^p \, du < \infty$, *for some* $p \geqq 1$ *and all* t, $|t| < t_0$ *where* $R(z) = \int_{-\infty}^{\infty} e^{zx} F(dx)$ *for arbitrary complex* z *with* $|\mathrm{Re}\, z| < t_0$;

ii) *F is lattice-valued, i.e. there are constants a and b such that F is concentrated on the set of points $aj + b$, $j = 0, \pm 1, \pm 2, \dots$.*

Then there is a sequence of functions $f_n(y_1, y_2, \dots, y_{2n})$ such that if $Y_1, Y_2, \dots, Y_n, \dots$ are independent standard normal variables, then the random variables

$$X_n = f_n(Y_1, Y_2, \dots, Y_{2n}), \qquad n = 1, 2, \dots$$

are independent, and distributed according to the given distribution F. In addition

$$P\left(\sup_{1 \leqq k \leqq n} |S_k - T_k| > C \log n + x\right) < K e^{-\lambda x}, \tag{1.1}$$

where $S_k = \sum_{j=1}^{k} X_j$, $T_k = \sum_{j=1}^{k} Y_j$, x and n are arbitrary, and the positive constants C, K, λ depend only on F.

Corollary. *For S_k, T_k given in Theorem 1*

$$P\left(\limsup_{n \to \infty} \frac{|S_n - T_n|}{\log n} \leqq C\right) = 1,$$

where C is the same constant as in Theorem 1.

Note that if i holds with $p=1$, then the density function $f(x)=dF/dx$ exists. On the other hand if $R(t)<\infty$ for $|t|<t_0$ and $f(x)$ is bounded (or it is square integrable) then i holds.

Theorem 2. *Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d.r.v.-s with a distribution function F different from the standard normal one, and let $Y_1, Y_2, \ldots, Y_n, \ldots$ be independent standard normal variables, then there is a positive constant C_0 such that*

$$P\left(\limsup_{n\to\infty} \frac{|S_n - T_n|}{\log n} \geqq C_0\right) = 1, \tag{1.2}$$

where the constant C_0 depends only on F.

Our construction for proving Theorem 1 is a certain diadic approximation. It provides a sequence of i.i.d.r.v.-s of distribution F in an arbitrary probability space where a sequence of independent standard normal variables is given. There is no need for further randomization or any other enlargement of the given probability space as is usual when one applies the original Skorohod's embedding scheme. The diadic scheme is also applicable for approximation of the sample distribution function, so we can improve the bounds in the theorem of Breiman [3] and Brillinger [4].

Theorem 3. *For a fixed n let $X_1, X_2, \ldots, X_n$ be i.i.d.r.v.-s with*

$$P(X_1 < t) = \begin{cases} 0 & \text{if } t<0 \\ t & \text{if } 0 \leqq t \leqq 1 \\ 1 & \text{if } t>1, \end{cases} \tag{1.3}$$

let $F_n(t)$ be the empirical distribution function based on the sample $X_1, X_2, \ldots, X_n$; and let $B_n(t)$ be a Browninan bridge. There is a version of $F_n(t)$ and $B_n(t)$ such that

$$P\left(\sup_{0\leqq t\leqq 1} |n(F_n(t)-t) - n^{\frac{1}{2}} B_n(t)| > C\log n + x\right) < K e^{-\lambda x} \tag{1.4}$$

for all x, where C, K, λ are positive absolute constants.

Corollary. *Let B_n, F_n be the same as in Theorem 3, and let ψ be a functional defined on the space $D(0,1)$ of piece-wise continuous function, satisfying a Lipschitz condition*

$$|\psi(u) - \psi(v)| \leqq L \sup_{0\leqq x\leqq 1} |u(x) - v(x)|.$$

Assume further that the distribution of the random variable $\psi(B_n)$ has a bounded density. Then

$$\sup_x |P(\psi(n^{\frac{1}{2}}(F_n(t)-t)) < x) - P(\psi(B_n) < x)| = O\left(\frac{\log n}{n^{\frac{1}{2}}}\right). \tag{1.5}$$

It is known, that the Skorohod representation cannot be used to prove rates of convergence better than $O(n^{-\frac{1}{4}})$, at least if the functional ψ depends essentially on a single time, e.g. ψ is the Kolmogorov statistic. The rate $O\left(\frac{\log^\beta n}{n^{\frac{1}{2}}}\right)$ was achieved first by Sawyer [12], who investigated functionals of type $\psi(u)=\int_0^1 h(u(x),x)\,dx$, where $h(s,x)$ and its partial derivatives of order one are of slow growth in x. Note that there is no evidence of the necessity of the term $\log n$ in (1.5), so we do not know whether our Corollary is best possible.

The diadic sceme can be extended to the twodimensional case, hence we can give a simultaneous representation of F_n, B_n for $n=1,2,\ldots$ – as was proposed by Kiefer [7]. Our following theorem is an improvement of Kiefer's result, we think however, that a $\log n$ factor in (1.6) is still superfluous.

Theorem 4. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d.r.v.-s with the same distribution as in Theorem 3, let $F_n(t)$ be the empirical distribution function based on the sample $X_1, X_2, \ldots, X_n$; and let $B_1(t), B_2(t), \ldots$ be a sequence of independent Brownian bridges. There is a version of the sequences $F_n(t)$, $B_n(t)$ such that*

$$P\left(\sup_{1\leqq k\leqq n}\sup_{0\leqq t\leqq 1}\left|k(F_k(t)-t)-\sum_{j=1}^{k}B_j(t)\right|>(C\log n+x)\log n\right)<K e^{-\lambda x} \tag{1.6}$$

for all x and n, where C, K, λ are positive absolute constants.

2. The Quantile Transformation and the Diadic Scheme

The construction of Csörgő and Révész [5] is based on a transformation which will be called quantile transformation. Assume that the distribution functions $F(x)$, $G(x)$ are continuous, we are given a random variable Y of distribution G and should like to construct an X of distribution F such that the difference $|X-Y|$ be as small as possible. The random variables $F(X)$, $G(Y)$ are both uniformly distributed, hence it is a natural suggestion to define X by the equation

$$F(X)=G(Y),$$

i.e. to define X by $F^{-1}(G(Y))$. (Note that if F is continuous and strictly monotone the definition of the inverse F^{-1} is straightforward, otherwise one has to be a little careful.) This transformation was proposed by Bártfai [2], who proved that for any pairs $\tilde{X}$, $\tilde{Y}$ with given marginal distributions F, G

$$E(\tilde{X}-\tilde{Y})^2 \geqq E(X-Y)^2,$$

where X, Y are the above constructed variables.

If the distributions F and G are near to each other then the difference $|X-Y|$ is small, namely if F is the distribution of the sum of n independent random variables having 0 expectation and a finite moment generating function, and G is the distribution of an appropriately normalized normal one, then – as it will be proved in the last section – there are positive constants C_1, C_2, ε such that

$$|X-Y|\leqq C_1\frac{X^2}{n}+C_2 \qquad \text{if } |X|<\varepsilon n.$$

This means that $|X-Y|$ is practically bounded, the error of the approximation of X by Y does not increase with the number of components in X. Especially, when the first $r+2$ moments of the components in X are equal to the moments of a standard normal distribution then the difference $|X-Y|$ tends to 0 on the order $n^{-\frac{r}{2}}$. Hence if someone wants to approximate the partial sums

$$S_n = X_1 + X_2 + \cdots + X_n \quad \text{by} \quad T_n = Y_1 + Y_2 + \cdots + Y_n,$$

then he may divide the components into blocks, approximate first the total sum

$$X_{n_i+1} + X_{n_i+2} + \cdots + X_{n_{i+1}}$$

within a block as a quantile transformation of the sum

$$Y_{n_i+1} + Y_{n_i+2} + \cdots + Y_{n_{i+1}},$$

define somehow the elements of these blocks taking into account that their sums are given, and then fit the different blocks together. He will then be producing two kinds of error: on the one hand, the sum of the errors of the quantile transformations, and on the other hand, the maximum of the maxima of the partial sums in the individual blocks. This was the way followed by Bártfai and then by Csörgő and Révész: the equality of moments assured that the error of "first type" was rather small, hence they could choose short blocks in order to have a small error of "second type".

The basic idea of our construction is to extend the quantile transformation into the individual blocks. Assume that we are already given the sum of $2n$ elements

$$U_{2n} = X_{m+1} + X_{m+2} + \cdots + X_{m+2n}$$

as the quantile transformation of the sum

$$V_{2n} = Y_{m+1} + Y_{m+2} + \cdots + Y_{m+2n},$$

where the Y_j-s are given standard normal variables, but the summands X_j-s are not defined yet. Motivated by the elementary observation that V_{2n} is the sum of

$$L = Y_{m+1} + \cdots + Y_{m+n}, \quad \text{and} \quad R = Y_{m+n+1} + \cdots + Y_{m+2n},$$

hence V_{2n} is independent of the difference

$$\tilde{V}_{2n} = L - R,$$

one can expect that U_{2n} is "almost independent" of the random variable

$$\tilde{U}_{2n} = (X_{m+1} + \cdots + X_{m+n}) - (X_{m+n+1} + \cdots + X_{m+2n}).$$

Thus it seems reasonable to define $\tilde{U}_{2n}$ first (in such a way that the difference $|\tilde{U}_{2n} - \tilde{V}_{2n}|$ be as small as possible) because it is not much effected by the given value of U_{2n}. The simple quantile transformation of $\tilde{V}_{2n}$ gives us a random variable $\hat{U}_{2n}$ with the desired distribution, but this $\hat{U}_{2n}$ is unfortunately strictly independent of U_{2n}, therefore the joint distribution of the pair U_{2n}, $\hat{U}_{2n}$ is not equal to the desired joint distribution.

This can be avoided by using the conditional quantile transformation, i.e. to transform the uniformly distributed random variable $\Phi((2n)^{-\frac{1}{2}} V_{2n})$ by the inverse (relative to x) of the conditional distribution function

$$F(x|y)=P(\tilde{U}_{2n}<x|U_{2n}=y).$$

In such a way we can define the sums

$$X_{m+1}+\dots+X_{m+n}; \quad X_{m+n+1}+\dots+X_{m+2n}$$

as $\frac{1}{2}(U_{2n}+\tilde{U}_{2n})$ and $\frac{1}{2}(U_{2n}-\tilde{U}_{2n})$ respectively, and follow the construction down to the individual X_j-s—if n is a power of 2. Our construction consists of two parts: in the "forward part" we define the sums of the blocks using the quantile transformation and then, in the "backward part" we use the conditional quantile transformation to define step by step the two halves of the blocks we have just defined down to the individual terms X_j.

Conditions i) and ii) of Theorem 1 assure that with local theorems we can handle the conditional distributions. One can get rid of these conditions, and we will return to this question in the second part of our paper. In the second part we shall also investigate the case when the moment generating function of X_1 does not exist.

3. Proof of Theorem 1

Introduce the following notations: $S_0=T_0=0$,

$$\begin{aligned}
&U_j=S_{2^j}, && U_{j,k}=S_{(k+1)2^j}-S_{k\cdot 2^j},\\
&\tilde{U}_j=U_{j+1}-U_j, && \tilde{U}_{n,k}=U_{n-1,2k}-U_{n-1,2k+1},\\
&V_j=T_{2^j}, && V_{j,k}=T_{(k+1)\cdot 2^j}-T_{k\cdot 2^j},\\
&\tilde{V}_j=V_{j+1}-V_j, && \tilde{V}_{n,k}=V_{n-1,2k}-V_{n-1,2k+1},\\
&F_j(x)=P(U_j<x), && F_n(x|y)=P(\tilde{U}_{n,0}<x|U_{n,0}=y),
\end{aligned}$$

$$G_j(t)=\sup\{x\colon F_j(x)\leqq t\},$$

$$G_n(t|y)=\sup\{x\colon F_n(x|y)\leqq t\},$$

$$(j=0,1,2,\dots;\ k=0,1,2,\dots;\ n=1,2,\dots).$$

Note that at the present level of the construction neither the X_j-s, nor the S_n-s are yet defined, we are given only the Y_j-s and T_n-s. We know, however, the distributions $F_j(x)$, $F_n(x|y)$, and the above notations served only the aim of making their meaning clear. The random variables X_j-s will be defined through these distributions.

Define first U_0 by

$$U_0=G_0(\Phi(V_0)),$$

where $\Phi(x)$ is the unit normal distribution function, and define $\tilde{U}_j$ by

$$\tilde{U}_j=G_j(\Phi(2^{-\frac{j}{2}}\cdot\tilde{V}_j)), \quad j=0,1,2,\dots.$$

Then the distribution function of U_0 is $F_0(x)=F(x)$, and that of $\tilde{U}_j$ is $F_j(x)$. In fact,

$$P(U_0<x)=P(G_0(\Phi(V_0))<x)=P(\Phi(V_0)<F(x))=F(x),$$

and the same argument works for $\tilde{U}_j$. The random variables $V_0, \tilde{V}_0, \tilde{V}_1, \dots$ are independent, hence the constructed $U_0, \tilde{U}_0, \tilde{U}_1, \dots$ are also independent.

Starting with

$$U_{j,1}=\tilde{U}_j, \quad j=1,2,\dots$$

for any $j\geqq 1$, $k\geqq 1$, for which $U_{j,k}$ is already defined, define $\tilde{U}_{j,k}$ by

$$\tilde{U}_{j,k}=G_j(\Phi(2^{-\frac{j}{2}}\cdot\tilde{V}_{j,k})|U_{j,k}),$$

and then define $U_{j-1,2k}$, $U_{j-1,2k+1}$ by

$$U_{j-1,2k}=\tfrac{1}{2}(U_{j,k}+\tilde{U}_{j,k}),$$
$$U_{j-1,2k+1}=\tfrac{1}{2}(U_{j,k}-\tilde{U}_{j,k}).$$

A little algebra shows that in such a way all the variables $U_{j,k}$ are defined for $j=0,1,2,\dots$; $k=2,3,\dots$, hence the variables $X_k=U_{0,k-1}$ are defined for $k=3,4,\dots$. Note that X_1 and X_2 are already defined as U_0 and $\tilde{U}_0$.

Now we prove that the just defined X_j-s are independent, and their distribution function is the given $F(x)$. Let $j\geqq 0$ be a fixed integer. The random variables $\{X_k;\ 2^{j-1}<k\leqq 2^j\}$ are defined as a function of the random variables $\{Y_k;\ 2^{j-1}<k\leqq 2^j\}$, hence the blocks $B_j=\{X_k;\ 2^{j-1}<k\leqq 2^j\}$ are independent for different j-s. We have already seen that the distribution of X_1 and X_2 is F, so we may assume that $j\geqq 2$. We have already shown that the distribution of $U_{j-1,1}=\tilde{U}_{j-1}$ is the desired $F_{j-1}(x)$. Now we prove step by step for decreasing $i<j-1$ that the $U_{i,k}$-s are independent and their distribution is $F_i(x)$ for $2^{j-i-1}\leqq k<2^{j-i}$, i.e. for all k-s for which $U_{i,k}$ is the sum of some elements of the block B_j. The random variables $U_{i,2k}$, $U_{i,2k+1}$ are defined as functions of $U_{i+1,k}$ and $\tilde{U}_{i+1,k}$, while $\tilde{U}_{i+1,k}$ is a function of $U_{i+1,k}$ and $\tilde{V}_{i+1,k}$.

Our construction is based on the elementary fact that the differences $\tilde{V}_{m,n}$ are independent, and $\tilde{V}_{m,n}$ is independent of any $V_{M,N}$ if

$$N\cdot 2^M\leqq n\cdot 2^m<(n+1)\cdot 2^m\leqq(N+1)\cdot 2^M.$$

Hence $\tilde{V}_{i+1,k}$ is independent of the differences $\tilde{V}_{i+1,m}$ for $m\neq k$, and the $\tilde{V}_{i+1,k}$-s are independent of the $U_{i+1,m}$-s. Thus the pair $U_{i,2k}$, $U_{i,2k+1}$ is independent of the other $U_{i,m}$-s. Finally we prove that the joint distribution of the pair $U_{i,2k}$, $U_{i,2k+1}$ is the prescribed one, consequently $U_{i,2k}$ and $U_{i,2k+1}$ are independent and their distribution is the desired $F_i(x)$. Indeed,

$$P(\tilde{U}_{i+1,k}<x\,|\,U_{i+1,k}=y)=P(G_{i+1}(\Phi(2^{-\frac{i+1}{2}}\cdot\tilde{V}_{i+1,k})|y)<x)=F_{i+1}(x|y),$$

since $\tilde{V}_{i+1,k}$ and $U_{i+1,k}$ are independent.

Note that the partial sums of the constructed $X_1, X_2, \dots$ variables have a direct expression with the constructed $\tilde{U}_j$, $\tilde{U}_{j,k}$. Let $m=(2k+1)\cdot 2^j$, then

$$2S_m=S_{k\cdot 2^{j+1}}+S_{(k+1)\cdot 2^{j+1}}+\tilde{U}_{j+1,k}.$$

This recursion implies that for $2^n < m \leqq 2^{n+1}$

$$S_m = \tilde{S}_m + \sum_{i=j+1}^{n} c(i)\, \tilde{U}_{i,k(i)}, \tag{2.1}$$

where the process $\tilde{S}_m$ is obtained by linear interpolation:

$$\tilde{S}_m = \frac{2^{n+1}-m}{2^n} S_{2^n} + \frac{m-2^n}{2^n} S_{2^{n+1}} \qquad (2^n \leqq m \leqq 2^{n+1}), \tag{2.2}$$

the coefficients $c(i)$ and indices $k(i)$ depend on m, $0 \leqq c(i) \leqq 1$, and $k(i)$ is defined by

$$k(i) \cdot 2^i < m \leqq (k(i)+1) \cdot 2^i. \tag{2.3}$$

A similar representation is valid for T_m:

$$T_m = \tilde{T}_m + \sum_{i=j+1}^{n} c(i) \cdot \tilde{V}_{i,k(i)},$$

where

$$\tilde{T}_m = \frac{2^{n+1}-m}{2^n} T_{2^n} + \frac{m-2^n}{2^n} T_{2^{n+1}} \qquad (2^n \leqq m \leqq 2^{n+1}), \tag{2.4}$$

and $c(i)$, $k(i)$ are the same as in (2.1). Thus

$$|(S_m - T_m) - (\tilde{S}_m - \tilde{T}_m)| \leqq \sum_{i=j+1}^{n} |\tilde{U}_{i,k(i)} - \tilde{V}_{i,k(i)}|. \tag{2.5}$$

Now we pass to the estimation of the differences $|S_m - T_m|$. We shall use the following

Lemma 1. *If F satisfies the conditions given in Theorem* 1, *then there are positive constants C_1, C_2, ε such that*

$$|\tilde{U}_j - \tilde{V}_j| < C_1 \cdot 2^{-j} \cdot \tilde{U}_j^2 + C_2, \qquad \text{if } |\tilde{U}_j| < \varepsilon \cdot 2^j, \tag{2.6}$$

$$|\tilde{U}_{j,k} - \tilde{V}_{j,k}| < C_1 \cdot 2^{-j} \cdot (\tilde{U}_{j,k}^2 + U_{j,k}^2) + C_2, \qquad \text{if } |\tilde{U}_{j,k}| < \varepsilon \cdot 2^j,\ |U_{j,k}| < \varepsilon \cdot 2^j, \tag{2.7}$$

where the U-s, V-s are defined above.

This lemma will be proved in the last section. Now we prove that there are positive constants α, β, γ such that

$$P\big(\sup_{1 \leqq m \leqq 2^N} |S_m - T_m| > x\big) \leqq \gamma\, e^{\alpha N - \beta x} \tag{2.8}$$

for arbitrary N and x. Since (2.8) implies (1.1) with $\lambda = \beta$, $K = \gamma$, and $C = \frac{2}{\log 2} \cdot \frac{\alpha}{\beta}$, this completes the proof. Fix N and x and define M by

$$\frac{x}{8\varepsilon} < 2^M \leqq \frac{x}{4\varepsilon}, \tag{2.9}$$

and take $M_0=\inf(M,N)$. Introduce the following notations:

$$\Delta=\sup_{1\leqq m\leqq 2^N}|S_m-T_m|,$$

$$\Delta_1=\sup_{0\leqq j<2^{N-M_0}}\sup_{1\leqq k\leqq 2^{M_0}}|S_{j\cdot 2^{M_0}+k}-S_{j\cdot 2^{M_0}}|,$$

$$\Delta_2=\sup_{0\leqq j<2^{N-M_0}}\sup_{1\leqq k\leqq 2^{M_0}}|T_{j\cdot 2^{M_0}+k}-T_{j\cdot 2^{M_0}}|,$$

$$\Delta_3=\sup_{M_0<j\leqq N}|S_{2^j}-T_{2^j}|,$$

$$\Delta_4=\sup_{0<j<2^{N-M_0}}|S'_{j\cdot 2^{M_0}}-T'_{j\cdot 2^{M_0}}|,$$

where $S'_m=S_m-\tilde{S}_m$, $T'_m=T_m-\tilde{T}_m$, and $\tilde{S}_m$, $\tilde{T}_m$ are defined by (2.2) and (2.4). With these notations

$$\Delta\leqq\Delta_1+\Delta_2+\Delta_3+\Delta_4.$$

Let A denote the event

$$A=\{|U_{M_0,k}|\leqq\varepsilon\cdot 2^M,\ k=0,1,\ldots,2^{N-M_0}-1\},$$

then

$$P(\Delta>x)\leqq P(\Delta_1>\varepsilon\cdot 2^M)+P(\Delta_2>\varepsilon\cdot 2^M)+P\left(\Delta_3>\frac{x}{4},A\right)+P\left(\Delta_4>\frac{x}{4},A\right), \quad (2.10)$$

where the comma indicates the union of the events. Here the terms on the right-hand side can be estimated by $\gamma_i e^{\alpha_i N-\beta_i x}$ with appropriate $\alpha_i,\beta_i,\gamma_i$ $(i=1,2,3,4)$.

$$P(\Delta_1>\varepsilon\cdot 2^M)\leqq 2^{N-M_0}\cdot 2^{M_0}\sup_{1\leqq k\leqq 2^{M_0}}\big(P(S_k>\varepsilon\cdot 2^M)+P(S_k<-\varepsilon\cdot 2^M)\big).$$

Here

$$\begin{aligned}P(S_k>\varepsilon\cdot 2^M)&=P(t_1(S_k-\varepsilon\cdot 2^M)>0)\leqq E\exp\{t_1(S_k-\varepsilon\cdot 2^M)\}\\&=(R(t_1))^k e^{-\varepsilon t_1\cdot 2^M}\leqq(R(t_1)\,e^{-\varepsilon t_1})^{2^M}=(\rho(\varepsilon))^{2^M},\end{aligned}$$

where $\rho(\varepsilon)$ is the Chernoff-function, and t_1 is the corresponding argument:

$$\rho(\varepsilon)=\inf_t R(t)e^{-t\varepsilon}=R(t_1)e^{-t_1\varepsilon}.$$

(We will see in the proof of Lemma 1 that $\varepsilon<t_0$.) A similar estimation holds for $P(S_k<-\varepsilon\cdot 2^M)$ with $\rho(-\varepsilon)$, and for $P(\Delta_2>\varepsilon\cdot 2^M)$ with the Chernoff function of the normal distribution, hence

$$P(\Delta_1>\varepsilon\cdot 2^M)\leqq 2^{N+1}\cdot e^{-\beta_1\cdot 2^M}\leqq 2^{N+1}\exp\left\{-\frac{\beta_1 x}{8\varepsilon}\right\},$$

$$P(\Delta_2>\varepsilon\cdot 2^M)\leqq 2^{N+1}\cdot e^{-\beta_2\cdot 2^M}\leqq 2^{N+1}\exp\left\{-\frac{\beta_2 x}{8\varepsilon}\right\},$$

where $\beta_1=-\sup(\log\rho(\varepsilon),\log\rho(-\varepsilon))$, $\beta_2=\frac{1}{2}\varepsilon^2$.

We assume in the sequel $M\leqq N$, otherwise $\Delta_3=\Delta_4=0$. For estimating the third term in (2.10) we use the following truncated variables:

$$\tau_j=\begin{cases}2^{-j}\cdot\tilde{U}_j^2 & \text{if } |\tilde{U}_j|<\varepsilon\cdot 2^j\\ 0 & \text{otherwise.}\end{cases}$$

The moment generating functions of these variables are bounded:

$$E e^{t\tau_j} = \int_0^{\varepsilon^2 \cdot 2^j} e^{ty} dP(\tau_j < y) = 1 + \int_0^{\varepsilon^2 \cdot 2^j} t e^{ty} P(\tau_j \geqq y)\, dy \leqq 1 + 2\int_0^{\infty} t e^{ty} e^{-\delta y} dy = q(t),$$

if $t < \delta$, because if $0 \leqq y \leqq \varepsilon^2 \cdot 2^j$, then

$$\begin{aligned} P(\tau_j \geqq y) &\leqq P(\tilde{U}_j^2 \geqq y \cdot 2^j) \leqq P(\tilde{U}_j \geqq (y \cdot 2^j)^{\frac{1}{2}}) + P(\tilde{U}_j \leqq -(y \cdot 2^j)^{\frac{1}{2}}) \\ &\leqq \rho(y^{\frac{1}{2}} \cdot 2^{\frac{j}{2}})^{2^j} + \rho(-y^{\frac{1}{2}} \cdot 2^{\frac{j}{2}})^{2^j} \leqq 2(1 - \delta y \cdot 2^{-j}) \leqq 2 e^{-\delta y}, \end{aligned}$$

where δ is an appropriate constant for which

$$\rho(u) \leqq 1 - \delta u^2 \quad \text{for } |u| \leqq \varepsilon.$$

Now we estimate Δ_3 using Lemma 1:

$$\begin{aligned} P\left(\Delta_3 > \frac{x}{4}, A\right) &\leqq P\left(\sum_{j=M}^{N} |\tilde{U}_j - \tilde{V}_j| > \frac{x}{4}, A\right) \leqq P\left(\sum_{j=M}^{N} (C_1 \cdot 2^{-j} \tilde{U}_j^2 + C_2) > \frac{x}{4}, A\right) \\ &\leqq P\left(\sum_{j=1}^{N} (C_1 \tau_j + C_2) > \frac{x}{4}\right) = P\left(\exp\left\{t\left(\sum_{j=1}^{N} (C_1 \tau_j + C_2) - \frac{x}{4}\right)\right\} > 1\right) \\ &\leqq E \exp\left\{t\left(\sum_{j=1}^{N} (C_1 \tau_j + C_2) - \frac{x}{4}\right)\right\} \leqq \exp\left\{N \log q(C_1 t) + C_2 N t - \frac{tx}{4}\right\}, \end{aligned}$$

which is again of the form $\gamma_3 e^{\alpha_3 N - \beta_3 x}$ for $t = \dfrac{\delta}{2C_1}$.

Estimating the fourth term we use (2.5) and (2.7), so we have to investigate the sums of type

$$\delta_m = \sum_{j=M+1}^{N} 2^{-j} (\tilde{U}_{j,k(j)}^2 + U_{j,k(j)}^2),$$

where $k(j)$ is defined in (2.3) and m is a multiple of 2^M. All these sums have the same distribution because the joint distributions of the random variables $\{\tilde{U}_{j,k(j)}, U_{j,k(j)};\ j = M+1, \ldots, N\}$ are the same for different m.

The first sum is:

$$\sum_{j=M+1}^{N} 2^{-j} (\tilde{U}_{j,0}^2 + U_{j,0}^2) = 2 \sum_{j=M+1}^{N} 2^{-j} (U_{j-1}^2 + \tilde{U}_{j-1}^2),$$

here the second term is what we have just investigated when estimating Δ_3, and the first term is dominated by the second:

$$\sum_{j=M}^{N-1} 2^{-j} U_j^2 \leqq \frac{2}{(\sqrt{2}-1)^2} \cdot \left(2^{-M} \cdot U_M^2 + \sum_{j=M}^{N-2} 2^{-j} \cdot \tilde{U}_j^2\right)$$

when we apply the relation

$$\begin{aligned} U_{j+M}^2 &= (U_M + \tilde{U}_M + \tilde{U}_{M+1} + \cdots + \tilde{U}_{M+j-1})^2 \\ &\leqq (q^{-j} U_M^2 + q^{-j+1} \tilde{U}_M^2 + \cdots + \tilde{U}_{M+j-1}^2) \cdot \sum_{j=0}^{\infty} q^j \quad \text{with } q = 2^{-\frac{1}{2}}. \end{aligned}$$

Hence

$$
\begin{aligned}
P\left(\Delta_4>\frac{x}{4}, A\right) &\leqq 2^{N-M}\cdot P\left(\sum_{j=M+1}^{N}\left(C_1\cdot 2^{-j}(\tilde{U}_{j,0}^2+U_{j,0}^2)+C_2\right)>\frac{x}{4}, A\right)\\
&\leqq 2^{N-M}\cdot P\left(C_3\left(2^{-M}U_M^2+\sum_{j=M}^{N-2}2^{-j}\tilde{U}_j^2\right)+NC_2>\frac{x}{4}, A\right)\\
&\leqq 2^{N-M}\cdot P\left(2C_3\sum_{j=1}^{N}\tau_j+NC_2>\frac{x}{4}\right)\\
&\leqq \exp\left\{N\log q(2C_3t)+C_2Nt-\frac{tx}{4}\right\},
\end{aligned}
$$

which is of the form $\gamma_4 e^{\alpha_4 N-\beta_4 x}$ for $t=\frac{\delta}{4C_3}$.

4. Proof of Theorem 2

Assume firstly that the moment generating function of X_1 does not exist at some $t>0$. Then

$$\sum_{j=1}^{\infty} P(e^{tX_j}>j)=\sum_{j=1}^{\infty} P(e^{tX_1}>j)=\infty,$$

hence

$$P\left(\limsup_{j\to\infty}\frac{X_j}{\log j}\geqq\frac{1}{t}\right)=1.$$

On the other hand

$$P\left(\lim_{j\to\infty}\frac{|Y_j|}{\log j}=0\right)=1,$$

hence

$$P\left(\limsup_{n\to\infty}\frac{|S_n-T_n|}{\log n}\geqq\frac{1}{2t}\right)=1,$$

that is (1.2) holds with $C_0=\frac{1}{2t}$. Similarly (1.2) holds with $C_0=-\frac{1}{2t}$ if the moment generating function of X_1 does not exist at some $t<0$. Namely, (1.2) holds with arbitrary large C_0 if there is no proper interval around the origin where the moment generating function of X_1 would exist.

If the moment generating function of X_1 is finite for $-\infty\leqq a_1<t<a_2\leqq\infty$, denote the interior of the range of the function $R'(t)/R(t)$ by D:

$$D=\left\{x:\ x=\frac{R'(t)}{R(t)},\ a_1<t<a_2\right\}.$$

The Chernoff function

$$\rho(x)=\inf_{a_1<t<a_2} e^{-tx}R(t)$$

is an analitic function in D. Denote the function $-\log\rho(x)$ by $\pi(x)$:

$$\pi(x)=-\log\rho(x).$$

A theorem of Erdős and Rényi [6] states that for any $x \in D$, $x > EX_1$,

$$P\left(\lim_{n\to\infty} \sup_{1\leq j\leq n} \frac{\pi(x)}{\log n} \sum_{k=j}^{j+\left[\frac{\log n}{\pi(x)}\right]} X_k = x\right) = 1.$$

Especially, for the variables Y_j, this theorem says that for all $y > 0$

$$P\left(\lim_{n\to\infty} \sup_{1\leq j\leq n} \frac{y^2}{2\log n} \sum_{k=j}^{j+\left[\frac{2\log n}{y^2}\right]} Y_k = y\right) = 1.$$

It is trivial that in case $EX_1 \neq 0$, (1.2) holds with arbitrarily large C_0. If $EX_1 = 0$, the theorem of Erdős and Rényi implies

$$P\left(\liminf_{n\to\infty} \sup_{1\leq j\leq n} \frac{1}{\log n}\left|\sum_{k=1}^{j+\left[\frac{\log n}{\pi(x)}\right]} (X_k - Y_k)\right| \geq \left|\frac{x}{\pi(x)} - \left(\frac{2}{\pi(x)}\right)^{\frac{1}{2}}\right|\right) = 1$$

for all $x \in D$, $x > 0$. For $x \in D$, $x < 0$ a similar statement holds but with "inf" instead of "sup". Hence (1.2) holds with

$$C_0 = \sup_{x\in D} \frac{|(2\pi(x))^{\frac{1}{2}} - x|}{2\pi(x)} = \sup_{a_1 < t < a_2} \frac{|(2(tr'(t) - r(t))^{\frac{1}{2}} - r'(t)|}{2(tr'(t) - r(t))},$$

where $r(t) = \log R(t)$. Since different distributions have different Chernoff functions, this C_0 is positive for arbitrary distribution different from the standard normal one.

Remarks. 1. Let F be a distribution function obeying the conditions of Theorem 1, and let X_n, S_n, T_n, C be the same as in Theorem 1. Then for arbitrary positive integer

$$P\left(\limsup_{n\to\infty} \frac{|S_{nk} - T_{nk}|}{\log n} \leq C\right) = 1.$$

Hence for the random variables

$$\tilde{X}_n = \frac{1}{\sqrt{k}} \sum_{j=(n-k)^{k+1}}^{nk} X_j$$

the best possible bound C_0 in (1.2) is less than or equal to $k^{-\frac{1}{2}} C$. This means that if we characterize the distance between an arbitrary distribution F and the standard normal one by the largest possible constant C_0 in (1.2) (assuming X_1 has the distribution F), then for arbitrary $\varepsilon > 0$ there is a distribution with a characteristic constant less than ε.

2. For empirical distribution functions a similar statement is valid. Let $F_n(t)$ be the empirical distribution function defined in Theorem 4, and let $B_1(t), B_2(t), \ldots$ be a sequence of independent Brownian bridges. Then Theorem 2 implies

$$P\left(\limsup_{n\to\infty} \sup_{0\leq t\leq 1} \frac{\left|n(F_n(t) - t) - \sum_{j=1}^{n} B_j(t)\right|}{\log n} \geq \frac{1}{8}\right) = 1.$$

In the same way as Theorem 2 was proved, one can prove

$$\lim_{n\to\infty} P\left(\sup_{0\le t\le 1} \frac{\left|n(F_n(t)-t)-\sum_{j=1}^{n} B_j(t)\right|}{\log n} \ge \frac{1}{6}\right)=1.$$

(We remark that the constant given here is not the best possible one.) That is, the construction of Theorem 3 has nearly the smallest possible error, since its error is about $C\log n$ but perhaps this C is not the smallest possible. The construction of Theorem 4 is, however, far from being best possible because it is very likely that

$$P\left(\limsup_{n\to\infty} \sup_{0\le t\le 1} \frac{\left|n(F_n(t)-t)-\sum_{j=1}^{n} B_j(t)\right|}{\log n} \le C\right)=1$$

is available.

5. Proof of Theorem 3 and 4

The proofs of Theorems 3 and 4 are essentially the same as that of Theorem 1. That is why we will only sketch them out here and try to use the same notations. First we start with the proof of Theorem 3. Here the first difference is the lack of the "forward part" of the construction, because now we are given a fixed n and we have to construct only the first n elements of the sample $X_1, X_2, \dots, X_n$. Assume that we are given a Brownian bridge $B_n(t)$ on the interval (0, 1). Then the random variables

$$V_{j,k}=B_n((k+1)\cdot 2^{-j})-B_n(k\cdot 2^{-j}) \qquad (k=0,1,\dots,2^j-1;\ j=1,2,\dots) \tag{5.1}$$

are normally distributed with expectation 0 and variance $(1-2^{-j})\cdot 2^{-j}$. The random variables

$$\tilde{V}_{j,k}=V_{j+1,2k}-V_{j+1,2k+1} \qquad (k=0,1,\dots,2^j-1;\ j=0,1,\dots) \tag{5.2}$$

are also normal with expectation 0 and variance 2^{-j}, especially $\tilde{V}_{0,0}$ is a standard normal variable, moreover the $\tilde{V}_{j,k}$-s are mutually independent. (Note that the index j has a different role here than in the first proof, which accounts for the difference in the definition of $\tilde{V}_{j,k}$.)

The first step of the construction is the definition of some $U_{1,0}$ which is a random variable of binomial distribution with parameters n and $\frac{1}{2}$ as a quantile transform of $\tilde{V}_{0,0}$. Let the function $H(t|n)$ be defined by

$$H(t|m)=\sup\left\{k: \sum_{i=0}^{k-1}\binom{m}{i} 2^{-m}\le t\right\} \qquad (m=0,1,\dots), \tag{5.3}$$

then $U_{1,0}$ is defined by

$$U_{1,0}=H(\Phi(\tilde{V}_{0,0})|n).$$

Note that in the present situation it is more natural to define the variables $U_{j,k}$ directly rather than *via* the variables $\tilde{U}_{j,k}$. Starting with $U_{1,0}$ which we have just defined and

$$U_{1,1}=n-U_{1,0},$$

we define step by step the variables $U_{j,2k}$ by

$$U_{j,2k}=H\big(\Phi(2^{\frac{j-1}{2}}\tilde{V}_{j-1,k})\,|\,U_{j-1,k}\big) \qquad (k=0,1,\ldots,2^{j-1}),$$

and $U_{j,2k+1}$ by

$$U_{j,2k+1}=U_{j-1,k}-U_{j,2k} \qquad (k=0,1,\ldots,2^{j-1}-1)$$

for $j=2,3,\ldots$, and finally $\tilde{U}_{j,k}$ is defined by

$$\tilde{U}_{j,k}=U_{j+1,2k}-U_{j+1,2k+1} \qquad (k=0,1,\ldots,2^j-1;\ j=0,1,\ldots).$$

In this construction $\sup\limits_{0\leq k\leq 2^j} U_{j,k}$ tends to 1 with probability 1. The sequence of intervals $(k\cdot 2^{-j},(k+1)\cdot 2^{-j})$ for pairs j,k with $U_{j,k}>0$ defines n points in $(0,1)$: let these points be the ordered sample $X_1^*, X_2^*, \ldots, X_n^*$ and define $X_1, X_2, \ldots, X_n$ by a random permutation π:

$$X_i=X^*_{\pi(i)} \qquad (i=1,2,\ldots,n).$$

Then for the empirical distribution $F_n(x)$

$$F_n(k\cdot 2^{-j})=\frac{1}{n}\sum_{i=0}^{k-1} U_{j,i} \qquad (k=1,2,\ldots,2^j;\ j=1,2,\ldots)$$

holds true. It is easy to see that the constructed X_i-s are i.i.d.r.v.-s with distribution given in Theorem 3. To estimate the errors of this construction we shall use the following

Lemma 2. *There are positive absolute constants C_1, C_2, ε such that*

$$|\tilde{U}_{j,k}-n^{\frac{1}{2}}\tilde{V}_{j,k}|<C_1\cdot 2^j\cdot n^{-1}\big(\tilde{U}^2_{j,k}+(U_{j,k}-n\cdot 2^{-j})^2\big)+C_2, \tag{5.4}$$

if

$$|\tilde{U}_{j,k}|<\varepsilon\cdot n\cdot 2^{-j}, \qquad |U_{j,k}-n\cdot 2^{-j}|<\varepsilon\cdot n\cdot 2^{-j}.$$

The details of the estimation will be omitted because they are just the same as in the first proof. The only difference is that the variables $U_{j,1}$ ($j=1,2,\ldots$) are not strictly independent. The effect of their dependence is however negligible as is easily shown by the standard representation of the empirical distribution function as a conditional Poisson process.

The construction of Theorem 4 is the following. Let $B_1(t), B_2(t), \ldots$ be a sequence of independent Brownian bridges. In the "forward part" of the construction the elements of this sequence will be divided into blocks of 2^n elements ($n=0,1,2,\ldots$), and the sums

$$\sum_{j=2^k+1}^{2^{k+1}} B_j(t)$$

will be transformed one by one into empirical distributions in the same way as in the proof of Theorem 3. The order of the elements of a block

$$\{X_{2^k+1}, X_{2^k+2}, \ldots, X_{2^{k+1}}\}$$

is, however, defined in this step only provisionally. These elements will be re-ordered step by step in the "backward part" of the construction. Here we are faced with the problem of dividing a given sample into two parts in such a way that the differences of the empirical distributions given by these parts be near to a given Brownian bridge. This construction is formulated in the following

Theorem 5. *For $n=2m$ let E_n be the n-dimensional space, C the space of all continuous functions on* (0, 1) *with the usual σ-algebra, and let Π_n be the set of permutations of the first n integers. There is a measurable mapping ψ_n from $E_n \times C \times \Pi_n$ to Π_n having the following properties: Let $X_1, X_2, \ldots, X_n$ be i.i.d.r.v.-s of distribution* (1.3), *$B_n(t)$ a Brownian bridge, p_n a random element of Π_n and let the X_i-s B_n, p_n be mutually independent. Let the permutation*

$$\psi_n(X_1, X_2, \ldots, X_n, B_n(t), p_n)$$

be denoted by π, and let $F_n^{(L)}(t)$, $F_n^{(R)}(t)$ be the empirical distribution functions defined by the samples

$$S^{(L)}=\{X_{\pi(1)}, \ldots, X_{\pi(m)}\}, \qquad S^{(R)}=\{X_{\pi(m+1)}, \ldots, X_{\pi(n)}\}. \tag{5.5}$$

Then π is a random element of Π_n i.e. all the elements of Π_n have the same probability in the distribution generated by ψ_n, and there are positive absolute constants C, K, λ such that

$$P\big(\sup_{0\leqq t\leqq 1}|n(F_n^{(L)}(t)-F_n^{(R)}(t))-n^{\frac{1}{2}}B_n(t)|>C\log n+x\big)<Ke^{-\lambda x}. \tag{5.6}$$

for all x.

In the "backward part" of construction of Theorem 4 the Brownian bridges used in the subsequent reordering of the blocks of X_i-s will be the process

$$\tilde{V}_{j,k}(t)=\sum_{i=2k\cdot 2^{j-1}+1}^{(2k+1)2^{j-1}} B_i(t)-\sum_{i=(2k+1)2^{j-1}+1}^{(2k+2)2^{j-1}} B_i(t)$$

(they are independent). Using an estimation similar to (2.5) one can prove the statement (1.6) simply by summing up the inequalities (5.6) for appropriate n-s. Hence in proving Theorem 5 the proof of Theorem 4 will be completed.

The proof of Theorem 5 is as follows. Let $V_{j,k}$, $\tilde{V}_{j,k}$ be again defined by (5.1) and (5.2), let $F_n(t)$ be the empirical distribution function defined by the sample $X_1, \ldots, X_n$, and let $W_{j,k}$ be defined by

$$W_{j,k}=n[F_n((k+1)\cdot 2^{-j})-F_n(k\cdot 2^{-j})] \qquad (k=0,1,\ldots,2^j-1; j=1,2,\ldots).$$

Actually there is no need to give the complete permutation π, what is used in (5.6) are the two samples $S^{(L)}$, $S^{(R)}$ given in (5.5). (The letters L and R stand for left and right.) We use the permutations only to make the formulation easier (especially in the proof of Theorem 4). That is why we divide the number of points $W_{j,k}$ step by step into two parts, denoting each by $U_{j,k}^{(L)}$ and $U_{j,k}^{(R)}$. We shall use a pure random halving, generated by the random variables $\tilde{V}_{j,k}$. Let us first define $U_{1,0}^{(L)}$: the number of the sample $S^{(L)}$ in the interval $(0,\frac{1}{2})$. By giving $U_{1,0}^{(L)}$, we have also defined $U_{1,0}^{(R)}$, $U_{1,1}^{(L)}$, $U_{1,1}^{(R)}$, because

$$\begin{aligned} &U_{1,0}^{(L)}+U_{1,0}^{(R)}=W_{1,0}, \qquad U_{1,1}^{(L)}+U_{1,1}^{(R)}=W_{1,1},\\ &U_{1,0}^{(L)}+U_{1,1}^{(L)}=m; \qquad U_{1,0}^{(R)}+U_{1,1}^{(R)}=m. \end{aligned} \tag{5.7}$$

If $W_{1,0}$, $W_{1,1}$ are given, the desired random variable $U_{1,0}^{(L)}$ is distributed according to a hypergeometric distribution. For any integers $A+B=C+D=N$ let the function $H(t|A;B;C;D)$ be defined by

$$H(t|A;B;C;D)=\sup\left\{k: \sum_{i=0}^{k-1} \frac{\binom{A}{k}\binom{B}{C-k}}{\binom{N}{C}} \leqq t\right\} \tag{5.8}$$

then $U_{1,0}^{(L)}$ is defined by the appropriate quantile transformation:

$$U_{1,0}^{(L)}=H(\Phi(\tilde{V}_{0,0})|W_{1,0};W_{1,1};m;m).$$

The other three variables $U_{1,0}^{(R)}$, $U_{1,1}^{(L)}$, $U_{1,1}^{(R)}$ are given by (5.7). Starting with these variables we define step by step the variables $U_{j,2k}^{(L)}$ by

$$U_{j,2k}^{(L)}=H(\Phi(2^{-\frac{j-1}{2}}\tilde{V}_{j-1,k})|W_{j,2k};W_{j,2k+1};U_{j-1,k}^{(L)};U_{j-1,k}^{(R)}),$$

and $U_{j,2k}^{(R)}$, $U_{j,2k+1}^{(L)}$, $U_{j,2k+1}^{(R)}$ by

$$U_{j,2k}^{(L)}+U_{j,2k}^{(R)}=W_{j,2k}, \qquad U_{j,2k+1}^{(L)}+U_{j,2k+1}^{(R)}=W_{j,2k+1},$$

$$U_{j,2k}^{(L)}+U_{j,2k+1}^{(L)}=U_{j-1,k}^{(L)}, \qquad U_{j,2k}^{(R)}+U_{j,2k+1}^{(R)}=U_{j-1,k}^{(R)},$$

for $k=0,1,\ldots,2^{j-1}-1$; $j=2,3,\ldots$. Finally define $\tilde{U}_{j,k}$ by

$$\tilde{U}_{j,k}=(U_{j+1,2k}^{(L)}-U_{j+1,2k+1}^{(L)})-(U_{j+1,2k}^{(R)}-U_{j+1,2k+1}^{(R)}).$$

The proof will be completed by the following

Lemma 3. *There are positive absolute constants* C_1, C_2, ε *such that*

$$\begin{aligned}|\tilde{U}_{j,k}-n^{\frac{1}{2}}\tilde{V}_{j,k}|&<C_1\cdot 2^j\cdot n^{-1}[(W_{j,2k}-n\cdot 2^{-j})^2\\&+(W_{j,2k+1}-n\cdot 2^{-j})^2+(U_{j-1,k}^{(L)}-n\cdot 2^{-j})^2+(U_{j-1,k}^{(R)}-n\cdot 2^{-j})^2]+C_2\end{aligned} \tag{5.9}$$

if

$$|W_{j,2k}-n\cdot 2^{-j}|<\varepsilon\cdot n\cdot 2^{-j}, \qquad |W_{j,2k+1}-n\cdot 2^{-j}|<\varepsilon\cdot n\cdot 2^{-j},$$

$$|U_{j-1,k}^{(L)}-n\cdot 2^{-j}|<\varepsilon\cdot n\cdot 2^{-j}, \qquad |U_{j-1,k}^{(R)}-n\cdot 2^{-j}|<\varepsilon\cdot n\cdot 2^{-j}.$$

6. Proof of the Lemmas

The proof of the lemmas is based on the theorem of large deviations, or rather on the central limit theorem with an effectiv error term. We use the following theorems of Petrov [9, 10].

Theorem A. *Let* $X_1, X_2, \ldots$ *be i.i.d.r.v.-s for which*

$$EX_1=0, \qquad EX_1^2=1, \qquad R(t)=E\exp\{tX_1\}<\infty \quad \text{for } |t|<t_0. \tag{6.1}$$

Then there is a positive η *such that for all* n

$$P(S_n<-nx)=\Phi(-n^{\frac{1}{2}}x)\exp\{-nx^3\lambda(-x)+O(x+n^{-\frac{1}{2}})\},$$

$$P(S_n>nx)=Q(n^{\frac{1}{2}}x)\exp\{nx^3\lambda(x)+O(x+n^{-\frac{1}{2}})\},$$

where

$$S_n = \sum_{j=1}^{n} X_j, \qquad Q(x) = 1 - \Phi(x),$$

the function $\lambda(x)$ *is analytic and the function* $O(x)$ *is uniform on the interval* $0 \leqq x \leqq \eta$.

Theorem B. *Let* $X_1, X_2, \ldots$ *be i.i.d.r.v.-s for which* (6.1) *holds, and for some* $p > 1$

$$\int_{-\infty}^{\infty} |R(t+iu)|^p \, du < \infty \qquad \text{for } |t| < t_0,$$

where $R(z) = E \exp\{z X_1\}$ *for arbitrary complex* z *with* $|\mathrm{Re}\, z| < t_0$.

Set $S_n = \sum_{j=1}^{n} X_j$, $F_n(x) = P(S_n < x)$ *and* $f_n(x) = \frac{d}{dx} F_n(x)$. *There is a positive* η *such that for all* n

$$f_n(nx) = n^{-\frac{1}{2}} \varphi(n^{\frac{1}{2}} x) \exp\{n x^3 \lambda(x) + O(|x| + n^{-\frac{1}{2}})\}$$

where the function $\lambda(x)$ *is analytic and the function* $O(x)$ *is uniform on the interval* $-\eta \leqq x \leqq \eta$.

Theorem C. *Let* $X_1, X_2, \ldots$ *be i.i.d.r.v.-s for which* (6.1) *holds, and* X_1 *is lattice-valued, i.e. there are constants* $a > 0$ *and* b *such that*

$$\sum_{j=-\infty}^{\infty} P(X_1 = aj + b) = 1.$$

Assume that a *is the largest such value. Then there is a positive* η *such that for all* n

$$P(S_n = nx) = n^{-\frac{1}{2}} \varphi(n^{\frac{1}{2}} x) \exp\{n x^3 \lambda(x) + O(|x| + n^{-\frac{1}{2}})\}$$

if x *is of the form* $x = \frac{1}{n}(aj + b)$, *where the function* $\lambda(x)$ *is analytic and* $O(x)$ *is uniform on the interval* $-\eta \leqq x \leqq \eta$.

Theorem A is given in [10], Theorems B and C are given in [9]. (Theorems B and C go back to Richter [10].) In fact Petrov stated these theorems for $x \to 0$ as $n \to \infty$, and for $|x| \leqq \eta$ his error term is a multiple of η. His proof is however also applicable for proving the above theorems.

Now we turn to the proof of formula (2.6) using Theorem A. We use the notation $n = 2^j$ in the sequel.

One has

$$F_j(\tilde{U}_j) \leqq \Phi(n^{-\frac{1}{2}} \tilde{V}_j) \leqq F_j(\tilde{U}_j + 0).$$

Denoting $\tilde{U}_j = nx$ (2.6) reduces to proving

$$\begin{aligned} -C_1 n x^2 - C_2 &\leqq nx - \sqrt{n}\, \Phi^{-1}(F_j(nx+0)) \\ &\leqq nx - \sqrt{n}\, \Phi^{-1}(F_j(nx)) \leqq C_1 n x^2 + C_2 \qquad \text{if } |x| < \varepsilon. \end{aligned}$$

Let us consider the case $x \geqq 0$. The previous inequality can be rewritten as

$$\begin{aligned} 1 - \Phi(\sqrt{n}x - u) &\geqq 1 - F_j(nx) \geqq 1 - F_j(nx+0) \\ &\geqq 1 - \Phi(\sqrt{n}x + u), \end{aligned}$$

where $u = C_1 \sqrt{n} x^2 + \frac{C_2}{\sqrt{n}}$, or by Theorem A

$$\log \frac{1-\Phi(\sqrt{n}x-u)}{1-\Phi(\sqrt{n}x)} \geqq n x^3 \lambda(x) + O\left(x + \frac{1}{\sqrt{n}}\right)$$
$$\geqq \log \frac{1-\Phi(\sqrt{n}x+u)}{1-\Phi(\sqrt{n}x)}.$$

But

$$\log \frac{1-\Phi(\sqrt{n}x-u)}{1-\Phi(\sqrt{n}x)} = u \frac{\varphi(\xi)}{1-\Phi(\xi)} > n x^3 \lambda(x) + O\left(x + \frac{1}{\sqrt{n}}\right),$$

where $\sqrt{n}x > \xi > \sqrt{n}x - u$.

The other side of the inequality and the case $x<0$ can be proved the same way.

To prove (2.7) we need the following relation

$$\begin{aligned} 1 - F_j(nx|ny) &= [1-\Phi(\sqrt{n}x)] \\ &\quad \cdot \exp\left[O\left(n|x|^3 + nx^2|y| + |y| + \frac{1}{\sqrt{n}}\right)\right] \quad \text{if } x \geqq 0 \\ F_j(nx|ny) &= \Phi(\sqrt{n}x) \\ &\quad \cdot \exp\left[O\left(n|x|^3 + nx^2|y| + |y| + \frac{1}{\sqrt{n}}\right)\right] \quad \text{if } x < 0 \end{aligned} \tag{6.2}$$

with $O(.)$ uniform for $|x|<\varepsilon$, $|y|<\varepsilon$.

Instead of (2.7) we prove the following somewhat stronger statement

$$|\tilde{U}_{j,k} - \tilde{V}_{j,k}| < C\left[\frac{\tilde{U}_{j,k}^2}{2^j} + \frac{|\tilde{U}_{j,k} U_{j,k}|}{2^j} + \frac{|U_{j,k}|}{2^{j/2}} + 1\right] \qquad \text{if } |U_{j,k}| < \varepsilon \cdot 2^j,\ |\tilde{U}_{j,k}| < \varepsilon \cdot 2^j.$$

Denoting $\tilde{U}_{j,k} = nx$, $U_{j,k} = ny$ this statement reduces to the following one:

$$\begin{aligned} \Phi(\sqrt{n}x+u) &\geqq F_j(nx+O|ny) \geqq F_j(nx|ny) \\ &\geqq \Phi(\sqrt{n}x-u) \quad \text{if } |x|<\varepsilon,\ |y|<\varepsilon, \end{aligned}$$

where $u = C\left[\sqrt{n}x^2 + \sqrt{n}xy + y + \frac{1}{\sqrt{n}}\right]$.

Taking logarithm and using formula (6.2) one proves this statement similarly to (2.6). (We may suppose that ε is small enough thus in the relation

$$\frac{1-\Phi(\xi-u)}{1-\Phi(x)} = u \frac{\varphi(\xi)}{\Phi(\xi)}$$

ξ is sufficiently near x.)

We shall prove (6.2) by integrating the conditional density function. In the following estimations K will always mean a sufficiently large constant, not necessarily the same in different formulas.

Denoting $\frac{d}{dx} F_j(x|y)$ by $f_j(x|y)$ and $\frac{d}{dx} F_j(x)$ by $f_j(x)$ one obtains

$$f_j(nx|ny)=\frac{f_{j-1}\left(\frac{n}{2}(y-x)\right) f_{j-1}\left(\frac{n}{2}(x+y)\right)}{2f_j(ny)}.$$

If $|x|<\eta$, $|y|<\eta$, then using Theorem B and the Taylor expansion of $\lambda(y\pm x)$ at the point y, one gets

$$\begin{aligned} f_j(nx|ny)&=n^{-\frac{1}{2}}\,\varphi(n^{\frac{1}{2}}x)\exp[n\mu(x,y)+O(|x|+|y|+n^{-\frac{1}{2}})]\\ &=n^{-\frac{1}{2}}\,\varphi(n^{\frac{1}{2}}x)\exp[O(n\,|x|^3+nx^2\,|y|+|y|+n^{-\frac{1}{2}})], \end{aligned}$$

where $\mu(x,y)=\frac{1}{2}(y+x)^3\,\lambda(y+x)+\frac{1}{2}(y-x)^3\,\lambda(y-x)-y^3\,\lambda(y)$.

We shall only prove the first relation of (6.2), a proof of the second being the same as the first.

For appropriate $A>0$

$$\begin{aligned} I_1&=F_j(nA|ny)-F_j(nx|ny)=\int_x^A nf_j(nt|ny)\,dt\\ &\leqq\int_x^A\sqrt{\frac{n}{2\pi}}\left(1+K\,|y|+\frac{K}{\sqrt{n}}\right)\exp\left(-\frac{nt^2}{2}(1-Kt-K\,|y|)\right)dt. \end{aligned}$$

On substituting $\frac{u^2}{2}=\frac{nt^2}{2}(1-Kt-K\,|y|)$ we get

$$\begin{aligned} I_1&\leqq\int_{\sqrt{n}\bar{u}}^{\infty}\left(1-\frac{Ku}{\sqrt{n}}+\frac{K}{\sqrt{n}}+K\,|y|\right)\varphi(u)\,du\\ &=\left(1+K\,|y|+\frac{K}{\sqrt{n}}\right)(1-\Phi(\sqrt{n}\bar{u}))+K\varphi(\sqrt{n}\bar{u})\cdot\frac{1}{\sqrt{n}}\\ &\leqq\left(1+K\,|y|+K\bar{u}+\frac{K}{\sqrt{n}}\right)(1-\Phi(\sqrt{n}\bar{u})), \end{aligned}$$

where $\bar{u}^2=x^2(1-Kx-K\,|y|)$.

Now

$$\log\frac{1-\Phi(\sqrt{n}\bar{u})}{1-\Phi(\sqrt{n}x)}=\sqrt{n}(x-\bar{u})\frac{\varphi(\xi)}{1-\Phi(\xi)};\qquad \sqrt{n}\bar{u}<\xi<\sqrt{n}x,$$

which gives

$$\frac{1-\Phi(\sqrt{n}\bar{u})}{1-\Phi(\sqrt{n}x)}\leqq\exp[O(nx^3+nx^2\,|y|+\sqrt{n}x\,|y|+\sqrt{n}x^2)].$$

These results imply that

$$I_1\leqq[1-\Phi(\sqrt{n}x)]\exp\left[K\left(nx^3+nx^2\,|y|+|y|+\frac{1}{\sqrt{n}}\right)\right].$$

Similarly

$$I_1\geqq[1-\Phi(\sqrt{n}x)]\exp\left[-K\left(nx^3+nx^2\,|y|+|y|+\frac{1}{\sqrt{n}}\right)\right].$$

We still must prove that the quantity

$$I_2 = 1 - F_j(nA|ny)$$

is negligible.

The proof of Theorem B implies that there is a positive constant K such that

$$f_{j-1}\left(\frac{n}{2}x\right) \leqq K f_{j-1}\left(\frac{n}{2}y\right) \cdot \exp\left(-\frac{n}{2}t(x-y)\right),$$

where t is defined by the equation $\dfrac{R'(t)}{R(t)} = y.$

Thus fixing a number y_0, $0 < y_0 < \dfrac{A}{2}$ and defining t_0 by the equation $\dfrac{R'(t_0)}{R(t_0)} = y_0$ we get for $x > A$, $y < \varepsilon$ the estimation

$$f_j(nx|ny) \leqq \frac{f_{j-1}\left(\frac{n}{2}(x+y)\right)}{f_j(ny)},$$

and taking into account the inequality

$$f_{j-1}\left(\frac{n}{2}(x+y)\right) = f_{j-1}\left(\frac{n}{2}(y_0 + x - y_0 + y)\right)$$

$$\leqq K f_{j-1}\left(\frac{n}{2}y_0\right) \exp\left[-\frac{n}{2}t_0(x+y-y_0)\right]$$

we finally obtain

$$f_j(nx|ny) \leqq \exp\left[-\frac{n}{4}t_0 x\right]$$

which ensures that I_2 is small enough.

If F fulfills condition i) of Theorem 1, the proof is based on Theorem C and it is just the same as the above proof.

In the case of Lemma 2 a binomial distribution of random parameter is approximated by the normal distribution. Let us denote the number of heads in m independent coin tossings by Z_m and let Y be a standard normal variable. It is a consequence of Lemma 1 that the quantile transformation provides a version of Z_m and Y such that

$$\left|\left(Z_m - \frac{m}{2}\right) - \frac{1}{2}m^{\frac{1}{2}}Y\right| \leqq \frac{C_1}{m}\left(Z_m - \frac{m}{2}\right)^2 + C_2 \quad \text{for } \left|Z_m - \frac{m}{2}\right| < \varepsilon m.$$

In the case when the number m is itself a random variable with expectation M, and we have to approximate $\left(Z_m - \dfrac{m}{2}\right)$ with $\frac{1}{2}M^{\frac{1}{2}}Y$, then the following estimation is applicable:

$$\left|\left(Z_m - \frac{m}{2}\right) - \frac{1}{2}M^{\frac{1}{2}}Y\right| \leqq \left(\frac{M}{m}\right)^{\frac{1}{2}} \left|\left(Z_m - \frac{m}{2}\right) - \frac{1}{2}m^{\frac{1}{2}}Y\right|$$

$$+ \left|\left(\frac{M}{m}\right)^{\frac{1}{2}} - 1\right| \cdot \left|Z_m - \frac{m}{2}\right|,$$

where

$$\left|\left(\frac{M}{m}\right)^{\frac{1}{2}}-1\right|\cdot\left|Z_m-\frac{m}{2}\right|\leqq\frac{(m-M)^2}{2M}+\frac{1}{2m}\left(Z_m-\frac{m}{2}\right)^2.$$

That is we can correct the effect of the random parameter by a linear transformation.

In the case of Lemma 3 the situation is the same: we approximate a hypergeometric distribution of random parameters by the normal distribution. In this case even the expectation differs from the theoretical value 0, the effect of the difference in the expectation and variance is however estimable by the terms given on the right hand side of (5.9). As for the approximation of a hypergeometric distribution with deterministic parameters through a normal distribution, Lemma 1 is not in fact applicable, however a direct application of Stirling's formula provides the desired estimation.

References

1. Bártfai, P.: Die Bestimmung der zu einem wiederkehrenden Prozeß gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation. Studia Sci. Math. Hungar. **1**, 161–168 (1966)
2. Bártfai, P.: Über die Entfernung der Irrfahrtswege. Studia Sci. Math. Hungar. **5**, 41–49 (1970)
3. Breiman, L.: Probability. Reading, Mass: Addison-Wesley, 1968
4. Brillinger, D.R.: An asymptotic representation of the sample *df.* Bull. Amer. Math. Soc. **75**, 545–547 (1969)
5. Csörgő, M., Révész, P.: A new method to prove Strassen type laws of Invariance Principle I.–II. Z. Wahrscheinlichkeitstheorie verw. Geb. **31**, 255–259 and 261–269 (1975)
6. Erdős, P., Rényi, A.: On a new law of large numbers. J. Analyse Math. **23**, 103–111 (1970)
7. Kiefer, J.: On the deviations in the Skorohod-Strassen approximation scheme. Z. Wahrscheinlichkeitstheorie verw. Geb. **13**, 321–332 (1969)
8. Kiefer, J.: Skorohod embedding of multivariate RV's, and the sample DF. Z. Wahrscheinlichkeitstheorie verw. Geb. **24**, 1–35 (1972)
9. Petrov, V.V.: On large deviations of sums of random variables. (Russian) Vestnik Leningrad. Univ. **16**, 25–37 (1961)
10. Petrov, V.V.: On sums of independent random variables. (Russian) Moscow, Nauka. 1972
11. Richter, W.: Local limit theorems for large deviations (Russian) Teor. Verojatnost. i Primenen **2**, 214–229 (1957)
12. Sawyer, S.: Rates of convergence for some functionals in probability. Ann. Math. Statist. **43**, 273–284 (1972)
13. Skorohod, A.V.: Studies in the theory of random processes. Reading, Mass: Addison-Wesley 1965
14. Strassen, V.: Almost sure behaviour of sums of independent random variables and martingales. Proc. 5th Berkeley Sympos. Math. Statist. Probab. (1965), Vol. II (part 1), 315–343. Berkeley: Univ. of Calif. Press 1967

J. Komlós
P. Major
G. Tusnády
Mathematical Institute of the
Hungarian Academy of Sciences
1053 Budapest
Hungary

(Received March 7, 1974)