The Annals of Probability
1977, Vol. 5, No. 6, 987–998

RECONSTRUCTING THE DISTRIBUTION FROM PARTIAL SUMS OF SAMPLES

BY G. HALÁSZ AND P. MAJOR

Mathematical Institute of the Hungarian Academy of Sciences

Let us observe an infinite sequence $z_1 = r_1 + \varepsilon_1$, $z_2 = r_2 + \varepsilon_2, \cdots$ where $r_1, r_2, \cdots$ are the partial sums of independent and identically distributed random variables and the sequence of random variables ε_k (the errors) is bounded by a function $f(k)$. Knowing the sequence z_n we want to determine the distribution function of the summands. We will show that this problem cannot be solved in general even if $f(k)$ is constant.

1. Introduction. Let a sequence of independent and identically distributed random variables $\xi_1, \xi_2, \cdots$ with an unknown distribution $F(x)$ be given. Denote the partial sums $\sum_{i=1}^{k} \xi_i$ by r_k, $k = 1, 2, \cdots$. We observe a sequence $r_k + \varepsilon_k$, $k = 1, 2, \cdots$, where ε_k is some error term. The error is bounded by a function $f(k)$ i.e., $\limsup \varepsilon_k/f(k) \leqq 1$ with probability 1, but we know nothing more about it. Can we recognize the unknown distribution $F(x)$? The answer is trivial if $f(k)$ tends to 0. In that case we know the individual terms ξ_i with an error tending to zero, thus the empirical distribution functions $F_n(x)$ based on the first n terms of the erroneous sample will tend to $F(x)$. Now what can be said about larger errors?

The first nontrivial answer was given by P. Bártfai who proved in [1] and [2] that if $\int \exp(tu)\, dF(u) < \infty$ for $|t| < t_0$ and $f(k) = o(\log k)$, then the unknown distribution can be recognized from one single realization with probability 1. On the other hand, Remark 1 of [3] implies that even if $F(x)$ is the standard normal distribution but $f(k) = \varepsilon \log k$, then on the basis of the sequence $r_k + \varepsilon_k$ one cannot decide whether $F(x)$ is normal or not.

In this paper we deal with the case when the existence of the moment generating function is not assumed. We substitute it by the following condition:

$$F(-x) + [1 - F(x)] < C_1 e^{-u(x)}, \quad x \geqq 0 . \tag{1}$$

We shall suppose throughout this paper that $u(x)$ is nonnegative, monotonically increasing, concave (i.e., $u'(x)$ is nonincreasing), $xu'(x) \nearrow \infty$ as $x \to \infty$, and $u'(x)$ is continuous. ($xu'(x) \to \infty$ implies that F has all moments.)

By a modification of the proof of Bártfai we prove

THEOREM 1. *If* $r_1, r_2, \cdots, s_1, s_2, \cdots$ *are sums of* i.i.d. rv's *with two different*

Received August 15, 1975.

AMS 1970 *subject classifications.* Primary 60G50, 62D05; Secondary 42A80.

Key words and phrases. Characteristic functions, mixing of distributions, partial sums of i.i.d. rv's, Cauchy's coefficient estimation.

distribution functions $F(x)$ *and* $G(x)$ *satisfying* (1) *with an* $u(x)$ *such that*

$$\int^{\infty} \frac{u(x)}{x^2}\, dx = \infty \tag{2}$$

then we have with probability 1 *with appropriate* $c > 0$ (c depending only on F and G)

$$|r_n - s_n| > \frac{\log n}{(\log \log n)^c}$$

for infinitely many n.

We do not know whether the $(\log \log n)^c$ can be omitted in this inequality. A slight modification of the proof would give the following

THEOREM 1a. *If* $F(x)$ *satisfies* (1) *with a function* $u(x)$ *satisfying* (2),

$$f(n) = o\left(\frac{\log n}{(\log \log n)^k}\right) \qquad \textit{for all} \quad k > 0$$

then $F(x)$ *can be reconstructed from the sequence* $r_k + \varepsilon_k$, $k = 1, 2, \cdots$ *with probability* 1.

Our Theorem 2 will show that the case $\int u(x)/x^2 < \infty$ is quite different.

THEOREM 2. *If*

$$\int^{\infty} \frac{u(x)}{x^2}\, dx < \infty$$

then we can find two distributions $F(x)$ *and* $G(x)$, $F(x) \not\equiv G(x)$ *and two samples* $\xi_1, \xi_2, \cdots$ *and* $\eta_1, \eta_2, \cdots$ *with distributions* F *and* G *in such a way that* $F(x)$ *and* $G(x)$ *satisfy* (1) *and the partial sums* $r_k = \sum_{i=1}^{k} \xi_i$, $s_k = \sum_{i=1}^{k} \eta_i$ *satisfy the relation*

$$|r_n - s_n| \leqq 1$$

for all n *with probability* 1.

We shall see that this theorem has the following

COROLLARY. *For any* $u(x)$, *for which* $\int u(x)/x^2\, dx < \infty$ *there is a distribution* $F(x)$ *satisfying* (1) *such that* $F(x)$ *cannot be recognized even if* $f(k) = C$ *whatever the positive constant* C *is.*

One may expect that these "hardly recognizable" distributions are pathological. We will show however that there are many pairs of "nice" distributions which satisfy Theorem 2.

Finally, we would like to mention that our investigations concern the situation when nothing is known about the structure of the error. If we know something of it, we may expect better results. The case when the error is caused by rounding off seems to have special interest. Therefore we quote the following result of P. Bártfai.

THEOREM OF BÁRTFAI [2]. *Let* $\mathscr{F}$ *be the class of distribution functions with an* *absolutely continuous density function of bounded variation. Let* $r_1, r_2, \cdots$ *be the partial sums of* i.i.d. rv's *with a distribution* $F \in \mathscr{F}$. *Then the unknown distribution* F *can be recognized from the sequence* $[r_k]$, $k = 1, 2, \cdots$ *with probability* 1. ($[\cdot]$ stands for integer part.)

The authors thank Mr. Tusnády for many useful discussions.

2. Proof of the theorems.

PROOF OF THEOREM 1. Suppose first that the characteristic functions $\varphi(t)$ and $\psi(t)$ of our distributions are infinitely differentiable satisfying

$$|a_\nu|, |b_\nu| \leqq D_\nu \qquad \nu = 1, 2, \cdots$$

where

$$\begin{matrix} a_\nu \\ b_\nu \end{matrix} = ((-i)^\nu/\nu!)\left[\log \begin{matrix}\varphi(t) \\ \psi(t)\end{matrix}\right]^{(\nu)}_{t=0}$$

(these are real numbers) and D_ν ($\geqq 1$) is logarithmically convex, i.e., $D_{\nu+1}/D_\nu$ is nondecreasing. Let

$$f_k(z) = \sum_{\nu=1}^{2k} a_\nu z^\nu .$$

This is to replace the logarithm of the moment generating function, the existence of which is not assumed here, and we try to approximate $\exp(lf_k(t))$ by

$$(1/m) \sum_{j=1}^{m} \exp(tR_j)$$

where

$$R_j = \sum_{(j-1)l<i\leqq jl} \xi_i \qquad j = 1, \cdots, m = [n/l] .$$

The integers k, l and the positive number t will be given later depending on n assumed to be large enough.

For $0 \leqq ex/m \leqq \frac{1}{2}$

$$\sum_{\nu=m}^{\infty} x^\nu/\nu! \leqq \sum_{\nu=m}^{\infty} (ex/\nu)^\nu \leqq 1$$

and we have

$$\exp(tR_j) = \sum_{\nu=0}^{k} (tR_j)^\nu/\nu! + O(1) ,$$

provided that $|tR_j| \leqq k/2e$. To estimate the probability of this event we use Markov's inequality

$$P(|R_j| > k/2et) \leqq (2et/k)^\mu E|R_j|^\mu .$$

Generally, putting

$$\exp(lf_k(z)) = \sum_{\mu=0}^{\infty} d_\mu z^\mu ,$$

$$d_\mu = (1/\mu!)ER_j{}^\mu \qquad 0 \leqq \mu \leqq 2k ,$$

the characteristic function of R_j being $\varphi^l(t)$. By Cauchy's coefficient estimation

$$|d_\mu| \leqq (1/r^\mu) \max_{|z|=r} |\exp(lf_k(z))| \leqq (1/r^\mu) \exp(l \textstyle\sum_{\nu=1}^{2k} D_\nu r^\nu)$$

for all $\mu = 0, 1, \cdots$. Generally, for fixed s and

$$0 \leqq t \leqq 1/2D_{2k}^{1/k} \leqq 1/2D_{2k}^{1/(2k-s-1)} \quad (k \geqq s+1)$$

$$\text{(3)} \qquad \sum_{\nu=s+1}^{2k} D_\nu t^\nu \leqq t^{s+1} \max_{s+1\leqq\nu\leqq 2k} D_\nu (2t)^{\nu-s-1} \sum_{\nu=0}^{\infty} (\tfrac{1}{2})^\nu \leqq c_2 t^{s+1} ,$$

since the maximum here is attained either for $\nu = s + 1$ or $\nu = 2k$, the sequence $D_\nu(2t)^{\nu-s-1}$ being logarithmically convex. Using this with $s = 0$ and

$$r = 1/2D_{2k}^{1/k} ,$$

$$|d_\mu| \leqq (1/r^\mu)\exp(c_2 lr) \qquad \mu = 0, 1, \cdots . \tag{4}$$

Hence

$$E|R_j|^\mu \leqq (\mu!/r^\mu)\exp(c_2 lr) \qquad \mu \text{ even, } 0 \leqq \mu \leqq 2k \tag{5}$$

and the Markov inequality gives

$$\begin{aligned} P(|R_j| > k/2et) &\leqq (2et/k)^\mu(\mu!/r^\mu)\exp(c_2 lr) \\ &\leqq \exp(c_2 lr - k) \end{aligned}$$

e.g., if $t \leqq 0.1r$ and $\mu = 2k$. Since j takes $m = [n/l] \leqq n$ values, we get

$$\exp(tR_j) = \textstyle\sum_{\nu=0}^{k}(tR_j)^\nu/\nu! + O(1) \tag{6}$$

for all $j = 1, \cdots, m$ simultaneously with the exception of an event of probability $\leqq n\exp(c_2 lr - k)$.

Now

$$\begin{aligned} E\textstyle\sum_{\nu=0}^{k}(tR_j)^\nu/\nu! &= \textstyle\sum_{\nu=0}^{k} d_\nu t^\nu = \exp(lf_k(t)) - \textstyle\sum_{\nu=k+1}^{\infty} d_\nu t^\nu \\ &= \exp(lf_k(t)) + O(1) \end{aligned}$$

for $k \geqq c_2 lr$, using again (4) and $t \leqq 0.1r$. By Cauchy's inequality and (5) we also have

$$\begin{aligned} E|\textstyle\sum_{\nu=0}^{k}(tR_j)^\nu/\nu!|^2 &\leqq (k+1)\textstyle\sum_{\nu=0}^{k}(t^{2\nu}/\nu!^2)E|R_j|^{2\nu} \\ &\leqq (k+1)\textstyle\sum_{\nu=0}^{k}(t^{2\nu}/\nu!^2)(2\nu)!\exp(c_2 lr)/r^{2\nu} \\ &\leqq (k+1)\exp(c_2 lr)\textstyle\sum_{\nu=0}^{\infty}(2\nu)!/\nu!^2 10^{2\nu} \leqq c_3 k\exp(c_2 lr) . \end{aligned}$$

We are now ready to apply Tchebycheff's inequality for

$$(1/m)\textstyle\sum_{j=1}^{m}(\sum_{\nu=0}^{k}(tR_j)^\nu/\nu!) .$$

The R_j's being independent, its variance is at most $c_3 k\exp(c_2 lr)/m$ which is thus also an upper bound for the probability that our quantity deviates by more than 1 from its expectation. Taking into account (6) at the same time, we have

$$(1/m)\textstyle\sum_{j=1}^{m}\exp(tR_j) = \exp(lf_k(t)) + O(1)$$

with probability at least $1 - n\exp(c_2 lr - k) - c_3 k\exp(c_2 lr)/m$. Choosing $k = [2\log n]$, $l = [k/5c_2 r]$ this probability is $\geqq 1 - 1/n^{\frac{1}{2}} \to 1$ as $n \to \infty$, if e.g.,

$$r = 1/2D_{2k}^{1/k} \geqq 1/k . \tag{7}$$

We proceed the same way with the η_i's, denoting by S_j and $g_k(z)$ quantities corresponding to R_j and $f_k(z)$, respectively. Let

$$M_n = \max_{1\leqq i\leqq n}|r_i - s_i| .$$

Then

$$|tR_j - tS_j| \leqq t2M_n \leqq 2M_n$$

and so

$$(1/m)\sum_{j=1}^{m}\exp(tR_j) = \exp(O(M_n))(1/m)\sum_{j=1}^{m}\exp(tS_j)\,.$$

Hence we obtain

$$\exp(lf_k(t)) + O(1) = \exp(lg_k(t) + O(M_n)) + O(\exp(O(M_n)))$$

and in case

$$lg_k(t) \geqq c_4(M_n + 1) \tag{8}$$

this yields

$$l(f_k(t) - g_k(t)) = O(M_n + 1)\,.$$

Suppose that $a_\nu \not\equiv b_\nu$ and let $\nu = s$ be the first index with $a_\nu \neq b_\nu$. Recalling $|a_\nu|, |b_\nu| \leqq D_\nu$ (3) shows that

$$|f_k(t) - g_k(t)| \geqq (|a_s - b_s|/2)t^s\,,$$

if we make the legitimate choice

$$t = c_5/D_{2k}^{1/k} \tag{9}$$

with c_5 sufficiently small, giving

$$M_n + 1 \geqq c_6 l t^s \geqq c_7 k/D_{2k}^{s-1/k} \tag{10}$$

with probability tending to 1. Our choice of D_ν will satisfy $D_{2k}^{1/k} \leqq \log^4 k$ for infinitely many k and so

$$M_n = \max_{1\leqq i\leqq n}|r_i - s_i| \geqq c_8 \log n/(\log\log n)^{4(s-1)}$$

a.s. for infinitely many n.

It remains to take care of the assumptions made in the course of the proof.

$D_{2k}^{1/k} \leqq \log^4 k$ for our k guarantees (7).

Without loss of generality we may suppose

$$b_1 = E\eta_i = 0 \qquad \text{and} \qquad b_2 = E(\eta_i - b_1)^2 > 0\,.$$

(3) with $s = 2$ shows that for c_5 in (9) sufficiently small

$$lg_k(t) \geqq c_9 l t^2$$

and if (8) did not hold, we would get (10) right away with $s = 2$.

Turning to a_ν and b_ν, we first estimate

$$\sup_{-\infty<t<\infty}|\varphi^{(\nu)}(t)| \leqq -\int_0^\infty x^\nu\, dP(|\xi_i| \geqq x) \leqq \nu \int_0^\infty x^{\nu-1} P(|\xi_i| \geqq x)\, dx\,,$$

and the same way for $\psi^{(\nu)}(t)$. Applying (1) and factoring out

$$x^{\nu-1}\exp(-u(x)/2) \leqq c_{10}\max_{x\geqq 0} x^\nu \exp(-u(x)/2)\,,$$

we get further

$$\leqq c_{10}\nu \max_{x\geqq 0} x^\nu \exp(-u(x)/2) \int_0^\infty \exp(-u(x)/2)\, dx\,.$$

Differentiation shows that the maximum is taken at $x = x(\nu) \nearrow +\infty$ defined by

$$\nu = xu'(x)/2$$

since by assumption $xu'(x) \nearrow +\infty$. This latter also implies $u(x)/\log x \to +\infty$ $(x \to \infty)$, hence the last integral is finite. Putting

$$D_\nu = x^\nu(\nu) \exp(-u(x(\nu))/2)/\nu^\nu$$

we thus have

$$\text{(11)} \qquad \sup_{-\infty<t<\infty} \frac{|\varphi^{(\nu)}(t)|}{|\psi^{(\nu)}(t)|} \leqq c_{11}\nu! \exp(2\nu)D_\nu .$$

The convexity of $\log D_\nu$ can be checked by considering ν as a continuous parameter and calculating

$$\frac{d \log D_\nu}{d\nu} = \log 2/u'(x(\nu)) - 1$$

which is in fact increasing since both $1/u'(x)$ and $x(\nu)$ are such. Hence $\nu!\, D_\nu$ is also logarithmically convex; and to infer from $\varphi(t) \not\equiv \psi(t)$ $a_k \not\equiv b_k$, or what is the same thing $\varphi^{(\nu)}(0) \not\equiv \psi^{(\nu)}(0)$, we can apply the Denjoy–Carleman theorem (see [5]) provided that

$$\text{(12)} \qquad \textstyle\sum_{\nu=1}^{\infty} \nu!\, D_\nu/(\nu+1)!\, D_{\nu+1} = \sum_{\nu=1}^{\infty} D_\nu/(\nu+1)D_{\nu+1} = +\infty .$$

But, D_ν being log-convex even for continuous ν, this is easily seen to be equivalent to

$$\begin{aligned} +\infty &= \textstyle\int^\infty \exp(-d \log D_\nu/d\nu)\, d\nu/\nu = (e/2) \int^\infty u'(x(\nu))\, d\nu/\nu \\ &= (e/2) \textstyle\int^\infty (u'(x)/xu'(x))\, d(xu'(x)) = (e/2) \int^\infty du'(x) + (e/2) \int^\infty u'(x)\, dx/x \\ &= O(1) + (e/2) \textstyle\int^\infty (u(x)/x^2)\, dx \end{aligned}$$

by partial integration and $u(x) = O(x)$.

Finally, from (11) (for $t = 0$) it follows the same way as in (3) that the polynomial

$$\pi(z) = \textstyle\sum_{\mu=1}^{\nu} \varphi^{(\mu)}(0) z^\mu/\mu! \qquad \text{or} \qquad \sum_{\mu=1}^{\nu} \psi^{(\mu)}(0) z^\mu/\mu!$$

satisfies

$$|\pi(z)| \leqq \tfrac{1}{2}$$

for

$$|z| = r = c_{12} \min(1/D_\nu^{1/\nu}, 1)$$

and as $\pi^{(\mu)}(0) = \varphi^{(\mu)}(0)$ and $\psi^{(\mu)}(0)$ $(\mu = 1, 2, \cdots, \nu)$, respectively, we get

$$|a_\nu| \quad \text{or} \quad |b_\nu| = |[\log(1 + \pi(z))]_{z=0}^{(\nu)}|/\nu! \leqq \log 2/r^\nu \leqq c_{13}^\nu D_\nu$$

by Cauchy's coefficient estimation, where the right-hand side can be made increasing with c_{13} large enough and this can play the role of D_ν in the proof. (12) and log-convexity implies

$$D_{\nu+1}/D_\nu \leqq D_{2k}/D_{2k-1} \leqq \log^2 k \quad (\nu \leqq 2k - 1)$$

for infinitely many k and by multiplying these $D_{2k}^{1/k} \leqq c_{14} \log^4 k$ and the proof is completed.

(The above method shows that under the conditions of Theorem 1a the a_ν's can be determined with probability 1. This remark proves Theorem 1a.)

In preparation for the next proof we remark that the Denjoy–Carleman theorem is sharp: If

$$\sum_{\nu=1}^{\infty} D_\nu/(\nu+1)D_{\nu+1} < +\infty ,$$

which we have seen to be equivalent to

$$\int^{\infty} (u(x)/x^2)\, dx < +\infty ,$$

then for any interval (α, β) we can find a function $h(t)$ vanishing outside (α, β) but not identically such that

$$\sup_{-\infty<t<\infty} |h^{(\nu)}(t)| \leqq \nu!\, D_\nu$$

([5]). If $(\alpha, \beta) \subset [-\pi, \pi]$, then, integrating by parts ν times, we get for the nth Fourier coefficient

$$\begin{aligned} |(1/2\pi) \textstyle\int_\alpha^\beta \exp(-int)h(t)\, dt| &\leqq (1/2\pi|n|^\nu) \textstyle\int_\alpha^\beta |h^{(\nu)}(t)|\, dt \\ &\leqq \nu!\, D_\nu/|n|^\nu = \nu!\, x^\nu(\nu) \exp(-u(x(\nu))/2)/|n|^\nu \nu^\nu \\ &\leqq c_{15} \exp(-u(|n|/2)/2) , \end{aligned}$$

choosing for ν an integer with $|n|/2 \leqq x(\nu) \leqq |n|$. This is possible if the variation of $xu'(x)/2$ is at least 1 over the interval $[|n|/2, |n|]$, which, e.g., by adding $x^{\frac{1}{2}}$ to $u(x)$, can always be achieved. Replacing $u(x)$ by $\alpha u(\beta x)$, the $\frac{1}{2}$'s can, of course, be replaced by and number.

PROOF OF THEOREM 2. The idea is based on the (known) fact that there exists a random variable γ that can be represented in two ways as the sum of two independent identically distributed random variables, $\gamma = \xi_1 + \xi_2 = \eta_1 + \eta_2$, say, the two distributions being different. If for $(\xi_{2l-1}, \xi_{2l}, \eta_{2l-1}, \eta_{2l})$ $(l = 1, 2, \cdots)$ we choose independent replicas of the quadruple $(\xi_1, \xi_2, \eta_1, \eta_2)$ then we have defined our sequences with $r_{2n} = s_{2n}$. If $\xi_1 - \eta_1$ can even be made bounded, then $r_n - s_n$ will also be bounded.

All our variables will take integer values with probabilities

$$P(\xi_1 = k) = P(\xi_2 = k) = p_k ,$$
$$P(\eta_1 = k) = P(\eta_2 = k) = q_k .$$

We must have

$$P(\gamma = m) = \sum_{k=-\infty}^{\infty} p_k p_{m-k} = \sum_{k=-\infty}^{\infty} q_k q_{m-k} . \tag{13}$$

p_k and q_k determine the conditional distributions of ξ_1 and η_1,

$$P_{n,m} = P(\xi_1 \leqq n \mid \gamma = m) = \sum_{k=-\infty}^{n} p_k p_{m-k}/P(\gamma = m) ,$$
$$Q_{n,m} = P(\eta_1 \leqq n \mid \gamma = m) = \sum_{k=-\infty}^{n} q_k q_{m-k}/P(\gamma = m) ,$$

but we are at liberty to prescribe their joint conditional distributions, e.g., as

$$(14)\qquad P(\xi_1 = i, \eta_1 = j \mid \gamma = m) = \text{length of the interval } [P_{i-1,m}, P_{i,m}] \cap [Q_{j-1,m}, Q_{j,m}]\,,$$

(This is the so called quantile transformation that has already played an essential role in [2] and [3], [4].) If for each n, m

$$Q_{n-1,m} \leqq P_{n,m} \leqq Q_{n+1,m}$$

or equivalently

$$(15)\qquad -q_n q_{m-n} \leqq \sum_{k=-\infty}^{n} (p_k p_{m-k} - q_k q_{m-k}) \leqq q_{n+1} q_{m-n-1}\,,$$

then (14) will be zero unless $i - j = 1, 0$ or -1, i.e., $P(|\xi_1 - \eta_1| \leqq 1) = 1$ and defining $\xi_2 = \gamma - \xi_1$, $\eta_2 = \gamma - \eta_1$ the proof will be completed if we can choose p_k and q_k with the above properties.

In terms of the characteristic functions

$$\varphi(t) = \sum_{k=-\infty}^{\infty} p_k \exp(ikt)\,, \qquad \psi(t) = \sum_{k=-\infty}^{\infty} q_k \exp(ikt) \qquad -\pi \leqq t \leqq \pi$$

condition (13) takes the form $\varphi^2(t) \equiv \psi^2(t)$. This will be fulfilled beforehand if we seek our functions as

$$\varphi(t) = f(t) + g(t)\,, \qquad \varphi(t) = f(t) - g(t)\,,$$
$$f(t) = \sum_{k=-\infty}^{\infty} a_k \exp(ikt) \quad (\text{with } f(0) = 1) \qquad \text{and}$$
$$g(t) = \sum_{k=-\infty}^{\infty} b_k \exp(ikt)$$

having disjoint support.

According to the remark made after the proof of Theorem 1 there is a Fourier series

$$h_1(t) = \sum_{k=-\infty}^{\infty} d_k \exp(ikt)$$

vanishing for $\pi/4 \leqq |t| \leqq \pi$ but not identically (we may suppose $d_0 \neq 0$) such that

$$d_k = O(\exp(-u(|k|)))\,.$$

The convolution

$$h_2(t) = (1/2\pi) \int_{-\pi}^{\pi} h_1(t+z)\overline{h_1(z)}\, dz$$

has nonnegative coefficients $|d_k|^2$ and vanishes for $\pi/2 \leqq |t| \leqq \pi$. We define $f(t)$ by

$$a_k = \sum_{l=-\infty}^{\infty} |d_l|^2 \exp(-u(|k-l|))$$

where we can assume $\sum_{k=-\infty}^{\infty} a_k = 1$. The property of vanishing for $\pi/2 \leqq |t| \leqq \pi$ is preserved since

$$f(t) = h_2(t) \sum_{k=-\infty}^{\infty} \exp(-u(|k|) + ikt)\,,$$

but we also have

$$a_k \geqq |d_0|^2 \exp(-u(|k|))$$

and since $u(|k-l|) \geqq u(|k|) - u(|l|)$, $u(x)$ being increasing and concave,

$$(16)\qquad a_k \leqq c_{16} \sum_{l=-\infty}^{\infty} \exp(-2u(|l|) - u(|k-l|)) \leqq c_{16} \exp(-u(|k|)) \sum_{l=-\infty}^{\infty} \exp(-u(|l|))$$

and so

$$c_{17}\exp(-u(|k|)) \leqq a_k \leqq c_{18}\exp(-u(|k|)) . \tag{17}$$

Let now

$$h_3(t) = \sum_{k=-\infty}^{\infty} e_k \exp(ikt)$$

be a function vanishing for $-\pi \leqq t \leqq \pi/2$ but not identically with

$$e_k = O(\exp(-4u(|k|)))$$

and put

$$g(t) = \delta(h_3(t) + \overline{h_3(-t)}) \quad (\not\equiv 0) .$$

(δ will be fixed later.) This is zero for $|t| \leqq \pi/2$ and has real coefficients satisfying

$$|b_k| \leqq c_{19}\delta \exp(-4u(|k|)) .$$

From (17) it follows that

$$0 < c_{20}\exp(-u(|k|)) \leqq \begin{matrix} p_k = a_k + b_k \\ q_k = a_k - b_k \end{matrix} \leqq c_{21}\exp(-u(|k|)) \tag{18}$$

for δ small enough.

Turning to condition (15) we see from the last estimations and the fact $|u(|k+1|) - u(|k|)| \leqq u(1)$ that it is equivalent to

$$\begin{aligned} |\textstyle\sum_{k=-\infty}^{n} (b_k a_{m-k} + b_{m-k} a_k)| &\leqq |\textstyle\sum_{k=-\infty}^{n} b_k a_{m-k}| + |\textstyle\sum_{k=m-n}^{\infty} b_k a_{m-k}| \\ &\leqq c_{22}\exp(-u(|n|) - u(|m-n|)) . \end{aligned} \tag{19}$$

Note that, $f(t)$ and $g(t)$ having disjoint support, $f(t)g(t) \equiv 0$, hence

$$\sum_{k=-\infty}^{\infty} b_k a_{m-k} = 0 .$$

This means that we can replace, in the case of the first sum, $\sum_{k=-\infty}^{n}$ by $\sum_{n+1}^{\infty}$. We do so if $n \geqq 0$ but leave it in the original form if $n < 0$. In any case

$$\begin{aligned} |\textstyle\sum_{k=-\infty}^{n} b_k a_{m-k}| &\leqq c_{23}\delta \textstyle\sum_{|k|\geqq n} \exp(-4u(|k|) - u(|m-k|)) \\ &\leqq c_{23}\delta \exp(-2u(|n|)) \textstyle\sum_{k=-\infty}^{\infty} \exp(-2u(|k|) - u((m-k))) , \end{aligned}$$

and as in (16) we get further

$$\leqq c_{24}\delta\exp(-2u(|n|) - u(|m|)) \leqq c_{24}\delta\exp(-u(|n|) - u(|m-n|))$$

using again $u(|m|) \geqq u(|m-n|) - u(|n|)$.

The second sum in (19) coincides with the first for n replaced by $m - n - 1$ and we see that (19) holds for δ sufficiently small. From (18) we get for the tail probabilities the upper bound

$$\begin{aligned} c_{21}\textstyle\sum_{|k|\geqq x} \exp(-u(|k|)) &\leqq c_{21}\exp(-u(x)/2) \textstyle\sum_{k=-\infty}^{\infty} \exp(-u(|k|)/2) \\ &\leqq c_{25}\exp(-u(x)/2) \end{aligned}$$

which is the statement with $u(x)/2$ in place of $u(x)$. □

3. Consequences of the theorems and problems. To prove the corollary we use a lemma. This lemma has already been proved in [4] (Lemma 1) but for the sake of completeness we give the proof here again.

LEMMA. *We are given distribution functions* $F_1(x)$, $F_2(x)$ *and* $G_1(x)$, $G_2(x)$ *and sums of* i.i.d. rv's $r_1^{(i)}, r_2^{(i)}, \cdots, s_1^{(i)}, s_2^{(i)} \cdots i = 1, 2$ *such that* $r_1^{(i)}$ *has distribution* $F_i(x)$ *and* $s_1^{(i)}$ *has distribution* $G_i(x)$, *and moreover*

$$P(|s_k^{(1)} - r_k^{(1)}| < a \text{ for all } k) = 1 ,$$
$$P(|s_k^{(2)} - r_k^{(2)}| < b \text{ for all } k) = 1 .$$

Let p be an arbitrary number $O \leqq p \leqq 1$ *and*

$$F(x) = pF_1(x) + (1-p)F_2(x) ,$$
$$G(x) = pG_1(x) + (1-p)G_2(x) .$$

Then there exist two sequences of sums of i.i.d. rv's $r_1, r_2, \cdots s_1, s_2, \cdots$ *such that* r_1 *has distribution* $F(x)$, s_1 *has distribution* $G(x)$, *and* $P(|s_k - r_k| < a + b$ *for all* $k) = 1$.

PROOF. We may suppose that the two sequences of pairs $\{r_k^{(1)}, s_k^{(1)}\}_{k=1}^\infty$ and $\{r_k^{(2)}, s_k^{(2)}\}_{k=1}^\infty$ are independent. Define a sequence of i.i.d. rv's $\varepsilon_1, \varepsilon_2, \cdots$ with distribution $P(\varepsilon_1 = 1) = 1 - P(\varepsilon_1 = 0) = p$ such that the sequence is independent of all the variables $r_k^{(1)}, s_k^{(1)}, r_k^{(2)}, s_k^{(2)}$.

Set $\tau(n) = \sum_{i=1}^n \varepsilon_i$, and

$$r_n = r_{\tau(n)}^{(1)} + r_{n-\tau(n)}^{(2)}$$
$$s_n = s_{\tau(n)}^{(1)} + s_{n-\tau(n)}^{(2)} .$$

Then the sequence r_n has the prescribed joint distribution, and the same holds for s_n. Furthermore,

$$|s_n - r_n| \leqq \sup_k |s_k^{(1)} - r_k^{(1)}| + \sup_k |s_k^{(2)} - r_k^{(2)}| < a + b .$$

PROOF OF THE COROLLARY. Applying Theorem 2 with $u(2^i x)$ instead of $u(x)$ we obtain two distributions $F_i(x)$, $G_i(x)$ and sums of i.i.d. rv's $r_1^{(i)}, r_2^{(i)}, \cdots$, $s_1^{(i)}, s_2^{(i)}, \cdots$ with distribution $F_i(x)$ resp. $G_i(x)$ such that

$$P\left(|r_n^{(i)} - s_n^{(i)}| < \frac{1}{2^i} \text{ for all } n\right) = 1$$

and $F_i(x)$, $G_i(x)$ satisfy (1).

Let $\sum_{i=1}^\infty p_i = 1$, $p_i > 0$, and let p_i tend to zero fast enough. Define $F(x) = \sum_{j=1}^\infty p_j F_j(x)$, $\bar{G}_i(x) = \sum_{j\neq i} p_j F_j(x) + p_i G_i(x)$. Then $F(x)$ and $\bar{G}_i(x)$ satisfy (1) and they cannot be distinguished if the error may exceed $1/2^{i+1}$, because the lemma enables us to construct two sums of i.i.d. rv's $r_1, r_2, \cdots, \bar{s}_1^{(i)}, \bar{s}_2^{(i)}, \cdots$ with distributions $F(x)$ and $\bar{G}_i(x)$ in such a way that

$$|r_n - \bar{s}_n^{(i)}| < \frac{1}{2^i} . \qquad \square$$

Our lemma enables us to construct a much wider class of distributions satisfying Theorem 2. Namely we have

THEOREM 3. *Let* $\int^{\infty} (u(x)/x^2)\,dx < \infty$, *and let the distribution function* $F(x)$ *have a density* $f(x)$ *satisfying the inequality* $f(|x|) \geqq Ce^{-u(x)}$, $x \geqq 0$. *Then there exist a distribution function* $G(x)$ *and two sums of* i.i.d. rv's $r_1, r_2, \cdots, s_1, s_2, \cdots$ *with distribution* $F(x)$ *and* $G(x)$ *in such a way that*

$$P(|r_n - s_n| \leqq 1 \ \textit{for all} \ \ n) = 1 .$$

PROOF. Let $F_1(x)$ and $G_1(x)$ be two distributions constructed in Theorem 2. Consider the convolutions

$$F_2(x) = F_1(x) * I(x)$$

and

$$G_2(x) = G_1(x) * I(x)$$

($I(x)$ denotes the uniform distribution on [0, 1].) Then $F_2(x)$ and $G_2(x)$, too, satisfy Theorem 2 and $F_2(x)$ has a density function $f_2(x)$, $f_2(x) \leqq C'e^{-u(|x|)}$. We can write $f(x)$ in the form

$$f(x) = \alpha f_2(x) + (1 - \alpha)h(x)$$

where $0 < \alpha \leqq 1$ and $h(x)$ is the density of a distribution function $H(x)$. Because of the lemma the distributions

$$F(x) = \alpha F_2(x) + (1 - \alpha)H(x)$$
$$G(x) = \alpha G_2(x) + (1 - \alpha)H(x)$$

satisfy Theorem 3.

One would also like to characterise the pairs of distributions $F(x)$, $G(x)$ satisfying Theorem 2. In all our previous examples the characteristic functions of these distributions $F(x)$ and $G(x)$ coincided in a neighbourhood of the origin. Using a slight generalization of the lemma we can get rid of this restriction. Another question is whether the moments of these distributions may differ. We can prove that if $F(x)$ and $G(x)$ have $4 + \varepsilon$th moment $\varepsilon > 0$, then the third moments must be the same. If the tails of the distributions satisfy (1) with $u(x) = x^{\frac{1}{2}+\varepsilon}$, then even the fourth moments must agree, etc. But we cannot answer the following question. Let $F(x)$ and $G(x)$ have all moments and assume that they satisfy Theorem 2. Are all the moments of $F(x)$ and $G(x)$ necessarily equal?

REFERENCES

[1] BÁRTFAI, P. (1966). 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.

[2] BÁRTFAI, P. Ph. D. thesis (in Hungarian).

[3] KOMLÓS, J. MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's and the sample df I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131.

[4] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent rv's and the sample df II. *Z. Wahrscheinlichkeitstheorie und. Verw. Gebiete* **34** 33–58.

[5] RUDIN, W. (1970). *Real and Complex Analysis.* McGraw-Hill, New York.

MATHTEMATICAL INSTITUTE OF
THE HUNGARIAN ACAD. OF SCIENCES
H-1053 BUDAPEST V.,
REALTANODA U. 13-15
HUNGARY