The proof of the central limit theorem and Fourier analysis I.

by Péter Major
The Mathematical Institute of the Hungarian Academy of Sciences

Summary: The central limit theorem belongs to the most important results of probability theory. There is another classical result of the probability theory, the law of large numbers which states in a slightly informal language that under some mild conditions the average of the partial sums of independent and identically distributed random variables can be well approximated by the expected value of these random variables. The central limit theorem gives better, more substantial information about the fluctuation of the average of these partial sums around their expected value. It states that if we multiply this fluctuation by $\sqrt{n}$, where n denotes the number of random variables in the average, then the distribution of the fluctuation with this normalization is close to a (non-degenerated) distribution independent of the number n. Moreover, and this is a most remarkable fact, this approximating distribution does not depend on the distribution of the random variables whose average is taken. This distribution is a "universal law" which is called the normal distribution in the literature. In the series of problems described in this text we shall formulate the above result in a more precise and more general form.

The method of the proofs also deserves special attention. This method, which is called the characteristic function method, is actually the application of the Fourier analysis in the proof of limit theorems. One goal of this series of problems was to show that this is a natural approach. The main difference between my text and the discussion of the central limit theorem in usual text-books is that I tried to supply a more definitive and detailed explanation about the relation between the method of the proof of the central limit theorem and some basic ideas of the Fourier analysis.

The theory of the Fourier series gives a natural method to prove the local version of the central limit theorem, i.e. to describe the asymptotic behaviour of the density and not of the distribution function of the normalized partial sums of independent random variables. Furthermore, if we understand what kind of additional problems arise if we want to prove the central limit theorem in its original (and not in its local) form and also understand the answers supplied by the results of the Fourier analysis to these questions, then we can prove the desired results.

Moreover, in such a way we get a natural method to prove some refinements of the central limit theorem. We also can investigate the question how good approximation is supplied by the central limit theorem for the distribution of normalized partial sums of indepen-

dent random variables, and what kind of better approximation can be given. But these problems are not discussed here. They will be the subject of the second part of this series of problems.

Finally, I mention some parts of this work which may be of special interest. They are the proof of the Stirling formula in a way suggested by the proof of the local central limit theorem, (problem 2), the necessary and sufficient condition of the relative compactness of a sequence of probability measures expressed by means of their characteristic functions (problem 22), a detailed discussion of the question how the properties of a function or a measure are reflected in the behaviour of their Fourier transform (problems 27–34 and the discussion about these problems).

CONTENT

A.) Local limit theorems.

Let us first consider the following problem: Let ξ_j, $j=1,2,\dots$, be a sequence of independent and identically distributed random variables which take integer values. We introduce the notation $P(\xi_1=k)=p(k)$, $k=0,\pm1,\pm2,\dots$, $\sum\limits_{k=-\infty}^{\infty} p(k)=1$ and define the partial sums $S_n=\sum\limits_{j=1}^{n}\xi_j$, $n=1,2,\dots$. Let us consider the probabilities $p_n(k)=P(S_n=k)$, $k=0,\pm1,\pm2,\dots$, $n=1,2,\dots$, and try to give a good approximation of the probabilities $p_n(k)$ in the case of large parameters n.

A good estimation can be given for these probabilities $p_n(k)$ by means of the following method. Let us define the Fourier series

$$P_n(t)=\sum_{k=-\infty}^{\infty} p_n(k)e^{ikt}, \quad -\pi\le t\le\pi. \tag{1}$$

By a basic formula of the theory of Fourier series the coefficients of the above Fourier series can be expressed by means of the formula

$$p_n(k)=\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}P_n(t)\,dt, \quad k=0,\pm1,\pm2,\dots. \tag{2}$$

Hence, if we give a good asymptotic formula for the Fourier series $P_n(t)$ and estimate well the integral in formula (2), then we get a good estimate for the probabilities $p_n(k)$ we are interested in. Let us observe that $P_n(t)=Ee^{itS_n}$, $-\pi\le t\le\pi$. Besides, as S_n is the sum of independent and identically distributed random variables,

$$P_n(t)=Ee^{it(\xi_1+\cdots+\xi_n)}=\left(Ee^{it\xi_1}\right)^n=(P_1(t))^n,$$

where $P_1(t)=\sum\limits_{k=-\infty}^{\infty} P(\xi_1=k)e^{ikt}$. Furthermore, $P_1(0)=1$, and as we shall see, $|P_1(t)|<1$ if $-\pi\le t\le\pi$ and $t\ne0$ under some natural conditions. Hence in the estimation of the probabilities $p_k(n)$ a so-called singular integral appears in formula (2), in which only a small neighbourhood of the origin gives an essential contribution to the integral. Such expressions can be well bounded by some standard methods of the analysis. Let us first show the following identities which we later need.

1.) Let us show that

$$\frac{1}{\sqrt{2a\pi}}\int_{-\infty}^{\infty} e^{-u^2/2a}\,du=1$$

for all real positive numbers $a>0$. Furthermore,

$$\frac{1}{\sqrt{2a\pi}}\int_{-\infty}^{\infty} e^{-(u-z)^2/2a}\,du=1$$

for all real positive numbers $a>0$ and *complex* numbers z.

First we apply the above method in a special case. Let us consider the Poisson distribution with parameter $\lambda=n$, that is the distribution of a random variable η for which $P(\eta=k)=P_n(k)=\frac{n^k}{k!}e^{-n}$, $k=0,1,2,\dots$. The distribution of a Poissonian random variable with parameter n agrees with the distribution of the sum of n independent Poissonian random variables with parameter 1. Hence the above sketched method makes possible to get a good estimate for the probabilities $P_n(k)$. We get a really good estimate in the case when the number k is close to the expected value of the random variable η, i.e. if $n\sim k$. In particular, by investigating the appropriate Fourier series we get a good bound on the number $P_n(n)$. We do not exploit explicitly the probabilistic content of the coefficients $P_n(k)$ of this Fourier series, but the calculations indicate what kind of technical problems has to be solved in the analogous problems of probability theory. Besides, we get in such a way the proof of an important result of the analysis, the proof of the Stirling formula. This is the content of the next problem.

2.) Let us calculate the value of the Poissonian distribution with parameter n in the point n with the help of the above discussed method by working with the Fourier series whose coefficients are the values of this Poisson distribution. Let us show with the help of this method that

$$n!=\left(\frac{n}{e}\right)^n\frac{2\pi}{\int_{-\pi}^{\pi}e^{n(e^{it}-1-it)}\,dt}. \tag{3}$$

Let us show that

$$\int_{-\pi}^{\pi}e^{n(e^{it}-1-it)}\,dt=\frac{\sqrt{2\pi}}{\sqrt{n}}\left(1+O\left(\frac{1}{n}\right)\right) \tag{4a}$$

and

$$n!=\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\left(1+O\left(\frac{1}{n}\right)\right). \tag{4b}$$

Let us prove the following improvement of the previous two statements:

$$\int_{-\pi}^{\pi}e^{n(e^{it}-1-it)}\,dt=\frac{\sqrt{2\pi}}{\sqrt{n}}\left(1+\frac{c_1}{n^{1/2}}+\frac{c_2}{n}+\cdots+\frac{c_k}{n^{k/2}}+O\left(\frac{1}{n^{(k+1)/2}}\right)\right) \tag{4c}$$

for arbitrary integer $k\ge1$ with explicitly calculable coefficients c_k. In particular, $c_1=0$. Furthermore,

$$n!=\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\left(1+\frac{\bar c_1}{n^{1/2}}+\frac{\bar c_2}{n}+\cdots+\frac{\bar c_k}{n^{k/2}}+O\left(\frac{1}{n^{(k+1)/2}}\right)\right) \tag{4d}$$

for arbitrary integer $k\ge1$ with explicitly calculable coefficients $\bar c_k$. In particular, $\bar c_1=0$.

Similarly to the solution of problem 2 a good asymptotic formula can be given for the probability of the event that the sum of independent and identically distributed random variables takes a definite value. This investigation is based on formula (2) where the integral of a Fourier series with periodicity 2π has to be estimated. But such cases may appear when this Fourier series also has a smaller periodicity. Such an example appears for instance if the partial sums of such random variables are considered which take only even or only odd values. The Fourier series determined by the distribution of such random variables has a smaller periodicity π. If we want to study the asymptotic behaviour of the distribution of partial sums of independent and identically distributed random variables in a systematic way, then we have to clarify which is the smallest period of the Fourier series in formula (2). In particular, we want to tell when the smallest periodicity of this Fourier series is precisely 2π. This is the reason why we introduce the following definition. Then we study the case when the smallest periodicity of the Fourier series in formula (2) is 2π. The investigation of partial sums of independent and identically distributed random variables whose distribution takes values on a lattice $kh+b$, $k=0,\pm1,\pm2,\dots$, with some numbers h and b can be relatively simply reduced to this special case.

Definition A. *The values of a random variable ξ are concentrated on the lattice of integers (as on the rarest lattice) if $\sum\limits_{k=-\infty}^{\infty} P(\xi=k)=1$, and for arbitrary integers $A>1$ and B $\sum\limits_{k=-\infty}^{\infty} P(\xi=Ak+B)<1$.*

More generally, we call a random variable ξ lattice valued if its values are concentrated with probability one on a set of the form $\{b+kh,\colon\ k=0,\pm1,\pm2,\dots\}$ with some real numbers $h>0$ and b. We say that the values of a random variable ξ are concentrated on a lattice of width h, $h>0$, (as on the rarest lattice) if there exists a real number b such that $\sum\limits_{k=-\infty}^{\infty} P(\xi=kh+b)=1$, and $\sum\limits_{k=-\infty}^{\infty} P(\xi=Akh+B)<1$ for all integers $A>1$ and real numbers B.

To carry out the further investigation we need the following result.

3.) If ξ is a non-constant random variable concentrated on a lattice, then there exist a smallest $h>0$ number such that it is concentrated on a lattice of width h as the rarest lattice.

Let ξ be a random variable concentrated on a lattice of width h (as on the rarest lattice). Let us choose a real number $b>0$ such that $\sum\limits_{n=-\infty}^{\infty} P(\xi=nh+b)=1$, and let us consider the Fourier series $P(t)=\sum\limits_{n=-\infty}^{\infty} e^{inh}P(\xi-b=nh)$ determined by the distribution of the random variable $\xi-b$. The periodicity of the Fourier series $P(t)$ equals $\frac{2\pi}{h}$, $P(0)=1$, $|P(t)|\le1$ for all real numbers t, and $|P(t)|<1$ if $|t|\le\frac{\pi}{h}$ and $t\neq0$. If the absolute value of the random variable $\xi-b$ has k-th moment, that is $E|\xi-b|^k<\infty$, then the function $P(t)$ is k times continuously differentiable, and

$\left.\frac{dP^k(t)}{dt^k}\right|_{t=0} = i^k E(\xi - b)^k$, (where $i = \sqrt{-1}$).

4.) Let $\xi_1, \xi_2, \dots,$ be a sequence of independent and identically distributed random variables whose values are concentrated with probability one on the lattice of the integers (as on the rarest lattice). Let $E\xi_1 = m$, $E\xi_1^2 = m_2 < \infty$, (that is we assume that the second moment of the random variable ξ_1 is finite), and put $\sigma^2 = m_2 - m_1^2$. (The number σ^2 denotes the variance of the random variable ξ_1.) Let us consider the partial sums $S_n = \xi_1 + \dots + \xi_n$, $n = 1, 2, \dots$. Then

$$P(S_n = k) = \frac{1}{\sqrt{2\pi n}\sigma} \exp\left\{-\frac{(k-nm)^2}{2n\sigma^2}\right\} + o\left(\frac{1}{\sqrt{n}}\right), \quad k = 0, \pm 1, \pm 2, \dots,$$

and $o(\cdot)$ is uniform in the variable k.

5.) Let us consider such a sequence $\xi_1, \xi_2, \dots,$ of independent and identically distributed random variables which satisfies the conditions of the previous problem together with the condition $E|\xi_1|^3 < \infty$. Under these conditions prove the following sharper form of the previous problem.

$$P(S_n = k) = \frac{1}{\sqrt{2\pi n}\sigma} \exp\left\{-\frac{(k-nm)^2}{2n\sigma^2}\right\} + \varepsilon(n,k), \quad k = 0, \pm 1, \pm 2, \dots,$$

where $|\varepsilon(n,k)| \le \frac{K}{n}$, and the constant K depends only on the distribution of the random variable ξ_1.

Historically first that special case of the above limit theorems was considered where the random variable ξ_1 is binomially distributed, i.e. $P(\xi_1 = 1) = 1 - P(\xi_1 = 0) = p$, $0 < p < 1$. In this case the distribution of the sum S_n can be expressed in a simple explicit form, and this expression can be well bounded with the help of the Stirling formula. As this special case has a particular importance in combinatorial applications, it deserves special discussion. This is done in the following problem.

5a.) Let us give an elementary proof of the statement of problem 5 in the special case when ξ_1 is binomially distributed. (Here we want to give a proof where the Stirling formula is applied instead of the Fourier analysis.)

6.) Let a sequence $\xi_1, \xi_2, \dots$ of independent and identically distributed random variables satisfy the conditions of problem 4 with the difference that now we assume that the values of the random variable ξ_1 are concentrated with probability 1 on a lattice of width h, $h > 0$, (which is the width of the rarest lattice). Let the values of the random variable ξ_1 be concentrated on the numbers $kh + b$, $k = \pm 1, \pm 2, \dots,$ with some real number b. Then with the notations of problem 4

$$P(S_n = kh + nb) = \frac{h}{\sqrt{2\pi n}\sigma} \exp\left\{-\frac{(kh+nb-nm)^2}{2n\sigma^2}\right\} + o\left(\frac{1}{\sqrt{n}}\right), \tag{5}$$
$$k = 0, \pm 1, \pm 2, \dots,$$

where $o(\cdot)$ is uniform in the variable k. If also the condition $E|\xi_1|^3<\infty$ is satisfied, then

$$P(S_n=kh+nb)=\frac{h}{\sqrt{2\pi n}\sigma}\exp\left\{-\frac{(kh+nb-nm)^2}{2n\sigma^2}\right\}+\varepsilon(k,n),$$
$$k=0,\pm1,\pm2,\dots,$$

where $|\varepsilon(n,k)|\le\frac{K}{n}$, and the constant K depends only on the distribution of the random variable ξ_1.

7.) If a sequence of random variables S_n, $n=1,2,\dots$, satisfies relation (5), (it has no importance what additional properties these random variables have), then

$$\lim_{n\to\infty}P\left(\frac{S_n-nm}{\sqrt{n}\sigma}<x\right)=\int_{-\infty}^x\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du$$

for all real numbers x. The convergence is uniform with respect to the parameter x.

The above results supply good information on the probabilities the partial sums of lattice valued independent and identically distributed random variables take different values. Then by "integrating out" this relation in problem 7 we have obtained an estimate about the limit behaviour of the distribution of the appropriately normalized partial sums of independent and identically distributed random variables if the summands are lattice distributed. These results can be sharpened, but first we want to generalize the above result and to investigate the limit behaviour of the distribution or density (if the density function exists) function of appropriately normalized partial sums of independent random variables which are not necessarily lattice distributed.

Let $\xi_1,\xi_2,\dots$, be a sequence of independent random variables, and let us consider the partial sums $S_n=\xi_1+\dots+\xi_n$, $n=1,2,\dots$. We want to show that if the random variable ξ_1 has a nice density function, then a good asymptotic formula can be given for the density function of the random sum S_n. Besides, we want to show under general conditions that the appropriate normalized version of the sums have a limit distribution as $n\to\infty$, and also want to describe this limit distribution function. In the case when the random variables ξ_n are lattice valued then the corresponding problems could be solved by means of the investigation of an appropriate Fourier series. The question arises whether this method can be adapted to solve the analogous problem in the general case. To carry out such a program we should have a relatively simple inversion formula which enables to calculate a density or distribution function by means of its Fourier transform.

In case of nice density functions there exists such a simple inversion formula. In case of general distribution function there are only complicated inversion formulas which are not really well applicable in the investigations we have in mind. To prove limit theorems for the appropriately normalized partial sums of independent random variables first we have to understand what means exactly convergence in distribution. Then we can prove limit theorems for the normalized partial sums of independent random variables by means of some basic results of the Fourier analysis. Before doing this we formulate an inversion formula for the Fourier transform which is useful in the investigation of the density functions.

Inversion formula for Fourier transforms. *Let $f(u)$ be an integrable function on the real line, i.e. let us assume that $\int_{-\infty}^{\infty} |f(u)|\,du < \infty$. Let us consider the Fourier transform*

$$\tilde f(t) = \int_{-\infty}^{\infty} e^{itu} f(u)\,du, \quad -\infty < t < \infty \tag{6a}$$

of this function $f(\cdot)$. If the function $\tilde f$ is also integrable, i.e. $\int_{-\infty}^{\infty} |\tilde f(t)|\,dt < \infty$, then the identity

$$f(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itu} \tilde f(t)\,dt \tag{6}$$

holds in almost all points of the real line with respect to the Lebesgue measure. Moreover, even the following stronger statement holds. Let μ be a finite measure, i.e. let $\mu(R^1) < \infty$. Let $\tilde f(t) = \int e^{itu}\,d\mu(u)$ denote the Fourier transform of this measure μ. If the function $\tilde f(\cdot)$ is integrable, then the measure μ has a density function, and it agrees with the function $f(\cdot)$ defined in formula (6). This result can be slightly generalized. It also holds if μ is a signed measure with bounded variation, i.e. it is the difference of two finite measures.

The above definition of the Fourier transform slightly differs from the definition usually given in the literature, where the integral in formula (6a) and its analog which defines the Fourier transform of a measure μ is divided by $\sqrt{2\pi}$. With such a normalization the Fourier transform and the formula expressing its inverse are more similar to each other. For us the formula given in (6a) is more convenient. We shall prove this inversion formula for Fourier transforms in the Appendix. Let us remark that the restriction that formula (6) holds only for almost all points u with respect to the Lebesgue measure is natural. Indeed, by modifying a function on a set of Lebesgue measure zero we get a new function which has the same Fourier transform.

Let us finally remark that the formula about the calculation of the Fourier coefficients by means of the Fourier transform suggest the inversion formula (6). Namely, let us approximate a function in the points $k\varepsilon$, $k = 0, \pm 1, \pm 2, \dots,$ in a natural way and let us write the Fourier series with these coefficients (in this case we consider the Fourier series in the interval $\left[-\frac{\pi}{\varepsilon}, \frac{\pi}{\varepsilon}\right]$). Then by expressing the Fourier coefficients of this Fourier series with the help of the Fourier series and by applying the limit procedure $\varepsilon \to 0$ we get, at least on a formal level, formula (6).

8.) Let $\xi_1, \xi_2, \dots,$ be a sequence of independent and identically distributed random variables, $E\xi_1 = 0, E\xi_1^2 = 1$, (i.e. we consider random variables with expectation zero and variance one), and put $S_n = \xi_1 + \dots + \xi_n$, $n = 1, 2, \dots$. Let us also assume that the random variable ξ_1 has a density function $f(x)$, and the Fourier transform $\varphi(t)$ of the density function $f(x)$ is integrable or at least it satisfies the following weakened condition: There exists an integer $k \ge 1$ such that $\varphi^k(t)$ is integrable.

Then the random variable $\frac{S_n}{\sqrt{n}}$ has a density function $f_n(x)$ which satisfies the relation

$$\lim_{n\to\infty} f_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{for all real numbers } x, \tag{7}$$

and the convergence is uniform in the parameter x.

9.) If a sequence of distribution functions $F_n(x)$, $n = 1, 2, \dots$ have density functions $f_n(x)$, and they satisfy formula (7), then

$$\lim_{n\to\infty} F_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \quad \text{for all real numbers } x.$$

The convergence is uniform with respect to the variable x.

10.) If the conditions of problem 8 are satisfied, and also the relation $E|\xi_1|^3 < \infty$ holds, then relation (7) holds with a better error term. Namely,

$$f_n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} + O\left(\frac{1}{\sqrt{n}}\right),$$

and $O(\cdot)$ is uniform with respect to the parameter x.

Remark: We have proved in problems 6 and 10 that if the absolute value of the random variables have finite third moment, then the density function of the appropriately normalized sums can be approximated with the normal density function with an accuracy of order $O\left(n^{-1/2}\right)$, where n is the number of summands in the sum. In the second part of this series of problems we shall prove an analog of this result about the approximation of the appropriately normalized distribution function of partial sums of independent and identically distributed random variables. This is an important result of the probability theory, which is called the Berry–Esseen inequality. It yields an estimate of the same order $O\left(n^{-1/2}\right)$ for the approximation of the normalized distribution function by the standard normal distribution function. Nevertheless, there is an essential difference between the estimate supplied by the Berry–Esseen inequality and the results of problems 6 and 10.

The Berry–Esseen inequality gives an upper bound of the form const. $\mu_3 n^{-1/2}$, where μ_3 is the third moment of the absolute value of the summands, and the const. is a universal number which can be chosen independently of the distribution of the summands. One cannot give a similar universal estimate on the normal approximation of the density function. Indeed, if for instance the density function of the summands take extremely large values in a small neighbourhood of the origin and the probability of this neighbourhood is relatively large, say larger than $\frac{1}{10}$, then for relatively small indices n the density function of the normalized sum of n independent random variables cannot be well approximated by a normal density function, although the third moment of the absolute value of the summands may be not large. A similar example can be given for the density function of the partial sums of lattice valued random variables if the width of the lattice may be very small.

The above sketched examples show that the analog of the Berry–Esseen inequality for density functions does not hold. This difference is related to the fact that in the first case we study distribution functions while in the second case their derivatives. It is a quite common experience in analysis that for the nice behaviour of a sequence of functions we have to impose much weaker conditions than for the nice behaviour of their derivatives.

A deficiency of the results of problems 8 and 9 is that it imposes conditions not for the density but for the characteristic function of the summands, while in typical applications we have some direct information about the density and not on the characteristic function. Hence we have to understand the relation between the properties of a density and the corresponding characteristic function. Later we shall return to this question. We shall see that relation (7) holds for the density function of normalized sums of independent and identically distributed random variables if the summands have "nice" density function.

B.) The definition of the normal distribution function and of the characteristic function. Some important results related to these notions.

Let us introduce the definition of the (standard) normal distribution. The solutions of the previous problems suggest that in limit distributions the normal limit distribution appears as the limit.

The definition of the normal distribution. *The $\Phi(x)$ standard normal distribution function is the distribution function whose density function the standard density function $\varphi(x)$ is of the form $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, that is*

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du.$$

More generally, a linear transform of the standard normal distribution function $\Phi(x)$ is called a normal distribution function. More explicitly, a normal distribution function can be characterized by two parameters, by a real number m and a positive real number σ. The normal distribution function with parameters m and σ equals $\Phi_{m,\sigma}(x)=\Phi\left(\frac{x-m}{\sigma}\right)$. Its density function is $\varphi_{m,\sigma}(x)=\frac{1}{\sigma}\varphi\left(\frac{x-m}{\sigma}\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-m)^2/2\sigma^2}$.

It follows from the solution of problem 1 that $\Phi(x)$ is really a distribution function. The normal distribution is also called the Gaussian distribution in the literature.

11.) Let us show that the expected value of a standard normal random variable equals zero, and its variance equals one. The expected value of a random variable with distribution function $\Phi_{m,\sigma}(x)$ equals m and its variance equals σ^2.

Let us consider the appropriately normalized partial sums of independent and identically distributed random variables with a nice density function. We have seen that the density functions of these normalized partial sums tend to the normal density and their distribution functions to the normal distribution function. A similar result holds for sums of lattice valued random variables. We say that the appropriate normalizations $\bar{S}_n = \frac{S_n - a_n}{b_n}$, $n = 1, 2, \dots$, of a sequence of random variables S_n, $n = 1, 2, \dots$, satisfy the local central limit theorem if either the random variables $\bar{S}_n$ have a density function for large n, and they converge uniformly to the standard normal distribution function, or for all large n there exists a lattice $kh_n + b_n$ of width h_n, $k = 0, \pm 1, \pm 2, \dots$, such that $h_n \to 0$ if $n \to \infty$, the random variable $\bar{S}_n$ is concentrated on this lattice (as on the rarest lattice), and $P(S_n = kh_n + b_n) = h_n \varphi(kh_n + b_n) + o(h_n)$, $k = 0, \pm 1, \pm 2, \dots$, $o(\cdot)$ is uniform in the variable k, where $\varphi(x)$ is the standard normal density function.

We have seen in problems 6 and 8 that the appropriately normalized sums of independent and identically distributed random variables with density function or lattice valued distribution satisfy the local central limit theorem under fairly general conditions. The normalization is made in a natural way. With the notation of the previous paragraph we choose $a_n = nE\xi_1 = ES_n$, $b_n = \sqrt{n \operatorname{Var} \xi_n} = \sqrt{\operatorname{Var} S_n}$, i.e. such a normalization is chosen with which the normalized partial sums have expected value zero and variance one. We also have seen that the local central limit theorem implies the (global) central limit theorem (problems 7 and 9). One would expect that the central limit theorem may hold also in such cases when its local version fails to hold. We want to show that this belief is right. Moreover, we want to prove the central limit theorem in more general cases when appropriately normalized partial sums of independent but not necessarily identically distributed random variables are considered.

We have proved the local central limit theorem with the help of the Fourier transform of the distribution functions we have investigated. We want to show that this method can be adapted to the proof of (global) central limit theorems. To do this we recall the definition of the Fourier transform of general (probability) measures. We shall call them, following the tradition of probability theory terminology, characteristic functions. Since later we also want to study partial sums of vector valued random variables, hence — to avoid some repetition — we define the multi-dimensional version of the characteristic functions.

The definition of characteristic functions. *Let $F(u) = F(u_1, \dots, u_k) = P(\xi_1 < u_1, \dots, \xi_k < u_k)$ denote the distribution function of a k-dimensional random vector $(\xi_1, \dots, \xi_k)$. The $\varphi(t) = \varphi(t_1, \dots, t_k)$, $t = (t_1, \dots, t_k)$, characteristic function of a k-dimensional distribution function F or of an F distributed k-dimensional random vector $\xi = (\xi_1, \dots, \xi_k)$, $k \ge 1$, is defined by the formula*

$$\begin{aligned}\varphi(t) = \varphi(t_1, \dots, t_k) &= Ee^{i(t,\xi)} = Ee^{i(t_1\xi_1 + \cdots + t_k\xi_k)} \\ &= \int e^{i(t,u)} F(\,du) = \int e^{i(t_1u_1 + \cdots + t_ku_k)} F(\,du_1, \dots, \,du_k),\end{aligned}$$

where $t = (t_1, \dots, t_k)$ is an arbitrary point of the k-dimensional Euclidean space. (Here we applied the following notation: If $u = (u_1, \dots, u_k)$ and $v = (v_1, \dots, v_k)$ are two k-dimensional vector, then $(u, v) = u_1v_1 + \cdots + u_kv_k$ denotes the scalar product of the vectors u and v.)

12.) A characteristic function $\varphi(t_1, \dots, t_k)$ is uniformly continuous in the k-dimensional

Euclidean space R^k, $\varphi(0,\dots,0)=1$ and

$$|\varphi(t_1,\dots,t_k)|\le 1 \quad \text{in all points } (t_1,\dots,t_k)\in R^k.$$

Let $\varphi(t)=\varphi(t_1,\dots,t_k)$, $t=(t_1,\dots,t_k)$, be the characteristic function of a random vector $\xi=(\xi_1,\dots,\xi_k)$. Let a be a real number and $m=(m_1,\dots,m_k)$ a k-dimensional vector. Then the characteristic function of the random vector $a\xi+m=(a\xi_1+m_1,\dots,a\xi_k+m_k)$ equals

$$e^{i(m,t)}\varphi(at)=e^{i(m_1t_1+\cdots+m_kt_k)}\varphi(at_1,\dots,at_k).$$

If $\xi_1,\dots,\xi_n$, $1\le j\le n$, is a sequence of independent k-dimensional random vectors with the notation $\xi_j=(\xi_j^{(1)},\dots,\xi_j^{(k)})$, and $\varphi_j(t)=\varphi_j(t_1,\dots,t_k)$, $j=1,\dots,n$, $t=(t_1,\dots,t_k)$, denotes their characteristic functions, then the characteristic function of the sum $\xi_1+\cdots+\xi_n$ equals $\prod\limits_{j=1}^n\varphi_j(t)=\prod\limits_{j=1}^n\varphi_j(t_1,\dots,t_k)$.

Let us calculate the characteristic function of some important distribution functions.

13.) Let us show that if the random variable ξ

a.) has standard normal distribution, i.e. it has a density function of the form $f(u)=\frac{1}{\sqrt{2\pi}}e^{-u^2/2}$, then its characteristic function equals $\varphi(t)=e^{-t^2/2}$.

b.) has uniform distribution in the interval $[0,1]$, i.e. it has a density function of the form $f(u)=1$ if $0\le u\le 1$ and $f(u)=0$ is otherwise, then its characteristic function equals $\varphi(t)=\frac{e^{it}-1}{it}$.

c.) has exponential distribution with parameter $\lambda>0$, i.e. it has a density function of the form $f(u)=\lambda e^{-\lambda u}$ if $u\ge 0$ and $f(u)=0$ if $u<0$, then its characteristic function equals $\varphi(t)=\frac{\lambda}{\lambda-it}$.

d.) has Cauchy distribution, i.e. it has a density function of the form $f(u)=\frac{1}{\pi}\frac{1}{1+u^2}$, then its characteristic function equals $\varphi(t)=e^{-|t|}$.

e.) has Poisson distribution with parameter $\lambda>0$, i.e. $P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,2,\dots$, then its characteristic function equals $\varphi(t)=\exp\left\{\lambda(e^{it}-1)\right\}$.

f.) has binomial distribution with parameters n and p where $n\ge 1$ is an integer, $0<p<1$, i.e. $P(\xi=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $k=0,1,\dots,n$, then its characteristic function equals $\varphi(t)=(1-p+pe^{it})^n$.

g.) has negative binomial distribution with parameters n and p, where $n\ge 1$ is an integer, $0<p<1$, i.e. $P(\xi=k)=\binom{n+k-1}{k}p^k(1-p)^n$, $k=0,1,2,\dots$, then its characteristic function equals $\varphi(t)=\left(\frac{1-p}{1-pe^{it}}\right)^n$.

h.) has γ distribution with parameter s, $s>0$, i.e., its density function is $\gamma_s(u)=\frac{1}{\Gamma(s)}u^{s-1}e^{-u}$, if $u>0$, and $\gamma_s(u)=0$, if $u<0$, where $\Gamma(s)=\int_0^\infty u^{s-1}e^{-u}\,du$, then its characteristic function equals $\varphi_s(t)=\frac{1}{(1-it)^s}$.

It may be worth mentioning that the density function $\gamma_s(u)$, considered in point h.) shows some similarity with the Poissonian distribution, and this can be exploited to prove, similarly to the solution of Problem 2, a good asymptotic formula for the function $\Gamma(s)$ if $s \to \infty$. This means a generalization of Stirling's formula, since one prove with the help of some partial integration that $\Gamma(n) = (n-1)!$.

Let us observe that $\gamma_{s_1}(u) * \gamma_{s_2}(u) = \gamma_{s_1+s_2}(u)$, where $*$ denotes convolution. This can be seen e.g. with the help of the form of the characteristic function of γ_s and some important properties of the characteristic functions discussed later. Further, we can write the identity $\gamma_s(s) = \frac{s^{s-1}e^{-s}}{\Gamma(s)} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-ts}}{(1-it)^s}\,ds$ with the help of the inverse Fourier transform formula for density functions. Then we get by giving a good asymptotic formula for the expression at right-hand side of this identity (similarly to the solution of Problem 2) that $\Gamma(s) \sim \sqrt{2\pi(s-1)}(\frac{s-1}{e})^{s-1}$ if $s \to \infty$.

C.) The definition of the convolution and some of its important properties.

In this series of problems we investigate the asymptotic behaviour of the distribution and density function of appropriately normalized partial sums of independent random variables. These distribution or density functions can be directly expressed by means of the distribution of density functions of the random variables in these partial sums. Hence the limit theorems discussed in this text also can be expressed in the language of distribution (and density) functions without speaking of sums of independent random variables. To do this we have to introduce the notion of the convolution operator. We introduce this notion in a slightly more general form and define the convolution of integrable (not necessary density) functions and signed measures. We shall not use the notion of the convolution in this series of problems. Hence this section could be omitted. Nevertheless, the discussion of limit theorems without the introduction of the convolution would not be complete, hence we introduce it. In several investigations of the analysis and probability theory this notion appears in a natural way, and it appears in the more general form introduced in this text.

The definition of the convolution operator. *If $f(x_1,\dots,x_k)$ and $g(x_1,\dots,x_k)$ are two k-dimensional measurable functions of k variables which are integrable, i.e. $\int |f(x_1,\dots,x_k)|\,dx_1\dots dx_k < \infty$ and $\int |g(x_1,\dots,x_k)|\,dx_1\dots dx_k < \infty$, then the convolution $f * g$ of the functions f and g is the function of k variables defined by the formula*

$$f * g(x_1,\dots,x_k) = \int f(u_1,\dots,u_k)g(x_1-u_1,\dots,x_k-u_k)\,du_1\dots du_k \tag{8}$$

in all such points $(x_1,\dots,x_k)$, where this integral is meaningful. (In the remaining points we can define the function $f * g$ in an arbitrary way.)

Let μ and ν be two signed measures of bounded variation on the measurable sets of the k-dimensional Euclidean space R^k. This means that we assume that there exist two representations $\mu = \mu_1 - \mu_2$ and $\nu = \nu_1 - \nu_2$ such that μ_i and ν_i, $i = 1,2$, are finite measures on the measurable subsets of the space R^k, i.e. $\mu_k(R^k) < \infty$ and $\nu_k(R^k) < \infty$,

$k=1,2$. Let $\mu\times\nu$ denote the direct product of these signed measures μ and ν on the product space $R^k\times R^k=R^{2k}$. Then the convolution $\mu*\nu$ is the signed measure on the measurable subsets of the space R^k defined by the formula

$$\mu*\nu(A)=\mu\times\nu\{(u,v)\colon\ u+v\in A\}\quad \text{for all measurable sets } A\subset R^k.$$

In other words, $\mu*\nu$ is the pre-image of the product (signed) measure $\mu\times\nu$ induced by the transformation $\mathbf{T}$: $R^k\times R^k\to R^k$ defined by the formula $\mathbf{T}(u,v)=u+v$.

Let $f(x_1,\dots,x_k)$ be a measurable and integrable function of k variables, ν a measure of bounded variation on the measurable subsets of R^k. The convolution $f*\nu(x_1,\dots,x_k)$ is the following function of k variables:

$$f*\nu(x_1,\dots,x_k)=\int f(u_1,\dots,u_k)\nu(x_1-\,du_1,\dots,x_k-\,du_k)$$

in all points $(x_1,\dots,x_k)\in R^j$, where this integral is meaningful. This means that we integrate the function $f(\cdot)$ with respect to the measure $\bar\nu_{x_1,\dots,x_k}$ defined by the formula

$$\bar\nu_{x_1,\dots,x_k}(A)=\nu((x_1,\dots,x_k)-A).$$

(In the remaining points where this integral is not meaningful we define the function $f*\nu$ in an arbitrary way.)

We have not defined the convolution $f*g(x_1,\dots,x_k)$ of two functions f and g or the convolution $f*\nu(x_1,\dots,x_k)$ of a function f and a measure ν in all points $(x_1,\dots,x_k)\in R^k$. But this restriction is not so disturbing as one might think at the first sight. As we shall see in problem 14, these convolutions exist for almost all points $(x_1,\dots,x_k)\in R^k$ with respect to the Lebesgue measure in the k-dimensional Euclidean space. On the other hand, in typical applications these convolutions appear as the density function of a signed measure which is absolutely continuous with respect to the Lebesgue measure, and such density functions are determined only almost everywhere.

14.) If $f(x_1,\dots,x_k)$ and $g(x_1,\dots,x_k)$ are two measurable and integrable functions on the space R^k, then the integral in formula (8) defining the convolution $f*g(x_1,\dots,x_k)$ is meaningful for almost all points $(x_1,\dots,x_k)\in R^k$ with respect to the Lebesgue measure. The convolution $f*g$ is a finite and integrable function on R^k.

If μ and ν are two signed measures of bounded variation on the space R^k, then their convolution $\mu*\nu$ has the same property.

If μ and ν are two signed measures of bounded variation on the space R^k, and the measure μ has a density function $f(u_1,\dots,u_k)$, that is $\mu(A)=\int_A f(u_1,\dots,u_k)\,du$ for all measurable sets $A\subset R^k$, then the convolution of these signed measures, the signed measure $\mu*\nu$ has a density function, and it equals the function $f*\nu$. In particular, the function $f*\nu(x)$ is integrable. If both signed measures μ and ν have a density function $f(u_1,\dots,u_k)$ and $g(u_1,\dots,u_k)$, then their convolution, the signed measure $\mu*\nu$ also has a density function, and it equals the convolution $f*g$.

15.) If ξ and η are two independent random vectors on the space R^k, the distribution of ξ is μ, the distribution of η is ν, then the distribution of the sum $\xi+\eta$ is the convolution $\mu*\nu$. If the random vector ξ has a density function $f(u_1,\dots,u_k)$, then the sum $\xi+\eta$ also has a density function, and it equals $f*\nu$. If ξ has a density function f and ν a density function g, then the sum $\xi+\eta$ also has a density function, and it equals the convolution $f*g$.

As a consequence, if ξ_j, $j=1,\dots,n$, are independent random vectors with distribution functions $F_j(x)=F_j(x_1,\dots,x_k)$, $j=1,\dots,n$, $\bar S_n=\frac{\sum\limits_{j=1}^n \xi_j - A}{B}$ with some norming factors $A=(A_1,\dots,A_k)$ and $B>0$, then the distribution of the expression $\bar S_n$ equals $F_1*\cdots*F_n(Bx+A)$. If the random vectors ξ_j have density functions f_j, $1\le j\le n$, then $\bar S_n$ has a density function of the form $Bf_1*\cdots*f_n(Bx+A)$.

$f*g=g*f$, $\mu*\nu=\nu*\mu$, $(f*g)*h=f*(g*h)$, $(\mu_1*\mu_2)*\mu_3=\mu_1*(\mu_2*\mu_3)$, i.e. the convolution operator is commutative and associative.

The next problem is about the relation between the convolution operator and Fourier transform.

16.) If $f(u_1,\dots,u_k)$ and $g(u_1,\dots,u_k)$ are two integrable functions on R^k with Fourier transforms

$$\varphi(t_1,\dots,t_k)=\int e^{i(t_1u_1+\cdots+t_ku_k)}f(u_1,\dots,u_k)\,du_1\dots du_k$$

and

$$\psi(t_1,\dots,t_k)=\int e^{i(t_1u_1+\cdots+t_ku_k)}g(u_1,\dots,u_k)\,du_1\dots du_k,$$

then the Fourier transform of the convolution $f*g(u_1,\dots,u_k)$ is the function $\varphi(t_1,\dots,t_k)\psi(t_1,\dots,t_k)$.

If μ and ν are two signed measures on R^k of bounded variation with Fourier transforms (or in other terminology with characteristic functions

$$\varphi(t_1,\dots,t_k)=\int e^{i(t_1u_1+\cdots+t_ku_k)}\mu(\,du_1,\dots,\,du_k)$$

and

$$\psi(t_1,\dots,t_k)=\int e^{i(t_1u_1+\cdots+t_ku_k)}\nu(\,du_1,\dots,\,du_k)$$

then the Fourier transform of the convolution $\mu*\nu$ equals $\varphi(t_1,\dots,t_k)\psi(t_1,\dots,t_k)$.

In an informal way the following problem can be formulated as the statement that the convolution is a smoothing operator. The convolution of two smooth functions is an even smoother function. For the sake of simplicity we shall consider only functions of one variable.

17.) Let $f(u)$ and $g(u)$ be two integrable functions. Let us assume that the derivatives $\frac{d^j f(u)}{du^j}$ exist, they are integrable functions, and $\lim\limits_{u\to-\infty} \frac{d^j f(u)}{du^j} = 0$ for all integers $0 \le j \le k$ with some integer $k \ge 1$. Let us also assume that the derivatives $\frac{d^j g(u)}{du^j}$ also exits, they are also integrable functions, and $\lim\limits_{u\to-\infty} \frac{d^j g(u)}{du^j} = 0$ for all integers $0 \le j \le l$ with all integers $l \ge 1$. Then the derivative $\frac{d^{k+l} f*g(u)}{du^{k+l}}$ also exists, it is an integrable function, and $\lim\limits_{u\to-\infty} \frac{d^{k+l} f*g(u)}{du^{k+l}} = 0$.

Let $f(u)$ and $g(u)$ be two integrable functions such that the function $f(u)$ has an analytic continuation to the domain $\{z\colon\ |\mathrm{Im}\, z| < A\}$ with some number $A > 0$. Let us further assume that $\int |f(u+ix)|\,du < \infty$ for all numbers $|x| < A$, and for all numbers $\varepsilon > 0$ and $B > 0$ there exists a constant $K = K(A, B, \varepsilon)$ such that $\int_{|u|>K} |f(y-u+ix)g(u)|\,du < \varepsilon$ if $|y| < B$ and $|x| < A$. Then the convolution $f * g$ is also an analytic function in the domain $\{z\colon\ |\mathrm{Im}\, z| < A\}$.

D.) Convergence in distribution.

Although there is no convenient inversion formula for the expression of a distribution function by means of its characteristic function, the proof of the local limit theorems suggests that the convergence of a sequence of distribution functions can be proved by means of the convergence of their characteristic functions. This statement really holds, but to prove it we have to understand the situation better.

First we have to understand the meaning of the convergence of distribution functions. Before giving the formal definition let us consider a simple example which may explain some details of the definition.

Let us consider a sequence of negative numbers x_n, $n = 1, 2, \dots$, which satisfies the relation $\lim\limits_{n\to\infty} x_n = 0$, and put $x_0 = 0$. Let us introduce the (degenerated) probability measures μ_n, $n = 0, 1, 2, \dots$, on the real line which are concentrated in the points x_n, $n = 0, 1, 2, \dots$, that is $\mu_n(\{x_n\}) = 1$, $n = 0, 1, 2, \dots$. It is natural to expect that by a right definition of convergence of probability distributions the sequence of these measures μ_n converges to the measure μ_0 as $n \to \infty$. Let us remark that the measures μ_n, $n = 0, 1, 2, \dots$, are determined by the distribution functions defined by the formula $F_n(u) = 0$ if $u \le x_n$, and $F_n(u) = 1$ if $u > x_n$, $n = 0, 1, 2, \dots$, i.e. $\mu_n([a, b)) = F_n(b) - F_n(a)$ for arbitrary pairs of numbers $a < b$. Observe that the relation $\lim\limits_{n\to\infty} F_n(x) = F_0(x)$ holds for all numbers $x \ne 0$, but this relation does not hold for $x = 0$, since $F_0(0) = 0$, and $F_n(0) = 1$ if $n \ne 0$. This example shows that the naive picture by which the convergence of a sequence of distribution functions to a limit distribution function would mean that these distribution functions converge to the limit distribution function in all points of the real line would not supply a good definition. In the previous example the convergence does not hold in the point $x = 0$, where the limit distribution function is not continuous. The right definition of the convergence of distribution functions given below does not demand convergence in the points of discontinuity of the limit distribution function. We shall give this definition in the multi-dimensional case.

Definition of convergence of distribution functions. *Let $F_n(x_1,\dots,x_k)$, $n=0,1,2,\dots$, be a sequence of k-dimensional, $k\ge 1$, distribution functions. We say that the distribution functions F_n converge to a distribution function F_0 in distribution, or in other words the probability measures μ_n, determined by the distribution functions F_n, $n=1,2,\dots$, converge in distribution to the measure μ_0 determined by the distribution function F_0, or in a third terminology the random vectors $\xi_n=(\xi_n^{(1)},\dots,\xi_n^{(k)})$, $n=1,2,\dots$, with distribution function F_n converge in distribution to a random vector $\xi_0=(\xi_0^{(1)},\dots,\xi_0^{(k)})$, with distribution function F_0 as $n\to\infty$, if*

$$\lim_{n\to\infty} F_n(x_1,\dots,x_k)=F_0(x_1,\dots,x_k)$$

in all points of continuity $(x_1,\dots,x_k)\in R^k$ of the distribution function $F_0(x_1,\dots,x_k)$. (Sometimes the convergence in distribution is called weak convergence in the literature.)

In later investigations the following Theorem A formulated below plays an important role. This result gives an equivalent condition for the convergence in distribution. The next problem is the proof of Theorem A. We remark that the same problem also appears in problem 1 of the series of problems *Weak convergence of probability measures in metric space.* (For the time being it exists only in Hungarian.)

Theorem A. *The distribution functions $F_n(x_1,\dots,x_k)$ converge in distribution to the distribution function $F_0(x_1,\dots,x_k)$ if and only if*

$$\int f(x_1,\dots,x_k)\,dF_n(x_1,\dots,x_k)\to\int f(x_1,\dots,x_k)\,dF_0(x_1,\dots,x_k)\quad \text{if } n\to\infty \tag{9}$$

for all continuous and bounded functions $f(x_1,\dots,x_k)$ in the k-dimensional Euclidean space R^k.

18.) Let us prove Theorem A.

Let us remark that the characterization of the convergence in distribution given in Theorem A also plays an important role in other investigations. It helps to find the good definition of convergence in distribution in general topological spaces. The original definition is closely related to the notion of distribution functions which exploits the simple geometrical structure of the Euclidean space. Hence this definition has no natural generalization to more sophisticated spaces. The result of Theorem A helps to overcome this difficulty, and this is the reason why Theorem A appears in the series of problems mentioned above.

It is also worth mentioning that condition (9) in Theorem A can also be interpreted as a particular case of the weak convergence introduced in functional analysis. The probability measures in the k-dimensional Euclidean space can also be considered as continuous linear functionals on the space of bounded and continuous functions on R^k endowed with the usual supremum norm. Then relation (9) means that the probability measures μ_n determined by the distribution functions F_n weakly converge to the measure μ_0 determined by the distribution function μ_0. Here we identify the measures

μ_n with the linear functionals on the space of continuous and bounded functions they induce and apply the usual terminology of weak convergence in functional analysis. This fact may explain why convergence in distribution is sometimes called weak convergence.

Theorem A suggests a natural approach to study limit theorems in distribution with the help of Fourier analysis. For arbitrary k-dimensional vector the trigonometrical function $e_t(x) = e_{t_1,\dots,t_k}(x_1,\dots,x_k) = e^{i(t,x)} = e^{i(t_1x_1+\cdots+t_kx_k)}$ is continuous and bounded. Hence relation (9) demands in particular that for all vectors $(t_1,\dots,t_k)$ the relation

$$\varphi_n(t_1,\dots,t_k) \to \varphi_0(t_1,\dots,t_k), \quad \text{if } n\to\infty \tag{9'}$$

should hold, where $\varphi_n(t_1,\dots,t_k)$ is the characteristic function of the distribution function $F_n(t_1,\dots,t_k)$, $n=0,1,2,\dots$.

Because of the nice properties of trigonometrical functions it is simpler to check relation $(9')$ than formula (9) dealing with general bounded and continuous functions. The question arises whether relation $(9')$ is sufficient to guarantee the validity of relation of (9) i.e. to prove convergence in distribution. We shall prove a more refined positive result in this direction, and this result serves as the basic tool in investigation of limit theorems in distribution. To prove this result we need such a result which states that the trigonometrical functions $e_{t_1,\dots,t_k}(x_1,\dots,x_k) = e^{i(t_1x_1+\cdots+t_kx_k)}$, or more precisely their finite linear combinations are a sufficiently rich sub-class of the space of continuous and bounded functions. In our discussion we shall apply Weierstrass second approximation theorem which is a result in this spirit. In the Appendix we also supply a proof of Weierstrass second approximation theorem.

Weierstrass second approximation theorem. *For all continuous and periodic by 2π functions $f(t)$ and real numbers $\varepsilon>0$ there exists a trigonometrical polynomial $P_n(t) = \sum\limits_{k=-n}^{n} a_k e^{ikt}$ such that*

$$\sup_{-\infty<t<\infty} |f(t)-P_n(t)| < \varepsilon.$$

(The degree of the polynomial P_n and the coefficients a_k in it depends both on the function $f(\cdot)$ and the real number $\varepsilon>0$. If the function $f(\cdot)$ is real valued, then the coefficients a_k can be chosen in such a way that $a_{-k}=\bar a_k$ for all indices $k=0,1,\dots,n$, where $\bar z$ is the conjugate of the number z. Then also the polynomial $P_n(t)$ is real valued.)

Also the following multi-dimensional version of this result holds. If $f(t_1,\dots,t_k)$ is a continuous function in the k-dimensional Euclidean space which is periodic in all of its variables, i.e. $f(t_1+2j_1\pi,\dots,t_k+2j_k\pi) = f(t_1,\dots,t_k)$ for all integers $j_1,\dots j_k$, and a real number $\varepsilon>0$ is fixed, then there exists a trigonometrical polynomial of k variables

$$P_n(t_1,\dots,t_k) = \sum_{(j_1,\dots,j_k)\colon |j_1|+\cdots+|j_k|\le n} a_{j_1,\dots,j_k} e^{i(j_1t_1+\cdots+j_kt_k)},$$

where $j_1,\dots,j_k$ are integers such that

$$|f(t_1,\dots,t_k)-P_n(t_1,\dots,t_k)| < \varepsilon \quad \textit{for all real numbers } t_1,\dots,t_k.$$

An application of Weierstrass second approximation theorem together with an appropriate scaling makes possible to approximate a k-dimensional continuous function $f(x_1,\dots,x_k)$ with arbitrary accuracy with a finite linear combination of trigonometrical sum $\sum a(t_1,\dots,t_k)e^{i(t_1x_1+\cdots+t_kx_k)}$ in an arbitrary finite domain. On the other hand, given a probability measure μ the integral of the approximating trigonometrical can be expressed by means of the characteristic function $\varphi(t_1,\dots,t_k)$ of the measure μ as $\sum a(t_1,\dots,t_k)\varphi(t_1,\dots,t_k)$. If the domain where the trigonometrical sum well approximates the original function f is chosen sufficiently large, then the above expression gives a natural approximation of the integral $\int f(x_1,\dots,x_k)\,d\mu(x_1,\dots,x_k)$.

In such a way Weierstrass second approximation gives a good approximation of a continuous function in a finite domain by a trigonometrical sum whose integral with respect to a probability measure can be expressed by means of the characteristic function of this measure. But it yields no information about the contribution of a "small neighbourhood of infinity" to the integrals in formula (9). This is not the deficiency of Weierstrass second approximation theorem, but this difficulty belongs to the essence of the problem we investigate. In the following problems we shall show that a sequence of distribution functions may converge in distribution only if the contribution of an appropriate neighbourhood of the infinity in formula (9) is uniformly small for all distribution functions we consider. We shall formulate this statement in a more precise form, and we shall also show that the problem about convergence of distribution functions can be well investigated by means of their characteristic functions. First we introduce the following definition.

Definition of the relative compactness and tightness of distribution functions. *Let a sequence of distribution functions $F_n(t_1,\dots,t_k)$, $n=1,2,\dots,$ be given in the k-dimensional Euclidean space, and let μ_n denote the probability measure on R^k determined by the distribution function F_n. We say that the sequence of distribution functions F_n or probability measures μ_n is relatively compact if all subsequences F_{n_k} (or μ_{n_k}), $k=1,2,\dots,$ of the original sequence F_n or μ_n have a (sub)subsequence $F_{n_{k_j}}$ (or $\mu_{n_{k_j}}$), $j=1,2,\dots,$ which is convergent in distribution.*

We say that a sequence of distribution functions F_n or of probability measures μ_n determined by these distribution functions is tight if for all numbers $\varepsilon>0$ there exists a number $K=K(\varepsilon)$ such that the k-dimensional cube $\mathbf{K}(K)^k=\underbrace{[-K,K]\times\cdots\times[-K,K]}_{k\text{-fold product}}$ satisfies the inequality $\mu_n(\mathbf{K}(K)^k)\ge 1-\varepsilon$ for all indices $n=1,2,\dots$.

19.) Let μ be a probability measure on the Borel measurable subsets of the k-dimensional Euclidean space, and let us fix a real number $\varepsilon>0$. Then there exists a number $K=K(\mu,\varepsilon)>0$ in such a way that the k-dimensional cube $\mathbf{K}^k(K)=\underbrace{-[K,K]\times\cdots\times[-K,K]}_{k\text{-fold product}}$ satisfies the inequality $\mu(\mathbf{K}(K)^k)>1-\varepsilon$. Let us show with the help of this statement and Weierstrass second approximation theorem that a probability measure on R^k is uniquely determined by its characteristic function, that is if two probability measures μ_1 and μ_2 on R^k have the same characteristic function, then $\mu_1 = \mu_2$.

Furthermore, all signed measures μ of bounded variation on the k-dimensional space R^k is uniquely determined by their Fourier transform

$$\varphi(t_1,\dots,t_k) = \int e^{i(t_1x_1+\cdots+t_kx_k)}\mu(\,dx_1,\dots,dx_k), \quad (t_1,\dots,t_k)\in R^k.$$

20.) Let us give an example which shows that the characteristic function of a distribution function in a finite interval does not determine this distribution function. That is let us show that for all numbers $T>0$ there exist two different characteristic functions $F_1(\cdot)$ and $F_2(\cdot)$ such that their characteristic functions $\varphi_i(t)=\int e^{itu}\,dF(u)$, $i=1,2$, satisfy the identity $\varphi_1(t)=\varphi_2(t)$ for all real numbers $-T\le t\le T$.

21.) Let μ_n, $n=1,2,\dots$, be a sequence of probability measures on the k-dimensional Euclidean space R^k. This sequence of probability measures is relatively compact if and only if it is tight. In particular, all sequences of probability measures on R^k which are convergent in distribution are tight.

The hard part in the proof of problem 21 is to show that tightness implies relative compactness. To prove this the limit of an appropriate subsequence of distribution functions has to be constructed, and it has to be checked that the limit is really a distribution function. The proof could be simplified with the help of a classical result of functional analysis (Riesz theorem) which represents finite measures on a compact topological space as the linear functionals on the space of continuous functions on this topological space. This result together with a one-point compactification of the Euclidean space and the characterization of the convergence in distribution given in Theorem A may yield a simpler proof. The tightness condition would guarantee that the limit measure we obtain has no mass in the "infinity". Nevertheless, I have described an elementary but more complicated proof.

I make a small détour. I describe a natural generalization of the results discussed in this section to sequences of probability measures on complete separable metric spaces. We shall not use these results in this note, but they are useful e.g. in the proof of the so-called functional central limit theorem. I shall explain these results without proof, although the results discussed in these section may be useful in these proofs.

Theorem A and formula (9) in it may give a natural idea for the definition of convergence in distribution of probability measures on a separable metric space. We say that a sequence of probability measures defined on a separable metric space (X,ρ) converges in distribution (or weakly, as this convergence is often called in the literature) to a probability measure μ on this metric space, if for all continuous and bounded functions on (X,ρ) the relation

$$\lim_{n\to\infty}\int f(x)\mu_n(\,dx) = \int f(x)\mu(\,dx).$$

holds.

It may be worth remarking that a meausure μ on a separable metric space is uniquely determined by the integrals $\int f(x)\mu(\,dx)$ of all continuous, bounded functions $f(x)$. This fact implies in particular, that the limit of probability measures on a metric space is uniquely determined. We can also define the relative compactness and tightness of sequences of probability measures on general metric spaces, and we can formulate theorems similar to the results of this section for them. At this point it is useful to restrict our attention to complete separable metric spaces.

A sequence of probability measurs μ_n on a complete separable metric space (X,ρ) is called relatively compact if each subsequence μ_{n_k} of μ_n has a sub-subsequence $\mu_{n_{k_j}}$ convergent in distribution (or in other words weakly). It can be proved that in a complete separable metric space of all probability measures μ and real numbers $\varepsilon>0$ there exists a compact set $\mathbf{K}$ such that $\mu(\mathbf{K})>1-\varepsilon$. Hence we can also define the tightness of a sequence of probability measures on a complete separable metric space. We say that a sequence μ_n, $n=1,2,\dots$, of probability meausures on a complete metric space (X,ρ) is tight if for all numbers $\varepsilon>0$ there exists a compact set $\mathbf{K}=\mathbf{K}(\varepsilon)$ such that $\mu_n(\mathbf{K})>1-\varepsilon$ for all indices $n=1,2,\dots$. The following theorem holds.

Theorem about the relation of tightness and relative compactness of probability measure sequences. *Let μ_n, $n=1,2,\dots$, be a sequence of probability measures on a complete separable metric space (X,ρ). This sequence of probability meausures is relatively compact if and only if it is tight. In particular, all sequences of probability measures on a compact, complete separable metric space are (relatively) compact.*

This result is useful in the study of convergence of probabilty measures on general metric spaces. (These problems are interesting, the functional central limit theorem e.g. can be obtained with the help of such an investigation.) To apply the above results in a concrete metric space we need a description of the compact sets of this metric space. The description of the compact sets in the $C([0,1])$ space (in the Banach space of continuous functions on the interval $[0,1]$ with the supremum norm.) We shall describe them in the next theorem.

Theorem about the charactrization of the compact sets in the space $C([0,1])$. *The compact sets of $C([0,1])$ (and in general of all metric spaces are closed.) A closed set $\mathbf{F}$ in the space $C(0,1])$ is compact if and only if*

(a) The set $C=\{x(0)\colon\, x(\cdot)\in\mathbf{F}\}$ is bounded, and

(b) The functions $x(\cdot)$ contained in the set $\mathbf{F}$ are universally uniformly continuous, i.e. for all numbers $\delta>0$ there is such a number $\eta=\eta(\delta)>0$ for which the relation

$$\sup_{x(\cdot)\in\mathbf{F}}\ \sup_{0\le s,t\le1,\,|t-s|\le\eta}|x(t)-x(s)|\le\delta.$$

holds.

In certain investigations, e.g. in the proof of the functional central limit theorem the following consequence of the above results is useful.

Consequence. *Let $X_n(t)$, $0 \le t \le 1$, $n = 1, 2, \dots$, be a sequence of $C([0,1])$-valued random variables. The sequence $X_n(t)$ converges weakly to some $C([0,1])$ valued random variable $X(t)$, $0 \le t \le 1$, as $n \to \infty$ if and only if*

(a) For all finite sequences of numbers $0 \le t_1 < \cdots < t_k \le 1$ the random vectors $(X_n(t_1), \dots, X_n(t_k))$ converge in distribution as $n \to \infty$.

(b) For all numbers $\varepsilon > 0$ and $\delta > 0$ there is a number $\eta = \eta(\delta, \varepsilon) > 0$ such that

$$P\left(\omega\colon \sup_{0 \le s,t \le 1,\, |t-s| \le \eta} |X_n(t,\omega) - X_n(s,\omega)|) \le \delta\right) > 1 - \varepsilon$$

for all indices $n = 1, 2, \dots$.

Remark. Point (a) of the above consequence guarantees that the distributionos of all convergent subsequence of the $C([0,1])$ valued random variables $X_n(\cdot)$ have the same limit. Furthermore, it can be proved with the help of point (b) and the convergence of the (real valued) random variables $X_n(0)$, $n = 1, 2, \dots$, that the μ_n distributions of the random variables $X_n(\cdot)$ constitute a relatively compact set.

E.) The relation between the convergence of distribution functions and the convergence of their characteristic functions.

To express the conditions of convergence in distribution by means of characteristic functions it is useful to give the necessary and sufficient condition of tightness (or of relative compactness which property is equivalent to tightness by the result of problem 20) of probability measures with the help of distribution functions. This is the content of the next problem.

22.) Let $F_n(u)$, $n = 1, 2, \dots$, be a sequence of distribution functions on the real line with characteristic functions $\varphi_n(t)$, $n = 1, 2, \dots$. This sequence of distribution functions $F_n(\cdot)$ is tight if and only if

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{2\delta} \int_{-\delta}^{\delta} \mathrm{Re}\, (1 - \varphi_n(t))\, dt = 0, \tag{10}$$

where $\mathrm{Re}\, z$ denotes the real part of the complex number z.

Although we shall not need the following observation let us also show that relation (10) is equivalent to the following relation (10′):

$$\lim_{\delta \to 0} \sup_{n} \frac{1}{2\delta} \int_{-\delta}^{\delta} \mathrm{Re}\, (1 - \varphi_n(t))\, dt = 0. \tag{10′}$$

With the help of the result of problem 22 and Weierstrass second approximation theorem we can give the necessary and sufficient condition of convergence of a sequence of distribution function in distribution in the language of characteristic function. We formulate and prove this result which we call because of its importance fundamental theorem.

The Fundamental Theorem about the convergence of distribution functions. *Let $F_n(u_1,\dots,u_k)$ be a sequence of distribution functions on the k-dimensional Euclidean space R^k with characteristic functions $\varphi_n(t_1,\dots,t_k)$, $n=1,2,\dots$. If the characteristic functions $\varphi_n(t_1,\dots,t_k)$ have the limit $\varphi_0(t_1,\dots,t_k)=\lim\limits_{n\to\infty}\varphi_n(t_1,\dots,t_k)$ in all points $(t_1,\dots,t_k)\in R^k$, and the limit function $\varphi_0(t_1,\dots,t_k)$ is continuous in the origin, then there exists a distribution function $F_0(u_1,\dots,u_k)$ in the k-dimensional space whose characteristic function is this limit function $\varphi_0(t_1,\dots,t_k)$. Besides, the distribution functions $F_n(u_1,\dots,u_k)$ converge in distribution to this distribution function $F_0(u_1,\dots,u_k)$. Moreover, the condition about the continuity of the limit function $\varphi_0(t_1,\dots,t_k)$ in the origin can be slightly weakened. The above statements also hold if we only demand that the restriction of the function $\varphi_0(t_1,\dots,t_k)$ to each coordinate axis is continuous in the origin.*

In the converse direction we state that if a sequence of k-dimensional distribution functions $F_n(u_1,\dots,u_k)$, $n=1,2,\dots$, converges in distribution to a k-dimensional distribution function $F_0(u_1,\dots,u_k)$, and $\varphi_n(t_1,\dots,t_k)$, $n=0,1,\dots$, denotes the characteristic function of the distribution function $F_n(u_1,\dots,u_k)$, then $\varphi_0(t_1,\dots,t_k)=\lim\limits_{n\to\infty}\varphi_n(t_1,\dots,t_k)$ in all points $(t_1,\dots,t_k)\in R^k$. Furthermore, this convergence is uniform in all compact subsets of the space R^k.

We have formulated the above fundamental theorem in a slightly stronger form than it is done in the literature. We remarked that if we want to deduce the convergence of the distribution functions from the convergence of the characteristic functions, then it is enough to know that the restriction of this limit function to the coordinate axes is continuous in the origin. We have made this remark because its proof causes no problem, and it is useful in certain multi-dimensional application of the result. In particular, it simplifies the solution of problem 46 in this series of problems.

We want to prove the above formulated Fundamental Theorem. To do this let us first prove the following problem interesting in itself.

23.) Let $\xi^{(n)}=(\xi_1^{(n)},\dots,\xi_k^{(n)})$ be a sequence of k-dimensional random vectors and let $\varphi_n(t_1,\dots,t_k)$ denote their characteristic functions. Prove (with the help of problem 22) that if the characteristic functions $\varphi_n(t_1,\dots,t_k)$ converge to a function $\varphi(t_1,\dots,t_k)$ in a small neighbourhood of the origin which is continuous in the origin, then the distribution functions of the random vectors $\xi^{(n)}=(\xi_1^{(n)},\dots,\xi_k^{(n)})$ are tight. Moreover, this tightness property also holds under the weaker condition that the restrictions of the limit function $\varphi(t_1,\dots,t_k)$ to the coordinate axes are continuous in the origin.

24.) Let us prove the Fundamental Theorem about the convergence of distribution functions.

25.) Let us show an example for a sequence of characteristic functions $\varphi_n(t)$, $n=0,1,2,\dots$, on the real line such that $\lim\limits_{n\to\infty}\varphi_n(t)=\varphi_0(t)$, this convergence is uniform on all finite intervals, but it is not uniform on the whole real line.

Let us make some comments about the above formulated Fundamental Theorem.

i.) The fundamental theorem gives conditions which automatically guarantee that the limit of characteristic functions is also a characteristic function. This result is very useful in a more systematic investigation of limit theorems. For instance it makes possible to describe all possible limit distributions which appear as the limit of appropriately normalized partial sums of independent random variables.

ii.) In the Fundamental Theorem we stated the limit of characteristic functions is again a characteristic function if it is continuous in the origin. On the other hand, we know that the characteristic functions are uniformly continuous in the whole space R^k (see for instances the result of problem 12). Let us remark that the class of the characteristic functions is a special class of all continuous functions. There is a non-trivial characterization of characteristic functions. A famous result of the analysis, the Bochner theorem yields such a characterization. This result plays an important role both in the analysis and in the probability theory. Although we shall not use this result in this series of problems, we shall prove Bochner's theorem in the second part of this series of problems. The information obtained in the proof of limit theorems are useful in the proof of Bochner's theorem.

iii.) The condition that the limit of the characteristic functions is continuous in the origin guarantees that the distribution functions corresponding to them are tight. Informally saying this means that "no mass flows out to the infinity". If some mass "may flow out to the infinity", then the asymptotic behaviour of the characteristic functions does not supply such a simple description of the asymptotic behaviour of the distribution functions corresponding to them as the Fundamental Theorem. The content of the next problem is to show this with the help of an example.

26.) Let us show an example for a sequence of probability measures μ_n, $n=1,2,\dots$, on the real line which satisfy the relation $\lim\limits_{n\to\infty}\mu_n(\mathbf{K})=0$ for all bounded sets $\mathbf{K}$, and the characteristic functions $\varphi_n(t)$ of the measures μ_n either

a.) have a limit in all points of the real line or

b.) do not have a limit function on the real line.

F.) The relation between the behaviour of a function or measure and its Fourier transform.

If we want to investigate the behaviour of distribution functions by means of their characteristic functions, or more generally if we want to study the properties of a function or a measure by means of its Fourier transform, then we need a good "dictionary" which describes how the properties of a function or a measure are reflected in their Fourier transform, and what kind of properties of the original function or measure follow from certain properties of their Fourier transform. The "dictionary" described in this section will be by no means complete, but it also contains some results which we do not need in our further investigation. Actually, if we only wanted to get a proof of the central limit theorem and were not interested in the more intricate questions related to this problem, then the proof of the first statement in problem 27 in this section would suffice for our purposes. The results of this section will be formulated only in the one-dimensional case, although the proof of their multi-dimensional generalization would cause no essential difficulty.

Informally the content of this “dictionary” can be formulated in the following way: The smoother a function or a measure (more precisely its density function) on the real line is, the faster its Fourier transform tends to zero at infinity. The less mass a probability measure on the real line contains in a small neighbourhood of the infinity (or in other words the more moments the absolute value of a random variable with this distribution has), the smoother the Fourier transform of this probability measure is. Besides, the derivatives of the Fourier transform at the origin determine the moments of a random variable whose distribution this probability measure is.

In the other direction: If the characteristic function of a distribution function is smooth, and it is enough to assume this smoothness property in a small neighbourhood of the origin, then the smoother the characteristic function is the less mass the probability measure with this characteristic function contains in the vicinity of the infinity. On the other hand, the faster the characteristic function tends to zero as its argument tends to infinity, the smoother the original distribution function is. Furthermore, some results also show that if a function is very smooth everywhere except some points where the function has a singularity, then the behaviour of its Fourier transform in the vicinity of the infinity very well reflects the character of these singularities. We shall discuss this last statement, — which we shall not need later — only superficially, and omit the proof.

Actually, the statement of problem 22 can also be considered as an element of this dictionary, and its content is in full agreement with the above sketched heuristic picture. The statement that a class of distribution functions is tight means that there is an appropriate neighbourhood of the infinity which has a small measure with respect to all measures determined by this class of distribution functions. On the other hand, formula (10) which is equivalent to this property has the content that the characteristic functions of this distribution functions satisfy some sort of uniform continuity in a small neighbourhood of the origin. (Let us recall that $1-\varphi(0)=0$.)

The results of problems 17 and 16 are also consistent with the above heuristics. The result of problem 17 formulates a statement by which the convolution of two smooth density functions is a function which is even smoother than the functions taking part in the convolution. On the other hand, the Fourier transform of the convolution of two functions equals the product of their Fourier transforms. By the above sketched heuristic argument the smoothness of a function depends on how fast its Fourier transform tends to zero in the neighbourhood of infinity. On the other hand, the product of two functions tending to zero in the vicinity of the infinity tends to zero faster than the functions in this product. This fact corresponds to the smoothing property of the convolution operator in the language of Fourier transforms.

27.) If the absolute value of the random variable ξ has finite k-th moment, i.e. $E|\xi^k| <$

∞, then the characteristic function $\varphi(t) = Ee^{it\xi}$ of the random variable ξ is k times continuously differentiable, and $\left.\frac{d^j\varphi(t)}{dt^j}\right|_{t=0} = i^j E\xi^j$ for all numbers $0 \le j \le k$.

A random variable ξ has exponential moments in a small neighbourhood of the origin if and only if the distribution function of the random variable $|\xi|$ tends to 1 exponentially fast in the infinity, i.e. $Ee^{u\xi} < \infty$ for all numbers $|u| \le t$ with some sufficiently small number $t > 0$ if and only if $P(|\xi| > x) < Ce^{-\alpha x}$ for all $x > 0$ with some appropriate constants $C > 0$ and $\alpha > 0$. In this case the characteristic function $\varphi(t) = Ee^{it\xi}$ has an analytic extension to the domain $\{z\colon\ |\mathrm{Re}\, z| < a\}$ with some appropriate number $a > 0$.

28.) If the absolute value of the function $f(u)$ is integrable on the real line, then the Fourier transform $\varphi(t) = \int e^{itu} f(u)\,du$ of the function $f(u)$ satisfies the relations $\lim\limits_{t\to\infty} \varphi(t) = 0$ and $\lim\limits_{t\to-\infty} \varphi(t) = 0$. (Riemann lemma.)

If the absolute value of the function $f(u)$ is integrable, the function $f(u)$ is k-times differentiable, and the absolute value of the functions $\frac{d^j f(u)}{du^j}$ is integrable for all numbers $1 \le j \le k$, then $\int_{-\infty}^{\infty} e^{itu} f(u)\,du = o\left((1+|t|)^{-k}\right)$ in case $|t| \to \pm\infty$.

If the function $f(u)$ has an analytic continuation to the domain $\{z\colon\ |\mathrm{Re}\, z| < A\}$ with some $A > 0$, $|\int |f(u+iv)|\,du < \infty$, if $|v| < A$, then $\varphi(t) = \int e^{itu} f(u)\,du = O\left(e^{-\alpha|t|}\right)$ as $t \to \pm\infty$ with some constant $\alpha > 0$.

29.) Let ξ be a random variable with non-degenerated distribution, (i.e. assume that ξ does not equal a constant with probability 1), and let $\varphi(t) = Ee^{it\xi}$ denote its characteristic function. There exists a constant $t \ne 0$ such that $|\varphi(t)| = 1$ if and only if ξ is a lattice valued random variable. (The notion of lattice valued random variable is introduced in Definition A.) If the values of the random variable ξ are concentrated on a lattice of width h (as on the rarest lattice), then the relation $|\varphi(t)| = 1$ holds for the characteristic function of the random variable ξ if and only if $t = 2\pi\frac{k}{h}$ with some $k = 0, \pm1, \pm2, \dots$. If ξ is a lattice valued, non constant random variable, then there exits a rarest lattice of width h where the distibution of ξ is concentrated. If the random variable ξ is not lattice valued, then $\sup\limits_{A\le|t|\le B} |\varphi(t)| < 1$ for all pairs of numbers $0 < A < B < \infty$.

30.) Let the distribution of the random variable ξ be given by the formula $P(\xi = 1) = P(\xi = \sqrt{2}) = P(\xi = -1) = P(\xi = -\sqrt{2}) = \frac{1}{4}$. Then the characteristic function $\varphi(t) = Ee^{it\xi}$ satisfies the inequality $|\varphi(t)| < 1$ if $t \ne 0$, and also the relation $\limsup\limits_{t\to\infty} |\varphi(t)| = 1$ holds. The relations $\limsup\limits_{t\to\infty} |\varphi(t)| = 1$ and $|\varphi(t)| < 1$ if $t \ne 0$ also hold in the more general case when the values of the random variable ξ are concentrated in finitely or countably many points but ξ is not a lattice valued random variable.

Remark 1: By a classical result of measure theory all (finite) measures can be decomposed (in a unique way) as the sum of an absolute continuous, a discrete (i.e. concentrated in countably many points) and a singular measure. (An absolute continuous measure has a density function with respect to the Lebesgue measure, a singular measure is concentrated on a set of Lebesgue measure zero, but all points have measure zero with respect to it.) By the result of problem 28 the Fourier transform $\varphi(t)$ of an absolute continuous measure tends to zero as $t \to \pm\infty$, and the smoother density function the measure has the faster this convergence is. The results of problems 29 and 30 tell that the Fourier transform of a discrete measure behaves in the opposite way in the neighbourhood of infinity. The lim sup of the absolute value of the Fourier transform of a discrete measure equals the measure of the real line with respect to this measure. The behaviour of a singular measure in the neighbourhood of the infinity cannot be characterized in a similar simple way.

Remark 2: Let us consider in the neighbourhood of infinity the behaviour of the Fourier transform $\tilde{f}(u)$ of an integrable function $f(s)$ which is sufficiently many times differentiable everywhere except a point $a \in R^1$. Let the function f have a singularity of the form $f(s) \sim C|s-a|^\alpha$, $\alpha > -1$, $\alpha \neq 2k$, $k = 0, 1, 2, \dots$, in the neighbourhood of the point a. In this case the Fourier transform of the function f behaves asymptotically as $\tilde{f}(u) \sim \bar{C} e^{iua} u^{-\alpha-1}$ if $u \to \infty$, where the constant $\bar{C} \neq 0$ can be given explicitly. We give a short heuristic explanation of this result.

Because of the smoothness properties of the function $f(s)$ the asymptotic behaviour of its Fourier transform in the neighbourhood of the infinity is determined by the singularity of the function f in the point a, and the Fourier transform equals asymptotically the expression

$$\begin{aligned}\tilde{g}(u) &= C \int_{-\infty}^{\infty} e^{ius} |s-a|^\alpha \, ds = C e^{iau} \int_{-\infty}^{\infty} e^{ius} |s|^\alpha \, ds \\ &= C e^{iau} u^{-\alpha-1} \int_{-\infty}^{\infty} e^{is} |s|^\alpha \, ds = \bar{C} e^{iau} u^{-\alpha-1}\end{aligned}$$

as $u \to \infty$.

The above calculation was rather incorrect. The main problem is that the integrals considered here are meaningless because of the factor $|s|^\alpha$ in the integrand, at least as usual Lebesgue integrals. Nevertheless, this calculation supplies a correct result. Moreover, by replacing the integral to the imaginary axis we can express the constant $\bar{C}$ by means of the $\Gamma(\cdot)$ function. If the function f has several similar singularities, then their effects sum up if the Fourier transform of the function in the neighbourhood of the infinity is described.

Although we shall not apply the results sketched in this Remark 2, such kind of results play an important role in certain investigations of probability theory and analysis. The cases $\alpha = 2k$, $k = 0, 1, \dots$, had to be excluded, because in this case the function $f(s)$ is smooth in a small neighbourhood of the point a. The exceptional behaviour of the function $|s-a|^\alpha$ with such parameters α has deep consequences in certain problems of statistical physics.

The results of the subsequent problems 31–34 yield information about the behaviour of a probability measure by means of the properties of its Fourier transform.

31.) If the characteristic function $\varphi(t) = Ee^{it\xi}$ of a random variable ξ is twice differentiable in the origin, then the random variable ξ has finite second moment, i.e.

$E\xi^2 < \infty$. Let us prove by induction that if the characteristic function of the random variable ξ is $2k$-times differentiable in the origin, then $E\xi^{2k} < \infty$.

32.) If the characteristic function $\varphi(t) = Ee^{it\xi}$ of the random variable ξ is differentiable in a small neighbourhood of the origin, and the derivative is a Lipschitz α function in a small neighbourhood of the origin with some parameter $\alpha > 0$, i.e. $|\varphi'(t) - \varphi'(s)| < C|t-s|^\alpha$ with an appropriate constant $C > 0$ if $|s| < \varepsilon$, $|t| < \varepsilon$ with some number $\varepsilon > 0$, then $E|\xi| < \infty$. Under these conditions also the following somewhat stronger result holds: $P(|\xi| > u) < \text{const.}\, u^{-1-\alpha}$ for all numbers $u > 0$.

If the characteristic function $\varphi(t)$ is $2k+1$-times differentiable in a small neighbourhood of the origin, and the $2k+1$-th derivative is a Lipschitz α function with some constant $\alpha > 0$, then $E|\xi|^{2k+1} < \infty$.

Remark: It follows from the results of problems 31 and 27 that the second moment of a random variable is finite if and only if the characteristic function of the random variable is twice differentiable in the origin. If the absolute value of a random variable has finite expected value, then the characteristic function of this random variable is differentiable in the origin (and everywhere). But to guarantee the existence of a finite expected value of the absolute value of the random variable ξ we have imposed a stronger condition than the differentiability of the characteristic function in the origin. (This would be the natural analog of the statement in problem 31.) Although the conditions of 32 could be weakened, the differentiability of the characteristic function of a random variable in the origin does not suffice for the finiteness of the expectation of the absolute value of this random variable.

Indeed, a necessary and sufficient condition of the differentiability of the characteristic function of a random variable can be expressed with the help of the distribution function of this random variable in a relatively simple way. Since this problem arises in a natural way in the investigation of the weak law of large numbers, this result is also contained in problem 12 of the series of problems *Convergence in probability and with probability one.* (For the time being it exists only in Hungarian). Let us formulate this result. The characteristic function of a random variable with distribution function F has a finite derivative ia, $-\infty < a < \infty$, in the origin if and only if

$$\lim_{x\to\infty} x\,[F(-x) + (1 - F(x))] = 0, \quad \text{and} \quad \lim_{u\to\infty} \int_{-u}^{u} xF(\,dx) = a.$$

33.) If the characteristic function of a random variable ξ has an analytic continuation to a small neighbourhood of the origin, then there exists a number $\alpha > 0$ such that $P(|\xi| > x) \le \text{const.}\, e^{-\alpha x}$ for all numbers $x > 0$.

34.) If the characteristic function $\varphi(u) = Ee^{iu\xi}$ of a random variable ξ is integrable, i.e. $\int |\varphi(u)|\,du < \infty$, then the random variable ξ has a density function $f(x)$. If $|\varphi(u)| < \text{const.}\, |u|^{-(k+1+\varepsilon)}$ with some number $\varepsilon > 0$, then the density function $f(x)$ has continuous and bounded k-th derivative. If $|\varphi(u)| < \text{const.}\, e^{-\alpha|u|}$ with some number $\alpha > 0$, then the density function $f(x)$ is analytic.

Remark 1: The statement of problem 34 can be slightly generalized. To guarantee the existence of a density function of the random variable ξ it is enough to assume that the square of the characteristic function $\varphi(u) = Ee^{iu\xi}$ is integrable. This statement holds because the inverse Fourier formula which expresses the density function of a random variable by means of its characteristic function in the form written down in relation (6) also holds if the characteristic function is only square integrable. Only in this case the integral in formula (6) has to be considered not as a usual Lebesgue integral. It has to be defined by the extension of the L_2 isomorphism between a function and its Fourier transform which follows from the Parseval formula. This is a classical result of the Fourier analysis, but we do not discuss it here. It implies the result of this Remark 1.

Remark 2: We remarked in the formulation of problem 8 that to guarantee the local central limit theorem the condition about the integrability of the characteristic function of the random variables we have considered can be weakened. It is enough to assume that a sufficiently large power of the characteristic function is integrable. The exponential distribution is an example for such a distribution function whose characteristic function is not integrable, but it is square integrable (see problem 13.c). Hence only the strengthened form of the result of problem 8 yields the local central limit theorem in this case. The results of this section may explain the deeper background of this example.

Let us observe that the density function of the exponential distribution function has a discontinuity in the origin, and the non-integrability of its characteristic function is caused by this discontinuity. The convolution of the exponential distribution function with itself yields a distribution function with a smoother density function whose characteristic function is the square of the exponential distribution function. This fact may explain why the square of the Fourier characteristic function of the exponential distribution is integrable. A similar picture arises in the cases of random variables with such a density function which is sufficiently smooth except some exceptional points where the density function has a not too strong singularity.

G.) The central limit theorem.

The previous results enable us to prove the central limit theorem by means of characteristic functions.

35.) Let $\xi_1, \xi_2, \dots,$ be a sequence of independent an identically distributed random variables, $E\xi_1 = 0$, $E\xi_1^2 = 1$. Put $S_n = \xi_1 + \cdots + \xi_n$. Let us show that the random variables $\frac{S_n}{\sqrt{n}}$, $n = 1, 2, \dots,$ converge in distribution to the standard normal distribution function.

We want to prove the central limit theorem in the more general case when the normalized partial sums of independent but not necessarily identically distributed random variables are considered. It is worth formulating this question in an even more general setting when for all positive integers k we consider the sum of independent random variables indexed with this number k, but we assume nothing about the relation between random variables with different indices k. If the random variables indexed with the same number k are uniformly small, then under some weak conditions the sums of the random variables with the same index k have a Gaussian limit as $k \to \infty$.

Essentially we can give the necessary and sufficient condition for the convergence of the distributions of the above sums to a Gaussian law. To formulate this statement in an explicit form we introduce the notion of triangular arrays. The central limit theorem for normalized partial sums of independent but not necessarily identically distributed random variables can be deduced as the special case of the central limit theorem for triangular arrays.

The definition of triangular arrays. *A system of random variables*

$$\begin{array}{l} \xi_{1,1},\dots,\xi_{1,n_1} \\ \vdots \qquad\quad \vdots \\ \xi_{k,1},\dots,\xi_{k,n_k} \\ \vdots \qquad\quad \vdots \end{array}$$

with $k=1,2,\dots$ is called a triangular array, if the random variables $\xi_{k,1}$, ... , ξ_{k,n_k} with the same first index k are independent. (We assume nothing about the relation of the random variables with different first index k.)

To prove the central limit theorem by means of characteristic function it is worth proving the following technical lemma. It gives an estimate of the difference of the function e^{it} and the sum of the first k terms of its Taylor series.

36.) For all non-negative integers k and real numbers t

$$\left|e^{it}-\left(1+\frac{it}{1!}+\cdots+\frac{(it)^k}{k!}\right)\right|\le\frac{|t|^{k+1}}{(k+1)!}. \tag{11}$$

If we want to prove the central limit theorem for triangular arrays with the help of characteristic functions, then the following approach is natural. Let us consider the characteristic functions $\varphi_{k,j}(\cdot)$ of the random variables $\xi_{k,j}$, $1\le j\le n_k$. We have to show that the product of these characteristic functions converge to the characteristic function of a Gaussian random variable. Let us consider the logarithm of these products. Then we have to investigate the sum of the functions $\log\varphi_{k,j}(t)$ with a fixed index k and fixed argument t. If we have a condition which says that the random variables $\xi_{k,j}$ are small in an appropriate sense, then it is natural to expect that $\varphi_{j,k}(t)\sim 1$ for a fixed number t, and the error of the approximation $\log\varphi_{k,j}(t)\sim(1-\varphi_{k,j}(t))$ gives a negligible error. This enables to reduce the proof to the investigation of a simpler sum. The goal of the next problem is to give a precise formulation and justification of the above heuristic argument. It makes possible to study both the necessary and sufficient conditions of the central limit theorem.

37.) Let us consider the characteristic function $\varphi(t)$ of a random variable ξ with a fixed argument t.

a.) If $E\xi=0$ and $E\xi^2\le\varepsilon$ with a sufficiently small positive number $\varepsilon=\varepsilon(t)>0$, then $|1-\varphi(t)|\le\frac{t^2}{2}E\xi^2$, and $|\log\varphi(t)+(1-\varphi(t))|\le t^4\left(E\xi^2\right)^2$.

b.) Let $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, be a triangular array such that $E\xi_{k,j}=0$, $k=1,2,\dots$, $1\le j\le n_k$, $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\xi_{k,j}^2=1$, and let the elements of triangular array satisfy the uniform smallness condition $\lim\limits_{k\to\infty}\left(\sup\limits_{1\le j\le n_k}E\xi_{k,j}^2\right)=0$. Let $\varphi_{k,j}(t)=Ee^{it\xi_{k,j}}$, $-\infty<t<\infty$, denote the characteristic function of the random variable $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$. The random sums $S_k=\sum\limits_{j=1}^{n_k}\xi_{k,j}$, $1\le k<\infty$, converge in distribution to the Gaussian distribution with expectation m and variance σ^2 as $k\to\infty$ if and only if

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}(\varphi_{k,j}(t)-1)=-\frac{\sigma^2t^2}{2}+imt \tag{12}$$

for all numbers $-\infty<t<\infty$.

Although we are first of all interested in the question when the partial sums S_k of a triangular array satisfying the conditions of problem 37 converge in distribution to the standard normal distribution, it was also useful to formulate the condition for the convergence of these random sums to a general Gaussian distribution with expectation m and variance σ^2. This knowledge helps us to find also the necessary conditions of the central limit theorem. We shall also show such examples where the appropriately normalized partial sums of independent random variables with expectation zero converge to a Gaussian distribution with expectation $m\neq0$.

In the investigation of the central limit theorem for triangular arrays the so-called Lindeberg condition formulated below plays a fundamental role. We shall see that the partial sums $S_k=\sum\limits_{j=1}^{n_k}\xi_{k,j}$ of a triangular array satisfying part b) of problem 37 converge to the standard normal distribution function if and only if this triangular array satisfies the Lindeberg conditions.

The definition of the Lindeberg condition: *Let $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, be such a triangular array for which $E\xi_{k,j}=0$, $k=1,2,\dots$, $1\le j\le n_k$, and $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\xi_{k,j}^2=1$. This triangular array satisfies the Lindeberg condition if and only if for all positive real numbers $\varepsilon>0$*

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}E\xi_{k,j}^2I\left(\{|\xi_{k,j}|>\varepsilon\}\right)=0,$$

where $I(A)$ denotes the indicator function of a set A.

38.) Let $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, be a triangular array which satisfies the relations $E\xi_{k,j}=0$, $k=1,2,\dots$, $1\le j\le n_k$, $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\xi_{k,j}^2=1$ and the Lindeberg condition. Then

a.) the triangular array $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, also satisfies the uniform smallness condition $\lim\limits_{k\to\infty}\left(\sup\limits_{1\le j\le n_k}E\xi_{k,j}^2\right)=0$.

b.) The random sums $S_k=\sum\limits_{j=1}^{n_k}\xi_{k,j}$, $1\le k<\infty$, converge in distribution to the standard normal distribution (i.e. to the normal distribution with expectation zero and variance 1) if $k\to\infty$.

The above result can be reversed in the following way.

39.) Let $\xi_{k,j}$, $k=1,2,\dots$, $1\le j\le n_k$, be a triangular array such that $E\xi_{k,j}=0$, $k=1,2,\dots$, $1\le j\le n_k$, $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\xi_{k,j}^2=1$, and it satisfies the uniform smallness condition $\lim\limits_{k\to\infty}\left(\sup\limits_{1\le j\le n_k}E\xi_{k,j}^2\right)=0$. Let us also assume that the random sums $S_k=\sum\limits_{j=1}^{n_k}\xi_{k,j}$, $1\le k<\infty$, converge in distribution to a normal distribution function with variance 1 and arbitrary expected value if $k\to\infty$. Then the triangular array $\xi_{k,j}$, $1\le k<\infty$, $1\le j\le n_k$, also satisfies the Lindeberg condition.

The content of the Lindeberg condition is that it guarantees that the too large values (which have the same order as the square-root of the variance of the sum) have a negligible influence in the central limit theorem. The contribution of such extremely large values have a small influence both on the variance and the distribution of the sum we investigate. The next problem expresses such a fact. This result together with the result of problem 42 formulated later have the following consequence: Let us consider a triangular array which satisfies the Lindeberg condition. If we truncate the too large values (those which are greater than a fixed positive number $\varepsilon>0$) and then normalize the truncated random variables in such a way that their expected value be zero, then the sum of these modified random variables have the same limit in distribution as the sums of the original random variables.

It is worth mentioning that some modification of the above argument may help to find a different proof of the central limit theorem if the Lindeberg condition holds. It can be seen that we may also truncate and normalize the elements of the k-th row of the triangular array at a level $\varepsilon_k>0$ with such a sequence $\varepsilon_k\to0$ which converges to zero sufficiently slowly instead of a fixed number $\varepsilon>0$ and we can guarantee that the new triangular array is equiconvergent with the original one. Then some not too hard calculation shows that all moments of the sums made from this truncated and normalized random variables converge to the corresponding moments of a random variable with standard normal distribution. On the other hand, it is known from the general theory that this fact implies the central limit theorem. Now we omit the details. We return to this remark in Part II. of this series of problems.

40.) Let a triangular array $\xi_{k,j}$, $1 \le j \le n_k$, $\lim\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} E\xi_{k,j}^2 = 1$, $E\xi_{k,j} = 0$, satisfy the Lindeberg condition. Let us fix a real positive number $\varepsilon > 0$, and define the random variable $\bar\xi_{k,j} = \bar\xi_{k,j}(\varepsilon) = \xi_{k,j} I(|\xi_{k,j}| < \varepsilon) - E\xi_{k,j} I(|\xi_{k,j}| < \varepsilon)$. Then $\lim\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} E\bar\xi_{k,j}^2 = 1$. Let us also define the partial sums $\bar S_k = \sum\limits_{j=1}^{n_k} \bar\xi_{k,j}$ and $S_k = \sum\limits_{j=1}^{n_k} \xi_{k,j}$. Then the differences $S_k - \bar S_k$ converge stochastically to zero if $k \to \infty$.

Let us emphasize that to guarantee the validity of the Lindeberg condition in problem 39 we have imposed not only the condition that the random sums S_k converge in distribution to a normal law. We also demanded that the limit distribution have the "right" variance 1. We also shall show examples where the triangular array $\xi_{k,j}$, $1 \le k < \infty$, $1 \le j \le n_k$, satisfies the uniform smallness condition $\lim\limits_{k\to\infty} \left(\sup\limits_{1\le j\le n_k} E\xi_{k,j}^2 \right) = 0$, it does not satisfy the Lindeberg condition, and the random sums S_k converge in distribution to a normal distribution. But in this example the variance of the limit distribution is less than 1. Before the discussion of such examples we present some results which give useful sufficient conditions for the validity of the Lindeberg conditions. We shall consider only normalized partial sums of independent random variables instead of triangular arrays. First we formulate the appropriate version of the Lindeberg condition for sequences of independent random variables.

The definition of the Lindeberg condition for sequences of independent random variables. *Let ξ_n, $n = 1, 2, \dots$, be a sequence of independent random variables, for which $E\xi_n = 0$, $\sigma_n^2 = E\xi_n^2 < \infty$, $n = 1, 2, \dots$, and the sequence of numbers $s_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$, $n = 1, 2, \dots$, satisfies the relation $\lim\limits_{n\to\infty} s_n^2 = \infty$. The sequence ξ_n, $n = 1, 2, \dots$, satisfies the Lindeberg condition if*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} E\xi_k^2 I(\{|\xi_k| > \varepsilon s_n\}) = 0.$$

for all numbers $\varepsilon > 0$.

Given a sequence ξ_n, $n = 1, 2, \dots$, of independent random variables for which $E\xi_n = 0$, $\sigma_n^2 = E\xi_n^2 < \infty$, and the sequence $s_n^2 = \sum\limits_{k=1}^{n} \sigma_k^2$ satisfies the relation $\lim\limits_{n\to\infty} s_n^2 = \infty$, define the triangular array $\xi_{k,j}$, $k = 1, 2, \dots$, $j = 1, \dots, n_k$, by the formulas $n_k = k$ and $\xi_{k,j} = \frac{\xi_j}{s_k}$, if $1 \le j \le k$ with the help of this sequence. The original sequence of random variables ξ_n, $n = 1, 2, \dots$ satisfies the Lindeberg condition if and only if the triangular array $\xi_{k,j}$ defined by its help satisfies it. Hence the central limit theorem and its converse formulated for triangular arrays can be reformulated for normalized partial sums of independent random variables. (In this reformulation the uniform smallness property of the summands have the form $\lim\limits_{n\to\infty}\frac{\max\limits_{1\le k\le n}\sigma_k^2}{s_n^2}=0$.) In the next problem we formulate some properties which imply the Lindeberg condition.

41.) Let ξ_n, $n=1,2,\dots$, be a sequence of independent random variables for which $E\xi_n=0$, $\sigma_n^2=E\xi_n^2<\infty$, $n=1,2,\dots$, $\lim\limits_{n\to\infty}s_n^2=\infty$, where $s_n^2=\sum\limits_{k=1}^n\sigma_k^2$. This sequence of random variables satisfies the Lindeberg condition if one of the following properties holds.

a.) $E|\xi_k|^{2+\alpha}<\infty$, for all numbers $k=1,2,\dots$ with some constant $\alpha>0$, and $\lim\limits_{n\to\infty}\frac{\left(\sum\limits_{k=1}^n E|\xi_k|^{2+\alpha}\right)^{2/(2+\alpha)}}{s_n^2}=0$. In particular, this condition holds if $E\xi_k^2\ge K$ with some constant $K>0$ for all indices $k=1,2,\dots$, and besides the relation $E\lim\limits_{k\to\infty}k^{-\alpha/2}E|\xi_k|^{2+\alpha}=0$ holds.

b.) The independent random variables ξ_n, $n=1,2,\dots$, are uniformly distributed. (This means that the result of problem 35 follows from the central limit theorem formulated under general conditions.)

We want to show an example for a sequence of independent random variables with expectation zero and finite variance which satisfies the uniform smallness condition, the normalized partial sums made with the help of this sequence converge in distribution to the standard normal distribution, but the normalization is not the natural one, that is we divide the partial sums not with the square-root of their variances. Such an example shows that the normalized partial sums of independent random variables may satisfy the central limit theorem (with a non-usual normalization) also if the Lindeberg condition does not hold. Before the construction we prove a simple result which is useful also in other investigations. This result states that a sequence of random variables convergent in distribution has the same limit as its small perturbations. Let us remark that a similar result also holds for random variables taking values in more general spaces.

42.) Let two sequences of random variables S_n and T_n, $n=1,2,\dots$, be given such that the sequence of random variables S_n, $n=1,2,\dots$, converges in distribution to a distribution function F and the sequence T_n, $n=1,2,\dots$, converges stochastically to zero, i.e. $P(|T_n|>\varepsilon)\to 0$ for all numbers $\varepsilon>0$ if $n\to\infty$. Then the sequence of random variables S_n+T_n, $n=1,2,\dots$, converges to the distribution function F.

43.) Let us construct a sequence of independent random variables ξ_n, $n=1,2,\dots$, for which $E\xi_n=0$, $E\xi_n^2=1$, $n=1,2,\dots$, and the sequence of partial sums $S_n=\sum\limits_{k=1}^n\xi_k$, $n=1,2,\dots$, satisfy one of the following statement:

a.) The sequence of normalized partial sums $\sqrt{\frac{2}{n}}S_n$, $n=1,2,\dots$, converge in distribution to the central normal distribution.

b.) The sequence of normalized partial sums $\sqrt{\frac{2}{n}}S_n$, $n = 1, 2, \dots$, converges in distribution to a normal random variable with expected value $m \neq 0$ and variance 1.

The result of problem 38 implies that an example satisfying problem 43 cannot satisfy the Lindeberg condition. Indeed, if the Lindeberg condition held, then the distribution of the normalized partial sums $\frac{S_n}{\sqrt{n}}$ would converge to the standard normal distribution function. Let us also remark that in the counter examples of problem 43 the limit distribution had a *strictly smaller* variance than the variance of the normalized partial sums. The following problem shows that no counter-example exists where the limit distribution has too large variance.

44.) Let a sequence of distribution functions F_n, $n = 1, 2, \dots$, converge in distribution to a distribution function $F_0(x)$. Then $\liminf\limits_{n\to\infty} \int u^2 F_n(\,du) \ge \int u^2 F_0(\,du)$.

H.) THE MULTI-DIMENSIONAL CENTRAL LIMIT THEOREM.

The multi-dimensional limit theorems can be investigated by means of the characteristic functions similarly to their one-dimensional analog. Moreover, the characteristic function method enables us to reduce the investigation of the multi-dimensional limit theorems to the one-dimensional case. This is the content of the subsequent two problems.

45.) Let $\mathbf{Z} = (Z_1, \dots, Z_m)$, $n = 1, 2, \dots$, be an m-dimensional random vector. Let us consider for all m-dimensional vectors $(a_1, \dots, a_m) \in R^m$ the random variable $Z = Z(a_1, \dots, a_m) = \sum\limits_{j=1}^m a_j Z_j$. The distribution functions of all one-dimensional random variables $Z = Z(a_1, \dots, a_m)$ also detbermine the distribution of the random vector $\mathbf{Z} = (Z_1, \dots, Z_m)$.

46.) Let $\mathbf{Z}_n = (Z_{1,n}, \dots, Z_{m,n})$, $n = 1, 2, \dots$, be a sequence of m-dimensional random vectors. The random vectors $\mathbf{Z}_n$ converge in distribution to an m-dimensional distribution as $n \to \infty$ if and only if for all vectors $(a_1, \dots, a_m) \in R^m$ the one-dimensional random variables $Z_n = Z_n(a_1, \dots, a_m) = \sum\limits_{j=1}^m a_j Z_{j,n}$, $n = 1, 2, \dots$, converge in distribution as $n \to \infty$. If the random vectors $\mathbf{Z}_n$ converge in distribution to a probability measure μ in the m-dimensional space, then this limit measure μ can be characterized in the following way. If $\mathbf{Z} = (Z_1, \dots, Z_m)$ is a random vector with distribution μ, and $(a_1, \dots, a_m) \in R^m$ is an arbitrary m-dimensional vector, then the distribution function of the random variable $Z = Z(a_1, \dots, a_m) = \sum\limits_{j=1}^m a_j Z_j$ equals the limit of the distribution functions of the random $Z_n = Z_n(a_1, \dots, a_m) = \sum\limits_{j=1}^m a_j Z_{j,n}$. In particular, we claim that these relations uniquely determine the measure μ.

Let us define the notion of multi-dimensional normal (Gaussian) distributions. It will be the natural analog of the one-dimensional normal distribution. The multi-dimensional normal distributions with expectation zero are exactly those distributions which appear as the limit in the multi-dimensional version of the central limit theorem. Before their formal definition let us recall some notions.

The expected value of an m-dimensional random vector $\mathbf{Z} = (Z_1, \dots, Z_m)$ is the m-dimensional vector $\mathbf{M} = (M_1, \dots, M_m)$ whose coordinates are the numbers $M_j = E\xi_j$, $1 \le j \le m$. The covariance matrix of the random vector $\mathbf{Z} = (Z_1, \dots, Z_m)$ is the $m \times m$ matrix $(D_{j,k})$, $1 \le j,k \le m$ whose elements are the numbers $D_{j,k} = EZ_jZ_k - EZ_jEZ_k = E(Z_j - EZ_j)(Z_k - EZ_k)$. In our notations a vector $\mathbf{b} = (b_1, \dots, b_m)$ will mean a row-vector, and we shall denote its transpose, which is a column vector, by $\mathbf{b}^*$. If $\mathbf{x} = (x_1, \dots, x_m) \in R^m$ and $\mathbf{y} = (y_1, \dots, y_m) \in R^n$ are two m-dimensional vectors, then $(\mathbf{x}, \mathbf{y})$ denotes their scalar product, i.e. $(\mathbf{x}, \mathbf{y}) = \sum\limits_{j=1}^{m} x_j y_j$.

The definition of the multi-dimensional normal distribution. *Let ξ_j, $1 \le j \le m$, be independent random variables with standard normal distribution. Then the random vector $\xi = (\xi_1, \dots, \xi_m)$ will be called an m-dimensional standard normal random vector and its distribution the m-dimensional standard normal distribution. If $B = (b_{j,k})$, $1 \le j,k \le m$, is an $m \times m$ matrix, $\mathbf{M} = (M_1, \dots, M_m)$ is an m-dimensional vector and $\xi = (\xi_1, \dots, \xi_m)$ is an m-dimensional standard normal random vector, then $\xi B + \mathbf{M}$ is called an m-dimensional vector with normal distribution. A probability measure μ on the measurable sets of the m-dimensional Euclidean space R^m is called an m-dimensional normal distribution if and only if it equals the distribution of an m-dimensional random vector with normal distribution defined in the above way with an appropriate $m \times m$ matrix B and a vector $\mathbf{M} \in R^m$.*

Let us first characterize the multi-dimensional normal distributions.

47.) The covariance matrix Σ of all m-dimensional random vectors is positive semi-definite, i.e. for all m-dimensional vectors $\mathbf{x} = (x_1, \dots, x_m)$ $\mathbf{x}\Sigma\mathbf{x}^* = (\mathbf{x}\Sigma, \mathbf{x}) \ge 0$. In the converse direction, for all m-dimensional vectors $\mathbf{M} = (M_1, \dots, M_m) \in R^m$ and $m \times m$ symmetric, positive semi-definite matrices Σ there exists an m-dimensional random vector with normal distribution whose expected value is this vector $\mathbf{M}$ and whose covariance matrix is this matrix Σ. The characteristic function of an m-dimensional normal distribution with expected value $\mathbf{M}$ and covariance matrix Σ equals

$$\varphi(t_1, \dots, t_m) = Ee^{i(\mathbf{t}, \xi)} = Ee^{i(t_1\xi_1 + \cdots + t_m\xi_m)} = \exp\left\{ -\frac{(\mathbf{t}\Sigma, \mathbf{t})}{2} + i(\mathbf{t}, \mathbf{M}) \right\}, \quad (13)$$

where $\mathbf{t} = (t_1, \dots, t_m)$, and $\xi = (\xi_1, \dots, \xi_m)$ denotes an m-dimensional random vector with expected value $\mathbf{M}$ and covariance matrix Σ. In particular, an m-dimensional normal distribution is uniquely determined by its expected value $\mathbf{M}$ and covariance matrix Σ.

The special form of the characteristic function of an m-dimensional characteristic function given in formula (13) has several simple but important consequences. It implies the most important properties of Gaussian distributions. Although these results do not belong to the subject of this series of problems, I discuss two problems that may be useful in such investigations.

48.) Let ξ_j, $1 \le j \le m$, be m independent random variables with standard normal distribution, $\mathbf{M} = (M_1, \dots, M_l)$ an l-dimensional (deterministic) vector, and B a rectangular matrix of size $l \times m$. Then $(\eta_1, \dots, \eta_l) = (\xi_1, \dots, \xi_m)B + \mathbf{M}$ is an l-dimensional random vector with normal distribution. In particular, if we preserve only l coordinate of a a normal random vector η of dimension m, then we get a normal random vector of dimension l.

49.) Let $\eta = (\eta_1, \dots, \eta_m)$ be such a random vector of dimension m with normal distribution whose covariance matrix $\Sigma = (\sigma_{p,q})$, $1 \le p, q \le m$, has the following property: The set $\{1, \dots, m\}$ has a partition $\{1, \dots, m\} = \bigcup\limits_{j=1}^{k} L_j$, $1 \le j \le k$, such that the non zero elements of the matrix Σ are concentrated in the union of the squares $L_1 \times L_1, \dots, L_k \times l_k$, i.e., $\sigma_{p,q} = 0$, if $p \in L_j$, $q \in L_{j'}$, and $j \ne j'$. In this case the random vectors $\bar\eta_j = (\eta_p,\ p \in L_j)$, $1 \le j \le k$, obtained by an appropriate grouping of the coordinates of the vector η are Gaussian random vectors, independent of each other.

Finally we formulate the multi-dimensional central limit theorem for appropriately normalized partial sums of independent vectors. We do not formulate this result in its most general form, and do not discuss its version for triangular arrays, although this would be also possible.

50.) Let $\xi_k = (\xi_{1,k}, \dots, \xi_{m,k})$, $k = 1, 2, \dots$, be a sequence of independent m-dimensional random vectors with expectation zero, i.e. we assume that the relation $\mathbf{M}_k = (E\xi_{1,k}, \dots, E\xi_{m,k}) = (0, \dots, 0)$ holds for all numbers $k = 1, 2, \dots$. Let us also assume that the random vectors $\xi_k = (\xi_{1,k}, \dots, \xi_{m,k})$ have a finite covariance matrix Σ_k for all indices $k = 1, 2, \dots$, and the relation $\lim\limits_{n\to\infty} \frac{1}{A_n^2} \sum\limits_{k=1}^{n} \Sigma_k = \Sigma$ holds with an $m \times m$ matrix Σ and with norming constants A_n, $n = 1, 2, \dots$, such that $A_n \to \infty$ if $n \to \infty$. If beside this all coordinates of the random vectors ξ_k, $k = 1, \dots$, satisfy the Lindeberg condition, i.e.

$$\lim_{n\to\infty} \frac{1}{A_n^2} \sum_{k=1}^{n} E\xi_{p,k}^2 I(|\xi_{p,k}| > \varepsilon A_n) = 0, \qquad \text{for all numbers } p = 1, \dots m, \tag{14}$$

for all $\varepsilon > 0$, then the normalized partial sums $\frac{1}{A_n}\mathbf{S}_n = \frac{1}{A_n}(S_{1,n}, \dots, S_{m,n}) = \frac{1}{A_n} \sum\limits_{k=1}^{n} (\xi_{1,k}, \dots, \xi_{m,k})$ converge in distribution to the normal distribution function with expected value $M = (0, \dots, 0)$ and covariance matrix Σ.

In particular, if $\xi_k = (\xi_{1,k}, \dots, \xi_{m,k})$, $k = 1, 2, \dots$, is a sequence of independent and identically distributed m-dimensional random vectors with expectation zero and finite covariance matrix Σ, then the normalized partial sums $\frac{1}{\sqrt{n}}(S_1, \dots, S_m) =$

$\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n}(\xi_{1,k},\dots,\xi_{m,k})$ converge in distribution to the normal distribution function with expectation zero and covariance Σ.

51.) Let $\xi_n=(\xi_{1,n},\dots,\xi_{m,n})$, $n=1,2,\dots$, be a sequence of m-dimensional random vectors with expectation zero and such that all random vectors $\xi_n=(\xi_{1,n},\dots,\xi_{m,n})$ have a finite covariance matrix Σ_n for all indices $n=1,2,\dots$. Let us further assume that the relation $\lim\limits_{n\to\infty}\frac{1}{A_n^2}\sum\limits_{k=1}^{n}\Sigma_n=\Sigma$ with an appropriate matrix Σ and norming factors A_n, $n=1,2,\dots$, such that $A_n\to\infty$, ha $n\to\infty$. Besides, we also assume that the uniform smallness condition $\lim\limits_{n\to\infty}\frac{\max\limits_{1\le j\le m}\max\limits_{1\le k\le n}E\xi_{j,k}^2}{A_n^2}=0$ holds. If the sequence of random vectors $\frac{S_n}{A_n}$ converges in distribution to an m-dimensional normal distribution whose covariance matrix is Σ, then also the Lindeberg condition formulated in relation (14) holds.

Some additional remarks

We formulate some problems whose investigation is a natural continuation of the study of the problems in this paper.

1.) The central limit theorem for appropriately normalized partial sums of independent random variables was deduced from the convergence of their characteristic functions. Actually not only the convergence of the characteristic functions can be proved, but also the speed of convergence can be given. Besides, the characteristic functions of the normalized partial sums can be better approximated if we add some correction terms to the characteristic function of the approximating normal distribution. These correction terms can be found by means of a natural Taylor expansion. It is natural to expect that these estimates also supply a good estimate about the accuracy of the normal approximation of the distribution function of normalized partial sums of independent random variables. Moreover, a better approximation of their characteristic functions also yields a better approximation of the distribution function of normalized partial sums of independent random variables if some appropriate correction terms are added to the normal approximation.

A natural analog of the above question is to give a good estimate on the Gaussian approximation of the density function of appropriately normalized partial sums of independent random variables and to find a better approximation for this density function by means of an appropriate asymptotic expansion. Naturally, in this investigation we have to impose some additional assumptions on the distribution of the random variables whose partial sums we investigate to guarantee the existence of a nice density function. This problem about the behaviour of density functions is simpler, because there exists a relatively simple inverse Fourier transformation formula to express a density function by means of its Fourier transform. On the other hand, there is no really useful formula for the calculation of distribution functions with the help of their characteristic function in a simple, well applicable form. But by convolving the distribution function of the normalized partial sums we want to investigate with an appropriate density function (essentially concentrated in a small neighbourhood of the origin) we get a new distribution function close to the original one which has a density function, and as a consequence this modified (smoothed) distribution function can be well estimated as the integral of its density function. This approach together with some additional ideas enables us to study the problems mentioned in the previous paragraph. These problems will be the main part of the (essentially shorter) second part of this series of problems.

2.) We have seen that the appropriately normalized partial sums of independent random variables converge in distribution to a normal distribution function under some mild conditions. On the other hand, we would like to get a complete picture about all possible limit theorems for the distribution of appropriately normalized partial sums or in a more general way, for the sums of the elements in the same rows of a triangular array of random variables. It is natural to impose some kind of uniform smallness condition which should guarantee that there are no such dominating terms in the random sums we investigate whose magnitude is comparable with the magnitude of the whole sum.

A fairly complete picture can be given about all possible limit theorems for the normalized partial sums of independent random variables. In the proof the *Fundamental Theorem about the convergence of distribution functions* proved also in this series of problems plays a key role. It makes possible to reformulate the problem to the language of characteristic functions. The first question to be understood in this problem is the description of all possible limit theorems. This leads to the study of the fixed points of certain operators in the space of distribution functions defined by means of convolution and rescaling. This study leads to the description of the so-called infinitely divisible distributions which appear as the limits in limit theorems. Further investigations make possible also to describe in which limit theorems a certain infinitely divisible distribution appears.

These problems are solved by means of certain methods of the analysis about the study of the behaviour of the characteristic functions, but the proof also contains several probabilistic ideas. We also remark that although there are several different kind of limit theorems, the central limit theorem is the only "universal law" for the limit distribution of normalized partial sums of independent random variables, where the limit distribution function "forgets" the distribution of the single terms in the partial sums. Here we do not explain the more precise meaning of this rather loose statement. A fairly complete discussion with complete proofs of the problems mentioned in this Section 2 can be found together with the explanation of the ideas behind the proofs in the text *Limit theorems and infinitely divisible distribution* on my homepage. (For the time being it exists only in Hungarian.) It also contains the explanation of the statement that the central limit theorem is the only "universal law" among the limit theorems for the limit distribution of partial sums of independent variables.

3.) Let us consider a sequence $\xi_1, \xi_2, \dots,$ of independent, identically distributed random variables with expectation zero and finite variance together with the partial sums $S_n = \sum\limits_{k=1}^n \xi_n$, $n=1,2,\dots$, defined with their help. The central limit theorem gives a good estimate on the probability $P\left(\frac{S_n}{\sqrt{n}} > x\right)$ for large indices n and fixed numbers x. We may ask whether a similar good estimate can be given for the probabilities $P\left(\frac{S_n}{\sqrt{n}} > x_n\right)$, that is we are interested in a good estimate on the probability of a similar event when the number x is replaced by a number x_n which may depend on n. The special case $x_n = x\sqrt{n}$, that is the investigation of the probability of the event that the average of independent, identically distributed random variables is larger than a fixed number x is a particularly important question. This problem belongs to an important part of the probability theory, called the theory of large deviations. The above mentioned problem is discussed together with some additional questions also in a series of problems *The theory of large deviation; the case of partial sums of real valued random variables* on my homepage. (For the time being it exists only in Hungarian). Let us remark that the estimates in the large deviations problems do not agree with the estimates suggested by a formal extension of the central limit theorem.

It would be natural to try to estimate the probability $P(S_n > nx)$ similarly to the estimate of the event $P(S_n > \sqrt{n}x)$ in the proof of the central limit theorem. Nevertheless, this method in itself does not suffice to estimate this probability. It is worth understanding why the method of the proof of the central limit theorem is not sufficient to estimate this probability. Let us first consider the case when the distribution function $P(S_n > nx)$ also has a nice density function $f_n(x) = \frac{d}{dx}P(S_n > nx)$, and let us first estimate this density function. Then we have to investigate the integral $f_n(x) = \frac{n}{2\pi}\int e^{-intx}\varphi^n(t)\,dt$, where $\varphi(t)$ is the characteristic function of the random variable ξ_1. This integral is similar to the integral appearing in the proof of the local central limit theorems. The only difference is that the factor $e^{it\sqrt{n}x}$ in the integral investigated in the proof of the local central limit theorem is replaced by the term $e^{it\sqrt{n}x}$ in this case.

In the proof of the local central limit theorem a singular integral had to be investigated which was strongly concentrated in a small neighbourhood of the origin. In the analogous large deviation problem a similar singular integral has to be investigated. But there is a small difference which makes the latter problem more difficult. The cause of this difficulty is that although the absolute value of the complex number valued integrand whose integral has to be estimated in the large deviation problem has a strongly localized maximum in the origin, but its imaginary part has a strong fluctuation in a small neighbourhood of the origin. This fluctuation causes a strong cancellation, hence we cannot claim that the contribution of a small neighbourhood of the origin yields the main contribution to the integral we are estimating. In the proof of the local central limit theorem this difficulty did not appear, since in this case the fluctuation of the imaginary part of the integrand is not so strong in a small neighbourhood of the origin. The reason for this difference is that the term $e^{-i\sqrt{n}tx}$ appears instead of the term e^{-intx} in the integral investigated in the proof of the local central limit theorem and $\left.\frac{d\varphi(t)}{dt}\right|_{t=0} = E\xi_1 = 0$ in this case.

Such kind of problems appear often in the analysis, and an important technique, the saddle point method was worked out to investigate such problems. To apply the saddle point method let us rewrite the integral we are interested in in the form $f_n(x) = \frac{n}{2\pi} \int e^{n(-itx+\log\varphi(t))}\,dt$. (In this heuristic explanation let us disregard such technical difficulties as the problem that we cannot always take logarithm of a function.) If the function $n(-itx + \log\varphi(t))$ in the exponent of the integrand we study is an analytic function of the variable t, then the saddle-point method suggests to replace the domain of integration to an appropriate new curve which goes through the saddle point, i.e. through a point z which satisfies the identity

$$\frac{d(-izx + \log\varphi(z))}{dz} = 0. \tag{$*$}$$

The integral we are investigating does not change by this replacement. On the other hand this new integral can be better estimated since the absolute value of the integrand on this new curve has a strongly localized (local) maximum in the saddle point, and its fluctuation around the saddle point is small. The proof of the large deviation estimates for averages of independent and identically distributed random variables are based on this idea. Naturally, it is useful to combine it with some additional observations. For instance, it is enough to look for the saddle point on the imaginary axis where the imaginary part of the left-hand side of the saddle point equation $(*)$ is automatically zero. If the saddle point equation has no solution, then an appropriate approximation and some additional ideas are needed to solve the large deviation problems.

Let us remark that in most probability text-books on the theory of large deviations the saddle point method is not discussed. In these text-books the notion of the so-called conjugated distributions are introduced, and the problem is solved with their help. On the other hand, a better understanding of the problem may help us to understand that the introduction of conjugated distributions can be interpreted as the application of the saddle point method to the investigation of large deviation problems formulated in the language of probability theory.

Finally we make the following remark. The above sketched saddle point method can be applied for the investigation of large deviation problems only if the characteristic function of the random variables whose averages we investigate has an analytic continuation. One may ask whether this condition does not mean an unnecessary restriction. A detailed investigation shows that the answer to this question is negative. It turns out that this condition about the analiticity of the characteristic function has a probabilistic content, and the behaviour of the probabilities we investigate in large deviation problems heavily depends on whether the characteristic function is analytic. If this analiticity property does not hold, then the probabilities investigated in the large deviation problems have an essentially different behaviour.

Solutions.

1.) Put $I=\frac{1}{\sqrt{2a\pi}}\int_{-\infty}^\infty e^{-u^2/2a}\,du$. By expressing the number I^2 as a double integral and rewriting it in polar coordinate system we get that

$$\begin{aligned}I^2&=\frac{1}{2a\pi}\int_{-\infty}^\infty\int_{-\infty}^\infty e^{-(u^2+v^2)/2a}\,du\\&=\int_{-\pi}^{\pi}\int_0^\infty\frac{1}{2\pi}\cdot\frac{r}{a}e^{-r^2/2a}\,dr\,d\varphi=\int_{-\pi}^{\pi}\frac{1}{2\pi}\,d\varphi=1.\end{aligned}$$

This relation implies the first statement of the problem.

Let us fix a number $a>0$. Put $F(z)=\frac{1}{\sqrt{2a\pi}}\int_{-\infty}^\infty e^{-(u-z)^2/2a}\,du$. With the substitution $\bar u=u-z$ we get that

$$F(z)=\frac{1}{\sqrt{2a\pi}}\int_{-\infty}^\infty e^{-\bar u^2/2a}\,d\bar u=1$$

for all *real numbers* z. We can generalize this identity for all complex numbers z with the help of one of the following argument of the theory of analytic functions.

First argument: Both functions $F(z)$ and $G(z)\equiv1$ are analytic function on the plane of complex numbers. As they agree for all real numbers they also agree for all complex numbers. To see that the function $F(z)$ is really analytic, observe that it can be obtained as the limit of analytic functions in such a way that the convergence is uniform on all compact subsets of the plane of complex numbers. Then the limit of these analytic functions, i.e. the function $F(z)$ is also analytic. Such analytic functions converging to $F(z)$ can be obtained by approximating the integrals $F(z)$ by the natural approximating sums of these integrals.

Second argument: We get with the change of variable $\bar u=u-z$ that

$$F(z)=\frac{1}{\sqrt{2a\pi}}\int_{-\infty-\operatorname{Im}z}^{\infty-\operatorname{Im}z}e^{-\bar u^2/2a}\,d\bar u=1.$$

To prove the above identity observe that $\lim\limits_{|u|\to\infty}e^{-(u+iv)^2/2a}=0$, and the convergence is uniform with respect to the variable v if $|v|$ is in a bounded interval. This fact together with the result of analytic functions by which the contour integral of an analytic function (containing no singular point) on a closed curve is zero imply that the value of the above integral does not change if we replace the route of integration $[-\infty-\operatorname{Im}z,\infty-\operatorname{Im}z]$ to the line $[-\infty,\infty]$. This implies the identity we wanted to prove.

2.) Let ξ be a random variable with Poissonian distribution with parameter $\lambda=n$, i.e. $P(\xi=k)=\frac{n^k}{k!}e^{-n}$, $k=0,1,2,\dots$. The Fourier series corresponding to the distribution of this random variable ξ equals

$$P_n(t)=\sum_{k=0}^\infty P(\xi=k)e^{itk}=\sum_{k=0}^\infty\frac{n^k}{k!}e^{-n+ikt}=e^{-n+ne^{it}}.$$

This identity together with relation (2) imply relation (3) with the choice $k = n$.

Let us observe that relation (4b) is a simple consequence of relations (3) and (4a) and the asymptotic formula $\frac{1}{1+x} = 1 - x + O(x^2) = 1 + O(x)$, if $|x| \le \frac{1}{2}$.

To prove formula (4a) let us first give an upper bound about the contribution of the domain $\{t\colon\ |t| \ge n^{-1/3}\}$ to the integral at the left-hand side of formula (4a). Then let us consider the restriction of this integral to the domain $\{t\colon\ |t| < n^{-1/3}\}$ and let us give a good asymptotical estimate on it.

To carry out the first estimate let us observe that

$$\left|e^{n(e^{it}-1-it)}\right| = e^{n\mathrm{Re}\,(e^{it}-1-it)} = e^{n(\cos t - 1)} < e^{-nt^2/4} < e^{-n^{1/3}/4} \text{ if } n^{-1/3} \le t \le \pi,$$

and this implies that

$$\left|\int_{\{n^{-1/3} \le |t| \le \pi\}} e^{n(e^{it}-1-it)}\,dt\right| = O\left(e^{-\text{const.}\, n^{1/3}}\right). \tag{2.1}$$

To carry out the second estimation let us give a good asymptotical formula on the integrand of relation (4a) in a small neighbourhood of the origin by means of a Taylor expansion. We get that $n(e^{it} - 1 - it) = n\left(-\frac{t^2}{2} - i\frac{t^3}{6} + O(t^4)\right)$ and

$$\begin{aligned} e^{n(e^{it}-1-it)} &= e^{-nt^2/2} e^{-int^3/6 + O(nt^4)} \\ &= e^{-nt^2/2}\left(1 - \frac{i(\sqrt{n}t)^3}{6\sqrt{n}} + O\left(\frac{(\sqrt{n}t)^4}{n}\right) + O\left(\frac{\sqrt{n}t)^6}{n}\right)\right), \end{aligned}$$

if $|t| \le n^{-1/3}$. By exploiting this estimate and making the substitution of variables $\bar t = \sqrt{n}t$ we get that

$$\begin{aligned} \int_{-n^{-1/3}}^{n^{-1/3}} e^{n(e^{it}-1-it)}\,dt &= \frac{1}{\sqrt{n}} \int_{-n^{1/6}}^{n^{1/6}} e^{-\bar t^2/2}\left(1 - i\frac{\bar t^3}{6\sqrt{n}} + \frac{O\left(\bar t^4 + \bar t^6\right)}{n}\right)d\bar t \\ &= \frac{1}{\sqrt{n}}\left(\int_{-\infty}^{\infty} - \int_{|\bar t| > n^{1/6}}\right). \end{aligned}$$

On the other hand,

$$\int_{|t| \ge n^{1/6}} e^{-t^2/2}\left(1 - i\frac{t^3}{6\sqrt{n}} + \frac{O(t^4 + t^6)}{n}\right)dt = O\left(e^{-n^{1/3}/4}\right),$$

and

$$\begin{aligned} &\int_{-\infty}^{\infty} e^{-t^2/2}\left(1 - i\frac{t^3}{6\sqrt{n}} + \frac{O(t^4 + t^6)}{n}\right)dt \\ &\qquad = \int_{-\infty}^{\infty} e^{-t^2/2}\,dt - \int_{-\infty}^{\infty} \frac{i}{6\sqrt{n}} t^3 e^{-t^2/2}\,dt + O\left(\frac{1}{n}\right) = \sqrt{2\pi} + O\left(\frac{1}{n}\right) \end{aligned}$$

because of the result of the first problem, and since the function $t^3e^{-t^2}$ is odd. These relations imply that

$$\int_{-n^{-1/3}}^{n^{1/3}} e^{n(e^{it}-1-it)}\,dt = \frac{\sqrt{2\pi}}{\sqrt{n}}\left(1+O\left(\frac{1}{n}\right)\right). \tag{2.2}$$

Then relation (4a) is a consequence of formulas (2.1) and (2.2).

Relation (4d) follows similarly from formulas (4c) as formula (4a) implies relation (4b). The only difference is that now we apply a better approximation $\frac{1}{1+x} = 1-x+x^2-\cdots+(-1)^kx^k+O\left(|x|^{k+1}\right)$ for $|x|\le\frac{1}{2}$. Also formula (4c) can be proved similarly to formula (4a). The difference is that now we consider the Taylor expansion of the functions $n(e^{it}-1-it)$ and $P_n(t)$ defined by the identity $e^{n(e^{it}-1-it)} = e^{-nt^2/2}(1+P_n(\sqrt{n}t))$ up to the k-th and not only to the first term in the interval $|t|\le n^{-1/3}$. Since $P_n(t) = \exp\left\{\frac{t^3}{\sqrt{n}}R\left(\frac{t}{\sqrt{n}}\right)\right\}$ with an analytic function $R(t)$ bounded on the real line, these calculations enable us to give an estimate for the integrand of the integral in (4.1) with an accuracy of $e^{-nt^2/2}O\left(\frac{\sum_{j=1}^k(\sqrt{n}|t|)^{j(l(j)}}{n^{(k+1)/2}}\right) = O\left(\frac{e^{-nt^2/4}}{n^{(k+1)/2}}\right)$, where $l(j)=\min\{l\colon\, lj\ge k+1\}$ with a function of the form $e^{-nt^2/2}(1+Q_n(t))$. The function $Q_n(t)$ can be written as $Q_n(t) = \sum\limits_{j=1}^k \frac{\bar{Q}_j(\sqrt{n}t)}{n^{j/2}}$, with some polynomials $\bar{Q}_j$ which do not depend on the parameter n, and they can be calculated explicitly. In particular, $\bar{Q}_1(t) = \frac{-i}{6}t^3$. Then by carrying out the substitution $\bar{t}=\sqrt{n}t$ we get the improvement (4c) of formula (4a).

3.) Let us first show that the distribution of a random variable ξ has the periodicity h if and only if $|Ee^{2\pi i\xi/h}|=1$. Indeed, if $|Ee^{2\pi tei\xi/h}|=1$, then $Ee^{2\pi i\xi/h}=e^{2\pi i/b}$ with some real number b. But this is possible if and only if the distribution of the random variable $2\pi\frac{\xi-b}{h}$ is concentrated in the points $n=0,\pm1,\pm2,\dots$, i.e. the random variable ξ is concentrated on the lattice $nh+b$, $n=0,\pm1,\pm2,\dots$. To see the statement in the converse direction observe that if the distribution of a random variable ξ is concentrated on a lattice $nh+b$, $n=0,\pm1,\dots$, of width h, then $|Ee^{2\pi i\xi}|=1$. Furthermore, if the random variable ξ concentrated on some lattice is not concentrated in a single point, then there are two numbers a and b, $a\ne b$, such that $P(\xi=a)>0$ and $P(\xi=b)>0$. Then we have $|Ee^{itx}|<1$ for all sufficiently small $t>0$, and we can even state that there exists a smallest number $t>0$ such that $|Ee^{it\xi}|=1$.

Take the smallest number $t>0$ such that $|Ee^{it\xi}|=1$. Then $h=\frac{2\pi}{t}$ is the greatest number h such that the distribution of the random variable ξ is concentrated on a lattice $nh+b$, $n=0,\pm1,\pm2,\dots$, of width h. Then the periodicity of the Fourier series $P(t)=Ee^{it(\xi-b)}=\sum\limits_{n=-\infty}^{\infty}P(\xi-b=nh)e^{itnh}$ of the random variable $\xi-b$ is $\frac{2\pi}{h}$. Further, from the definition of the number h follows that $|P(t)|<1$, if

$0 < t < \frac{2\pi}{h}$. Since $P(-t) = P\left(\frac{2\pi}{h} - t\right)$ the above relation can be rewritten as $|P(t)| < 1$, if $0 < |t| < \frac{\pi}{h}$. Clearly, $P(0) = 1$.

Formal differentiation by terms of the infinite sum $P(t)$ yields that $\frac{P^{(k)}(t)}{dt^k} = \sum\limits_{n=-\infty}^{\infty} i^k (nh)^k e^{itnh} P(\xi - b = nh)$, and $\left.\frac{P^{(k)}(t)}{dt^k}\right|_{t=0} = \sum\limits_{n=-\infty}^{\infty} i^k (nh)^k P(\xi - b = nh) = i^k E(\xi - b)^k$. Under the condition $E|\xi - b|^k < \infty$ this formal calculation is allowed. Indeed, the approximating partial sums of the k-th derivative of the function $P(t)$ satisfies the inequality $\sum\limits_{n=-N}^{N} |i^k (nh)^k e^{itnh} P(\xi - b = nh)| \le E|\xi - b|^k$, and this property allows (k-fold) differentiation by terms.

4.) By relation (2) and the formula written after it

$$P(S_n = k) = \int_{-\pi}^{\pi} \frac{1}{2\pi} e^{-ikt} P^n(t)\, dt = \int_{|t| < \frac{1}{\varepsilon\sqrt{n}}} + \int_{\frac{1}{\varepsilon\sqrt{n}} < |t| < \varepsilon} + \int_{\varepsilon < |t| < \pi} = I_1 + I_1 + I_3,$$

where $P(t) = \sum\limits_{k=-\infty}^{\infty} e^{ikt} P(\xi_1 = k)$, and $\varepsilon > 0$ is an arbitrary (small) positive real number. We solve the problem if we give a good estimate for the integrals I_1, I_2 and I_3 for small $\varepsilon > 0$.

It is easy to bound the integral I_3. It follows from the result of problem 3 and the continuity of the function $P(t)$ that $\sup\limits_{\varepsilon \le |t| < \pi} |P(t)| < 1$ with some number $0 < q = q(\varepsilon) < 1$. Hence

$$|I_3| = \left| \int_{\varepsilon < |t| < \pi} \frac{1}{2\pi} e^{-ikt} P^n(t)\, dt \right| < q^n$$

with some number $0 < q < 1$. To estimate the integrals I_1 and I_2 we need a good bound on the function $P^n(t)$ if $|t| < \varepsilon$. It is more convenient to work with the function $\log P(t)$ instead. (For small numbers $\varepsilon > 0$ this is allowed, since in this case the value of the function $P(t)$ in the interval $[-\varepsilon, \varepsilon]$ is in a small neighbourhood of the number 1.) Simple calculation yields that

$$\frac{d \log P(t)}{dt} = \frac{P'(t)}{P(t)}, \qquad \left.\frac{d \log P(t)}{dt}\right|_{t=0} = im,$$
$$\frac{d^2 \log P(t)}{dt^2} = \frac{P''(t)P(t) - P'(t)^2}{P^2(t)}, \qquad \left.\frac{d^2 \log P(t)}{dt^2}\right|_{t=0} = -m_2 + m^2 = -\sigma^2,$$

hence a Taylor expansion around the origin yields that

$$\log P(t) = imt - \frac{\sigma^2}{2} t^2 + o(t^2), \quad \text{if } |t| < \varepsilon.$$

This estimate together with the relation $|P^n(t)| = e^{n\mathrm{Re}\,\log P(t)} = e^{-n\sigma^2 t^2/2 + no(t^2)} \le e^{-n\sigma^2t^2/3}$, if $|t| < \varepsilon$ and $n \ge n(\varepsilon)$ imply that

$$|I_2| \le \frac{1}{2\pi} \int_{\frac{1}{\varepsilon\sqrt{n}} < |t| < \varepsilon} |P^n(t)|\,dt \le \frac{1}{2\pi} \int_{\frac{1}{\varepsilon\sqrt{n}} < |t| < \varepsilon} e^{-n\sigma^2t^2/3}\,dt$$
$$\le \frac{1}{\pi\sqrt{n}} \int_{-\frac{1}{\varepsilon}}^{\infty} e^{-\sigma^2t^2/3}\,dt \le \frac{e^{-\sigma^2/4\varepsilon^2}}{\sqrt{n}}.$$

Furthermore,

$$\begin{aligned}
I_1 &= \int_{-\frac{1}{\varepsilon\sqrt{n}}}^{\frac{1}{\varepsilon\sqrt{n}}} \frac{1}{2\pi} e^{-ikt+inmt-n\sigma^2t^2/2+o(nt^2)}\,dt \\
&= \int_{-\frac{1}{\varepsilon}}^{\frac{1}{\varepsilon}} \frac{1}{2\pi\sqrt{n}} e^{i(mn-k)t/\sqrt{n}-\sigma^2t^2/2+o(t^2)}\,dt \\
&= \int_{-\frac{1}{\varepsilon}}^{\frac{1}{\varepsilon}} \frac{1}{2\pi\sqrt{n}} e^{i(mn-k)t/\sqrt{n}-\sigma^2t^2/2}(1+o(t^2))\,dt \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{n}} e^{i(nm-k)t/\sqrt{n}-\sigma^2t^2/2}\,dt - \int_{|t|>\frac{1}{\varepsilon}} \frac{1}{2\pi\sqrt{n}} e^{i(nm-k)t/\sqrt{n}-\sigma^2t^2/2}\,dt \\
&\qquad + o\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

On the other hand

$$\left| \int_{|t|>\frac{1}{\varepsilon}} \frac{1}{2\pi\sqrt{n}} e^{i(nm-k)t/\sqrt{n}-\sigma^2t^2/2}\,dt \right| \le \frac{e^{-\sigma^2/4\varepsilon^2}}{\sqrt{n}},$$

and by completing the quadratic term in the exponent of the last formula we get that

$$\int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{n}} e^{i(nm-k)t/\sqrt{n}-\sigma^2t^2/2}\,dt$$
$$= \frac{e^{-(nm-k)^2/2n\sigma^2}}{2\pi\sqrt{n}} \int_{-\infty}^{\infty} \exp\left\{ -\frac{\sigma^2}{2}\left(t - i\frac{nm-k}{\sqrt{n}}\right)^2 \right\} dt = \frac{e^{-(nm-k)^2/2n\sigma^2}}{\sqrt{2\pi n}\sigma}$$

by the result of problem 1. These estimates imply that

$$\left| I_1 - \frac{1}{\sqrt{2\pi n}\sigma} \exp\left\{ -\frac{(k-nm)^2}{2n\sigma^2} \right\} \right| \le \text{const.}\, \frac{e^{-\sigma^2/4\varepsilon^2}}{\sqrt{n}}$$

if $n > n(\varepsilon)$. As the estimates given for the expressions I_1, I_2 and I_3 hold for all $\varepsilon > 0$ if n is sufficiently large, hence they imply the result of problem 4.

5.) The solution of this problem is similar to that of problem 4. Since the random variable ξ_1 has three finite moments in the present case, hence we can make the following better approximation of the function $\log P(t)$: $\log P(t) = imt - \frac{\sigma^2}{2}t^2 + O(t^3)$. Hence $P^n(t) = e^{imnt - n\sigma^2 t^2/2 + O(nt^3)}$. Then we can solve problem 5 by making estimations similar to those in the proof of problem 4. The only essential difference is that now we define the domain of integration in the definition of the expressions in I_1 and I_2 in a different way. Now put $I_1 = \int_{|t|<n^{-1/3}}$ and $I_2 = \int_{n^{-1/3}\le|t|<\varepsilon}$. The reason the domain of integration in the definition of expression I_1 was chosen in the above way is that in the domain is that $e^{O(nt^3)} = 1 + O(nt^3)$ in the domain $\{|t| < n^{-1/3}\}$, hence the approximation $e^{-ikt}P^n(t) = e^{i(mn-k)t - n\sigma^2 t^2/2}\left(1 + O(nt^3)\right)$ holds in this domain. Then by applying a natural adaptation of the calculation in problem 4 that by omitting the error term $O(nt^3)$ from the integral approximating the expression I_1 we get an error of order $O\left(\frac{1}{n}\right)$. The error of all other remaining estimations is essentially smaller. They are of order $e^{O(-\text{const.}\, n^{1/3})}$. Hence these estimates yield the solution of problem 5.

5a.) We have to estimate the expression $P(S_n = k) = \binom{n}{k}p^k(1-p)^{n-k}$. Let us first consider the case $\alpha n < k < \beta n$ with some numbers $0 < \alpha < \beta < 1$. By the Stirling formula

$$
\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} = \frac{\left(\frac{n}{e}\right)^n}{\left(\frac{n-k}{e}\right)^{n-k}\left(\frac{k}{e}\right)^k}\sqrt{\frac{n}{2\pi k(n-k)}}\left(1 + O\left(\frac{1}{n}\right)\right) \\
&= \left(1 - \frac{k}{n}\right)^{-(n-k)}\left(\frac{k}{n}\right)^{-k}\frac{1}{\sqrt{2\pi n}\sqrt{\frac{k}{n}\left(1-\frac{k}{n}\right)}}\left(1 + O\left(\frac{1}{n}\right)\right).
\end{aligned}
$$

A Taylor expansion of the function $\log x$ around the point p yields that

$$
\begin{aligned}
p^k\left(\frac{k}{n}\right)^{-k} &= \exp\left\{k\left(\log p - \log\frac{k}{n}\right)\right\} \\
&= \exp\left\{-\frac{k}{p}\left(\frac{k}{n} - p\right) + \frac{k}{2p^2}\left(\frac{k}{n} - p\right)^2 + O\left(n\left(\frac{k}{n} - p\right)^3\right)\right\}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&(1-p)^{n-k}\left(\frac{n-k}{n}\right)^{-(n-k)} \\
&\quad = \exp\left\{\frac{n-k}{1-p}\left(\frac{k}{n} - p\right) + \frac{n-k}{2(1-p)^2}\left(\frac{k}{n} - p\right)^2 + O\left(n\left(\frac{k}{n} - p\right)^3\right)\right\}.
\end{aligned}
$$

By multiplying the last two expressions and exploiting that in the formula got in such a way the coefficient of the term $\left(\frac{k}{n} - p\right)^2$ is

$$
\frac{k}{2p^2} + \frac{n-k}{2(1-p)^2} = \frac{(k-np)^2}{2p(1-p)} + (1-2p)\frac{(k-np)^3}{n^2}
$$

we get that

$$p^k\left(\frac{k}{n}\right)^{-k}(1-p)^{n-k}\left(\frac{n-k}{n}\right)^{-(n-k)}=\exp\left\{-\frac{(kn-p)^2}{2np(1-p)}+O\left(n\left(\frac{k}{n}-p\right)^3\right)\right\}.$$

Since $\frac{k}{n}\left(1-\frac{k}{n}\right)=p(1-p)\left(1+O\left(\frac{k-np}{n}\right)\right)$ the above calculations yield that

$$P(S_n=k)=\frac{\exp\left\{-\frac{(kn-p)^2}{2np(1-p)}+O\left(n\left(\frac{k}{n}-p\right)^3\right)+O\left(\frac{k}{n}-p\right)+O\left(\frac{1}{n}\right)\right\}}{\sqrt{2\pi p(1-p)n}}$$

Since $E\xi_1=p$, $\operatorname{Var}\xi_1=p(1-p)$ the last formula yields an estimate stronger than we want in the case $|k-np|<\gamma n$ with a sufficiently small number $\gamma>0$. Indeed, by introducing the quantity $z=\frac{k-np}{\sqrt{n}}$ we get the demanded estimate with an error term $\varepsilon(n)$ of the following form:

$$\varepsilon(n)=\varepsilon(n,z)\le\frac{C}{\sqrt{n}}e^{-K_1z^2}\left[\exp\left\{K_2\frac{|z|^3}{\sqrt{n}}+K_3\frac{|z|}{\sqrt{n}}+\frac{K_4}{n}\right\}-1\right]$$

with appropriate constants $C>0$, and $K_j>0$, $j=1,\dots,4$. Then we have to show that $\varepsilon(z,n)\le\frac{\text{const.}}{n}$, if $|z|\le\gamma\sqrt{n}$. This estimate holds for $n^{1/6}>|z|<\gamma\sqrt{n}$, since in this case $\varepsilon(n,z)\le e^{-K_1z^2/2}$. If $|z|\le n^{1/6}$ then $e(n,z)\le\frac{C}{\sqrt{n}}e^{-K_1z^2}\left(\frac{|z|^3+|z|+1}{\sqrt{n}}\right)\le\frac{\text{const.}}{n}$, that is the demanded estimate holds also in this case.

If $|k-np|\ge\gamma n$, then the result of problem 5a follows from the relations $P(S_n=k)\le\frac{\text{const.}}{n}$ and $e^{-(k-np)^2/2np(1-p)}<\frac{\text{const.}}{n^2}$. Actually, even stronger estimates could be proved. The first estimate is a consequence of the Chebishev inequality, since $P(S_n=k)\le P(|S_n-ES_n|\ge\gamma n)\le\frac{\operatorname{Var}S_n}{\gamma^2n^2}\le\frac{\text{const.}}{n}$. The second inequality is obvious.

6.) Let us introduce the random variables $\bar\xi_j=\frac{\xi_j-b}{h}$, $j=1,\dots,n$, and $\bar S_n=\sum\limits_{j=1}^n\bar\xi_j$. Then $E\bar\xi_j=\frac{m-b}{h}$ and $\operatorname{Var}\bar\xi_j=\frac{\sigma^2}{h^2}$. Since $P(S_n=kh+nb)=P(\bar S_n=k)$, and the random variable $\bar S_n$ is concentrated on the lattice of the integers as on the rarest lattice, statement formulated in this problem follows from the results of problems 4 and 5.

7.) It follows from formula (5) that the relation

$$\lim_{n\to\infty}P\left(A<\frac{S_n-nm}{\sqrt{n}\sigma}<B\right)=\int_A^B\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du,\tag{2.3}$$

holds if $-\infty<A<B<\infty$, and the limit in this relation is uniform for all pairs of numbers (A,B) such that $C_1\le A<B\le C_2$ with some fixed numbers

$-\infty < C_1 < C_2 < \infty$. Indeed,

$$
\begin{aligned}
P\left(A < \frac{S_n - nm}{\sqrt{n}\sigma} < B\right) &= P(A\sqrt{n}\sigma + nm - nb < S_n - nb < B\sqrt{n}\sigma + nm - nb) \\
&= \sum_{k\colon\, k\in\mathcal{K}(A,B)} P(S_n = kh + nb) \\
&= \frac{h}{\sqrt{2\pi n}\sigma} \sum_{k\colon\, k\in\mathcal{K}(A,B)} \exp\left\{-\frac{(kh+nb-nm)^2}{2n\sigma^2}\right\} + \sqrt{n}o\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{h}{\sqrt{n}\sigma} \sum_{l(k,n)\in\mathcal{L}(A,B)} \frac{1}{\sqrt{2\pi}} e^{-l(k,n)^2/2} + o(1)
\end{aligned}
$$

where $\mathcal{K}(A,B) = \{k\colon\ A\sqrt{n}\sigma < kh + nb - nm > B\sqrt{n}\sigma\}$, $l(k,n) = \frac{kh-nm+nb}{\sqrt{n}\sigma}$, and

$$
\mathcal{L}(A,B) = \left\{ l(k,n) = \frac{kh - nm + nb}{\sqrt{n}\sigma},\ k = 0, \pm 1, \pm 2, \dots \right\} \cap (A,B),
$$

i.e. the points of the set $\mathcal{L}(A,B)$ are the points of the lattice of width $\frac{h}{\sqrt{n}\sigma}$ and containing the point $\frac{nb-nm}{\sqrt{n}\sigma}$ which fall into the interval (A,B). This implies formula (2.3), since for a fixed number n the probability at the left-hand side is an approximating sum of the integral at the right-hand side plus an error which tends to zero as $n\to\infty$.

We prove that relation (2.3) also holds for $A = -\infty$. Indeed, for all $\varepsilon > 0$ we can choose a number $K = K(\varepsilon)$ such that $\int_{-K}^{K} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du > 1 - \varepsilon$. Then $\lim\limits_{n\to\infty} P\left(\left|\frac{S_n - nm}{\sqrt{n}\sigma}\right| < K\right) > 1 - \varepsilon$, and $\lim\limits_{n\to\infty} P\left(\frac{S_n-nm}{\sqrt{n}\sigma} < -K\right) < \varepsilon$. Then

$$
\begin{aligned}
&\limsup_{n\to\infty} \left| P\left(\frac{S_n - nm}{\sqrt{n}\sigma} < x\right) - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \right| \\
&\qquad \le \limsup_{n\to\infty} \left| P\left(-K \le \frac{S_n - nm}{\sqrt{n}\sigma} < x\right) - \int_{-K}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du \right| + \varepsilon \le \varepsilon.
\end{aligned}
$$

Since the last relation holds for all $\varepsilon > 0$, it implies the statement of problem 7.

8.) Let $\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx = Ee^{it\xi_1}$ denote the Fourier transform of the density function of the random variable ξ_1. Then the Fourier transform of the density function of the random sum $S_n = \xi_1 + \cdots + \xi_n$ equals $Ee^{it(\xi_1+\cdots+\xi_n)} = \left(Ee^{it\xi_1}\right)^n = \varphi^n(t)$. Since $|\varphi(t)| \le 1$, hence under the conditions of problem (8) the function $\varphi^n(t)$ is integrable for $k \ge n$, and the density function $f_n(t)$ of the random sum can be expressed by formula (6) if we replace the function $\varphi(t)$ by $\varphi^n(t)$. Moreover, this relation also holds if we only assume that the function $\varphi^k(t)$ is integrable, and $n \ge k$. The above calculation makes possible to prove problem 8 similarly to problem 4 with some natural modification. Now we have to estimate the integral

(6) instead of the integral (2) (with the modification that we write $\varphi^n(t)$ instead of the function $\varphi(t)$ in (6). Further, because of the condition a $E\xi^2<\infty$ the Fourier transform $\varphi(t)$ is twice differentiable, $\varphi'(0)=iE\xi_1$, $\varphi(0)''=-E\xi_1^2$. This means that the analogs of the relations applied in the solution of problem 4 holds in this case. (Later we shall discuss the properties of the Fourier transform $\varphi(t)$ in the general case.)

The only essential difference in the estimation of the integral we have to investigate is that the integral $I_1=\int_{\varepsilon<|t|<\pi}e^{-ikt}P^n(t)\,dt$ introduced in the solution of problem 4 now we write $I_1'=I_1'(x)=\int_{\varepsilon<|t|<\infty}e^{-itx}\varphi^n(t)\,dt$. Observe that

$$
\begin{aligned}
I_1' \le \int_{\varepsilon<|t|<\infty} |\varphi(t)|^n\,dt &\le \sup_{\varepsilon\le|t|<\infty}|\varphi(t)|^{n-k}\int_{\varepsilon<|t|<\infty}|\varphi(t)|^k\,dt\\
&\le \text{const.}\sup_{\varepsilon\le|t|<\infty}\varphi(t)|^{n-k},
\end{aligned}
$$

since $\varphi^k(\cdot)$ is an integrable function. For a fixed number t, $t\ne0$, $|\varphi(t)|<1$. Further, by the Riemann lemma $\lim\limits_{|t|\to\infty}|\varphi(t)|=0$, and $\varphi(t)$ is a continuous function. (This series of problem also contains the proof of the Riemann lemma.) These facts imply that $\sup\limits_{\varepsilon<|t|<\infty}|\varphi(t)|<q<1$. This relation together with the previous estimates imply that $|I_1'|\le\text{const.}\,q^n$. The only further difference in the solution of problem 8 when compared to the solution of problem 4 is that now in the integrals defining the expressions I_1 and I_2 the functions $e^{-ikt}P^n(t)$ are replaced by the function $e^{-itx}\varphi^n(t)$. These expressions can be estimated in the same way as the analogous integrals in problem 4.

9.) The solution of this problem is similar to that of problem 6, only the notations are simpler. The condition directly implies that

$$
\lim_{n\to\infty}(F_n(B)-F_n(A))=\int_A^B\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du,
$$

and the convergence is uniform for pairs of numbers (A,B) in a bounded set. Then we can show similarly to the argument in problem 6 that the number A in the above formula can be replaced by $-\infty$.

10.) The proof is a slight modification of the proof sketched in the solution of problem 8. This modification is similar to the modification made in the solution of problem 5 compared to problem 4. Since $E|\xi_1|^3<\infty$, hence the approximation

$$
\log\varphi(t)=itE\xi_1-\frac{t^2}{x}E\xi_1^2+O(t^3)
$$

holds in a small neighbourhood of the origin. This makes possible to get the solution of the problem by modifying the domain of integration in the expressions I_2 and I_3 as it was done in the solution of problem 5.

11.) Let ξ be a random variable with standard normal distribution. Then

$$E\xi=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}ue^{-u^2/2}\,du=0,$$

since the integrand is an odd function. On the other hand, integration by parts yields that

$$\begin{aligned}E\xi^2&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}u^2e^{-u^2/2}\,du=-\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}u\left(\frac{d}{du}e^{-u^2/2}\right)\,du\\&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du=1.\end{aligned}$$

If ξ is a $\Phi_{m,\sigma}$ distributed random variable, then its transform $\frac{\xi-m}{\sigma}$ has standard normal distribution, i.e. it has expectation zero and variance 1. Hence ξ has expectation m and variance σ^2.

12.) Let $\varphi(t)=\varphi(t_1,\dots,t_k)=Ee^{i(t,\xi)}=Ee^{i(t_1\xi_1+\cdots+t_k\xi_k)}$ denote the characteristic function of a k-dimensional random vector $\xi=(\xi_1,\dots,\xi_k)$, where $t=(t_1,\dots,t_k)$, and $(t,\xi)=\sum\limits_{j=1}^{k}t_j\xi_j$. Then $|\varphi(t)|\le E|e^{i(t,\xi)}|=1$. For arbitrary number $\varepsilon>0$ there exists a constant $R=R(\varepsilon)$ such that $P(|\xi|>R)=P\left(\sum\limits_{j=1}^{k}\xi_j^2>R^2\right)<\frac{\varepsilon}{2}$. Put $\delta=\frac{\varepsilon}{2R(\varepsilon)}$ and consider such numbers $t=(t_1,\dots,t_k)$ for which $|t|^2=\sum\limits_{j=1}^{k}t_j^2<\delta$. Then $|e^{i(t,\xi)}-1|\le|(t,\xi)|\le\frac{\varepsilon}{2}$ for $|x|<R(\varepsilon)$, and $|\varphi(t)-\varphi(\bar t)|=|Ee^{i(t,\xi)}-Ee^{i(\bar t,\xi)}|\le E|e^{i(t-\bar t,\xi)}-1|\le E|e^{i(t-\bar t,\xi)}-1|I(|\xi|\le R)+P(|\xi|>R)\le E|(t-\bar t,\xi)|I(|\xi|\le R)+\frac{\varepsilon}{2}\le\varepsilon$ if $|t-\bar t|\le\delta$, where $I(A)$ denotes the indicator function of a set A. Hence the function $\varphi(t)$ is uniformly continuous.

The characteristic function of a random vector $a\xi+m$ in a point $t\in R^k$, where $a\in R$, $m\in R^k$ is the function $Ee^{i(t,a\xi+m)}=e^{(it,m)}Ee^{i(at,\xi)}=e^{(it,m)}\varphi(at)$, where φ denotes the characteristic function of the random vector ξ.

If ξ_j, $j=1,\dots,n$, are independent random vectors with characteristic functions $\varphi_j(t)$, then the characteristic function of the random sum $\xi_1+\cdots+\xi_n$ in a point $t\in R^k$ equals $Ee^{i(t,\xi_1+\cdots+\xi_n)}=Ee^{i(t,\xi_1)}\dots e^{i(t,\xi_n)}=Ee^{i(t,\xi_1)}\dots Ee^{i(t,\xi_n)}=\prod\limits_{j=1}^{k}\varphi_j(t)$.

13.) a.) If the random variable ξ has standard normal distribution, then

$$Ee^{it\xi}=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{itu-u^2/2}\,du=e^{-t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(it-u)^2/2}\,du=e^{-t^2/2}$$

by the result of problem 1.

b.) If the random variable ξ has uniform distribution in the interval $[0,1]$, then

$$Ee^{it\xi} = \int_0^1 e^{itu}\,du = \frac{e^{it}-1}{it}.$$

c.) If the random variable ξ has exponential distribution with parameter $\lambda>0$, then

$$Ee^{it\xi} = \int_0^\infty \lambda e^{itu-\lambda u}\,du = \frac{\lambda}{\lambda-it}.$$

d.) If ξ is a random variable with Cauchy distribution, then

$$Ee^{it\xi} = \int_{-\infty}^\infty \frac{1}{\pi}\frac{e^{itu}}{1+u^2}\,du.$$

This integral can be calculated by means of the residium theorem in the theory of analytic functions.

The function $g(z)=g(z,t)=\frac{e^{itz}}{\pi(1+z^2)}$ is analytic in the plane of complex numbers with two poles $z=\pm i$. The residium of the function $g(z)$ in the point i equals e^{-t}, and in the point $-i$ it equals e^t. Let us consider the following contour integral. Let us first integrate the function $g(z)=g(z,t)$ on the interval $[-R,R]$ and then on the half-circle $|z|=R$, $\operatorname{Im} z\ge 0$ if $t\ge 0$ and on the half-circle $|z|=R$, $\operatorname{Im} z\le 0$ if $t\le 0$. The above contour integral equals the residium of the function $g(z)$ in the point i if $t>0$ and the residium of this function in the point $-i$ if $t<0$. On the other hand the restriction of the integral to the half-circle of radius R tends to zero if $R\to 0$. This implies that $Ee^{it\xi}=\int_{-\infty}^\infty g(t,u)\,du = e^{-|t|}$.

The following argument gives another different a little bit artificial but correct proof of this statement. The characteristic function of the density function $f(x)=\frac12 e^{-|x|}$ equals

$$\frac12\int_{-\infty}^\infty e^{-|x|+itx}\,dx = \frac12\left(\frac{1}{1+it}+\frac{1}{1-it}\right) = \frac{1}{1+t^2}.$$

Since the function $\frac{1}{1+t^2}$ is integrable the inverse Fourier transformation formula (6) implies the desired statement.

e.) If the random variable ξ has Poissonian distribution with parameter $\lambda>0$, then

$$Ee^{it\xi} = e^{-\lambda}\sum_{k=0}^\infty \frac{\lambda^k}{k!}e^{ikt} = \exp\left\{\lambda(e^{it}-1)\right\}.$$

f.) If the random variable ξ has binomial distribution with parameters n and p, then

$$Ee^{it\xi} = e^{-\lambda}\sum_{k=0}^n \binom{n}{k} p^k e^{ikt}(1-p)^{n-k} = \left(1-p+pe^{it}\right).$$

g.) If ξ is a random variable with negative binomial distribution with parameters n and p, then its distribution agrees with the distribution of the random sum $\xi_1 + \cdots + \xi_n$, where ξ_j, $1 \le j \le n$, are independent random variables with negative binomial distribution with parameters 1 and p. (To see this property observe that a possible the random variable ξ has the following probabilistic interpretation: If we make independent experiment after each other which are successful with probability p, then ξ denotes the number of unsuccessful experiments up to the n-th successful experiment. If ξ_j denotes the number of the unsuccessful experiments between the $j-1$-th and j-th successful experiments, then we get the above representation.) Hence $Ee^{it\xi} = \left(Ee^{it\xi_1}\right)^n$. On the other hand,

$$Ee^{it\xi_1} = \sum_{k=0}^{\infty}(1-p)p^k e^{itk} = \frac{1-p}{1-pe^{it}}.$$

h.) We get with the change of variables $\bar u = (1-it)u$ that

$$\varphi_s(t) = \frac{1}{\Gamma(s)}\int_0^\infty e^{-u+itu}u^{s-1}\,du = \frac{1}{\Gamma(s)}\frac{1}{1-it)^s}\int_0^\infty e^{-\bar u}\bar u^{s-1}\,d\bar u = \frac{1}{(1-it)^s}.$$

In this calculation we applied some complex analysis argument. At the change of variables step $\bar u = (1-it)u$ of the calulation the domain of integration became the line $(1-it)u$, $u>0$, instead of the positive abscissa axis. But we can turn back the domain of integration to the positive abscissa axis by means of a usual complex analysis argument, by which the integral of an analytic function on a closed curve equals zero. At this step we have to exploit that the function e^{-z} tends to zero fast as $\operatorname{Re} z \to \infty$.

14.) If $f(x_1,\dots,x_k)$ and $g(x_1,\dots,x_k)$ are two integrable functions, then

$$\begin{aligned}
\infty &> \int\int |f(x_1,\dots,x_k)||g(u_1,\dots,u_k)|\,dx_1\cdots dx_k du_1\dots du_k \\
&\qquad\qquad\qquad\text{(with substitution } \bar u_j = x_j + u_j\ j=1,\dots,k,) \\
&= \int\int |f(x_1,\dots,x_k)||g(\bar u_1 - x_1,\dots,\bar u_k - x_k)|\,dx_1\cdots\,dx_k\,d\bar u_1\dots\,d\bar u_k \\
&= \int\left(\int |f(x_1,\dots,x_k)||g(\bar u_1 - x_1,\dots,\bar u_k - x_k)|\,dx_1\cdots dx_k\right)d\bar u_1\dots du_k \\
&= \int |f|*|g|(x_1,\dots,x_k)\,dx_1\dots dx_k.
\end{aligned}$$

This relation implies that the function $|f*g(x_1,\dots,x_k)| \le |f|*|g|(x_1,\dots,x_k)$ is bounded in almost all points $(x_1,\dots,x_k)\in R^k$. It also implies that $f*g$ is an integrable function. In the sequel we write x instead of $(x_1,\dots,x_k)$ and u instead of $(u_1,\dots,u_k)$.

If μ and ν are two measures of bounded variation, then there exists a representation $\mu=\mu_1-\mu_2$, $\nu=\nu_1-\nu_2$ such that μ_i and ν_i, $i=1,2$, are finite measures. The identity $\mu*\nu=(\mu_1*\nu_1+\mu_2*\nu_2)-(\mu_1*\nu_2+\mu_2*\nu_1)$ holds. Since $\mu_i*\nu_j(R^k)<\infty$ for all indices $i,j=1,2$, this implies that $\mu*\nu$ is a measure of bounded variation.

If the measure μ has a density function f, then we have for all measurable sets $A\subset R^k$

$$\begin{aligned}\int_A f*\nu(x)\,dx&=\int_A\left(\int f(u)\nu(x-\,du)\right)\,dx=\int_A\left(\int f(x-u)\nu(\,du)\right)\,dx\\&=\int\left(\int_A f(x-u)\,dx\right)\nu(\,du)=\int\left(\int I(x\colon\ x\in A)f(x-u)\,dx\right)\nu(\,du)\\&=\int\left(\int I(v\colon\ u+v\in A)f(v)\,dv\right)\nu(\,du)\\&=\int\int I(v\colon\ u+v\in A)\mu(\,dv)\nu(\,du)\\&=\mu\times\nu\left(\{(u,v)\colon\ u+v\in A\}\right)=\mu*\nu(A)\end{aligned}$$

and this means that the function $f*\nu$ is the density function of the convolution $\mu*\nu$ of the measures μ and ν.

Let us observe that the above calculations also imply that the function $f*\nu(x)$ is finite in almost all points $x\in R^k$, moreover it is integrable. Indeed, the above calculation implies this result with the choice $A=R^k$ if μ and ν are (bounded) positive measures. The general case can be reduced to this case if we decompose the measures μ and ν as the difference of two positive finite measures. (We may assume that the measures in the decomposition $\mu=\mu_1-\mu_2$ have density function.)

If the measure μ has a density function f and the measure ν has a density function g then let us define the measures $\bar\nu_x(A)=\nu(x-A)$ for all $x\in R^k$. The density function of the measure $\bar\nu_x(A)$ equals $g(x-u)$ in the point $u\in R^k$, and the density function of the measure $\mu*\nu$ in the point x equals

$$\int f(u)\bar\nu_x(du)=\int f(u)g(x-u)\,du=f*g(x)$$

by the already proved results.

15.) It follows from the definition of the convolution that if ξ an η are independent random variables with distributions μ and ν, then the distribution of the random sum $\xi+\eta$ equals $\mu*\nu$. It follows from the result of the previous problem that if the distribution μ of the random variable ξ has a density function f, then the distribution $\mu*\nu$ of the random variable $\xi+\eta$ has a density function $f*\nu$. If also the measure ν has a density function g, then this density function equals $f*g$.

If the distribution of the random variable Z is $F(x)$, then the distribution of the random variable $\bar Z=\frac{Z-A}{B}$ with $B>0$ equals $F(Bx+A)$. If the random variable Z has a density function $f(x)$, then the random variable $\bar Z$ has a density function $Bf(Bx+A)$. The previous results imply the statements formulated for the distribution and density function of the random sum $\bar S_n$.

The relation $\mu * \nu = \nu * \mu$ follows from the definition of the convolution. The statement $(\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3)$ follows from the fact that the identity

$$(\mu_1 * \mu_2) * \mu_3(A) = \mu_1 * (\mu_2 * \mu_3)(A) = \mu_1 \times \mu_2 \times \mu_3(\{(u, v, w)\colon\ u + v + w \in A\})$$

holds for all measurable sets A. The analog statements about the convolution of functions can be reduced to these statement if we represent the convolution of functions as the convolution of the density function of the corresponding signed measures. Otherwise these statement also can be simply proved by simple calculation.

16.) If μ and ν are two signed measures of bounded variation with Fourier transforms $\tilde{f}(t_1, \dots, t_k)$ and $\tilde{g}(t_1, \dots, t_k)$, then

$$\begin{aligned} &\tilde{f}(t_1, \dots, t_k)\tilde{g}(t_1, \dots, t_k) \\ &\qquad = \int e^{i(t_1(u_1+v_1)+\cdots+t_k(u_k+v_k))} \mu(\,du_1, \dots,\, du_k)\nu(\,dv_1, \dots,\, dv_k). \end{aligned}$$

The transformation $\mathbf{T}(u_1, \dots, u_k, v_1, \dots, v_k) = (u_1+v_1, \dots, u_k+v_k)$, $(u_1, \dots, u_k) \in R^k$, $(v_1, \dots, v_k) \in R^k$, is a measurable transformation from the space $(R^k \times R^k, \mathcal{B}_{2k}, \mu \times \nu)$ to the space $(R^k, B_k, \mu * \nu)$, where $\mathcal{B}_{2k}$ and $\mathcal{B}_k$ denote the σ-algebras in the spaces R^{2k} and R^k. By applying the measure theoretical result which describe how measurable transformations transform integrals for the functions

$$h(x_1, \dots, x_k) = e^{it(x_1+\cdots+x_k)})$$

and

$$g(u_1, \dots, u_k, v_1, \dots, v_k) = h(\mathbf{T}(u_1, \dots, u_k, v_1, \dots, v_k)) = e^{i(t_1(u_1+v_1)+\cdots+t_k(u_k+v_k))}$$

with the above defined transformation $\mathbf{T}(u_1, \dots, u_k, v_1, \dots, v_k)$ we get from the relation written at the beginning of the solution that

$$\tilde{f}(t_1, \dots, t_k)\tilde{g}(t_1, \dots, t_k) = \int e^{i(t_1x_1+\cdots+t_kx_k)} \mu * \nu(\,dx_1, \dots,\, dx_k).$$

This implies the statement about the Fourier transform of the convolution of measures. The analogous statement about the Fourier transform of the convolution of density functions follows from this statement and the relation between the convolution of measures and their density functions.

17.) By differentiating the identity $f * g(x) = \int f(x-u)g(u)\,du$ k times we get that

$$\frac{df * g^k(x)}{dx^k} = \int \left.\frac{df^k(v)}{dv^k}\right|_{v=x-u} g(u)\,du = \int \left.\frac{df^k(v)}{dv^k}\right|_{v=u} g(x-u)\,du.$$

The conditions of the problem allow the above successive differentiations. Further, working with the right-hand side of the last formula we can carry out l additional differentiations and get that

$$\frac{df * g^{k+l}(x)}{dx^{k+l}} = \int \left.\frac{df^k(v)}{dv^k}\right|_{v=u} \left.\frac{dg^l(v)}{dv^l}\right|_{v=x-u} du.$$

If $f(u)$ is analytic function which also satisfies the other conditions of the problem, then the function

$$F(z) = \int f(z-u)g(u)\,du$$

is an analytic continuation of the convolution $f*g(x)$ to the domain $\{z\colon \operatorname{Im} z < A\}$.

18.) a.) Convergence in distribution implies the convergence of the integrals:

Since $F(x_1,\dots,x_k) \to 1$ if $x_j \to \infty$ for all $j=1,\dots,k$, and $F(x_1,\dots,x_k) \to 0$ if $x_j \to -\infty$ for one of the indices $1 \le j \le k$ hence for all $\varepsilon > 0$ there exists a k-dimensional rectangle $\mathbf{K} = \mathbf{K}(\varepsilon)$ such that $\mu_F(\mathbf{K}) > 1-\varepsilon$. (Given a distribution function F in the sequel we shall denote by μ_F the probability measure on R^k induced by the distribution function F.) We also may assume, by enlarging the rectangle $\mathbf{K}$ if it is necessary, that the boundary of the rectangles $\mathbf{K}$ has zero μ_F measure. Indeed, the projection of the distribution function F to the j-th coordinate is a one-dimensional distribution function, and as a consequence it has at least countably many atoms (points with positive measure with respect the measure induced by this distribution) for all indices $j=1,\dots,k$. This implies that we can choose a larger rectangle $\mathbf{K}$ if this is necessary whose boundary has μ_F measure zero.

Because of the boundedness of the function f the relation

$$\left|\int_{R^k\setminus\mathbf{K}} f(x_1,\dots,x_k)\,dF(x_1,\dots,x_k)\right| < \text{const.}\,\varepsilon$$

holds, and also $\limsup\limits_{n\to\infty}\left|\int_{R^k\setminus\mathbf{K}} f(x_1,\dots,x_k)\,dF_n(x_1,\dots,x_k)\right| < \text{const.}\,\varepsilon$ for all $n = 1,2,\dots$ because of the zero μ_F boundary of the rectangle $\mathbf{K}$ and the convergence of the distribution functions F_n to the distribution function F. Furthermore, the function f is uniformly continuous on the rectangle $\mathbf{K}$ hence there exists some constant $\delta > 0$ such that $|f(x)-f(y)| < \varepsilon$ if $|x-y| \le \delta$, and $x,y \in \mathbf{K}$. The rectangle $\mathbf{K}$ can be decomposed to finitely many rectangles Δ_j, $j=1,\dots,p(\mathbf{K})$, of diameter less than δ without joint interior points, and such that all these rectangles Δ_j have boundaries with μ_F measure zero. These properties imply that $\lim\limits_{n\to\infty}\mu_{F_n}(\Delta_j) = \mu_F(\Delta_j)$ for all indices $j=1,\dots,p(\mathbf{K})$, and because of the uniform continuity of the function f on the rectangle $\mathbf{K}$

$$\limsup\left|\int_{\mathbf{K}} f\,dF_n - \int_{\mathbf{K}} f\,dF\right| < \varepsilon.$$

The above inequalities imply that $\limsup \left|\int f\,dF_n - \int f\,dF\right| < \text{const.}\,\varepsilon$ with a const. independent of ε. Since this inequality holds for all $\varepsilon > 0$, it implies the statement we wanted to prove.

b.) The convergence of the integrals implies convergence in distribution:

Let $x = (x_1, \dots, x_k)$ be a point of continuity of the distribution function F. Then for all numbers $\varepsilon > 0$ there exists a number $\delta = \delta(\varepsilon) > 0$ such that the points $y = (y_1, \dots, y_k) = (x_1 - \delta, \dots, x_k - \delta)$ and $z = (z_1, \dots, z_k) = (x_1 + \delta, \dots, x_k + \delta)$ satisfy the inequalities $F(y) > F(x) - \varepsilon$ and $F(z) < F(x) + \varepsilon$. There exist continuous functions $f_1(u)$ and $f_2(u)$ an the k-dimensional Euclidean space R^k which satisfy the following properties: $0 \le f_i(u) \le 1$ for all $u \in R^k$, $i = 1, 2$. Further $f_1(u) = 1$ for $u = (u_1, \dots, u_k)$ if $u_j \le y_j$ for all indices $j = 1, \dots, k$, and $f_1(u) = 0$ if $u_j \ge x_j$ for some of the indices $1 \le j \le k$. The function $f_2(\cdot)$ satisfies the following relations: $f_2(u) = 1$ if $u_j \le x_j$ for all $j = 1, \dots, k$, and $f_2(u) = 0$ if $u_j \ge z_j$ for some $1 \le j \le k$. Then

$$
\begin{aligned}
\limsup_{n\to\infty} F_n(x) &\ge \lim_{n\to\infty} \int f_1(u)\,dF_n(u) = \int f_1(u)\,dF(u) \ge F(x) - \varepsilon \\
\liminf_{n\to\infty} F_n(x) &\le \lim_{n\to\infty} \int f_2(u)\,dF_n(u) = \int f_2(u)\,dF(u) \le F(x) + \varepsilon.
\end{aligned}
$$

Since these relations hold for all $\varepsilon > 0$, they imply the statement of part b.).

19.) Since $\bigcup\limits_{K=1}^{\infty} \mathbf{K}(K)^k = R^k$, and the rectangles $\mathbf{K}(K)^k$, $K = 1, 2, \dots$, constitute a monotone increasing series of sets, hence $\lim\limits_{K\to\infty} \mu(\mathbf{K}(K)^k) = \mu(R^k) = 1$, i.e. $\mu(\mathbf{K}(K)^k) \ge 1 - \varepsilon$ if $K \ge K(\varepsilon)$.

To show that the characteristic function of the probability measure μ is determined by its characteristic function let us first observe that the integrals $\int f(u)\,d\mu(u)$ determine the measure μ if we take all continuous functions $f(\cdot)$ with a bounded support. Indeed, the measure μ of those rectangles $\mathbf{P} = [K_1, L_1) \times \cdots \times [K_k, L_k)$ whose boundary has μ measure zero determine the measure μ. Besides, we claim that for all numbers $\varepsilon > 0$ and rectangles $\mathbf{P}$ there exists a function $f_{\varepsilon,\mathbf{P}}(\cdot)$ such that $0 \le f_{\varepsilon,\mathbf{P}}(u) \le 1$ for all points $u \in R^k$, $f_{\varepsilon,\mathbf{P}}(u) = 1$ if $u \in \mathbf{P}$, and $f_{\varepsilon,\mathbf{P}}(u) = 0$ if $\rho(u, \mathbf{P}) > \varepsilon$. (In the sequel $\rho(\cdot,\cdot)$ denotes the usual Euclidean distance in the space R^k.) Then the relation $\mu(\mathbf{P}) = \lim\limits_{\varepsilon\to 0} \int f_{\mathbf{P},\varepsilon}\,d\mu(u)$ implies the above property.

A possible construction of a function $f_{\mathbf{P},\varepsilon}$ with the above properties is the following: Put $f_{\mathbf{P},\varepsilon}(u) = 1 - g_{\mathbf{P},\varepsilon}(u)$ and $g_{\mathbf{P},\varepsilon}(u) = \min\left(1, \frac{1}{\varepsilon}\rho(u, \mathbf{P})\right)$.

Given a continuous function $f(\cdot)$ of compact support together with a sufficiently large number $K > 0$ for which the cube $[-K, K] \times \cdots \times [-K, K]$ contains the support of the function $f(\cdot)$ let us define the periodic extension of the function $f(\cdot)$ with period $2K$ by the formula $f_K(u_1 + 2Kj_1, \cdots, u_k + 2Kj_k) = f(u_1, \cdots, u_k)$, $-K \le u_j < K$, $l_j = 0, \pm 1, \pm 2, \dots$, $j = 1, \dots, k$. The integrals $\int f_K(u)\,d\mu(u)$ of these functions obtained by periodic extension determine the measure μ, since

$\int f(u)\,d\mu(u) = \lim\limits_{K\to\infty} \int f_K(u)\,d\mu(u)$ because of the tightness property of the measure μ formulated in the already proven part of problem 19.

Finally, by Weierstrass second approximation theorem for all continuous functions $f_K(\cdots)$ K with period K and real number $\varepsilon > 0$ there exists a trigonometrical polynomial

$$g_\varepsilon = g_{\varepsilon, f_K}(u_1, \cdots, u_k) = \sum c^\varepsilon_{j_1,\dots,j_k} e^{i\pi(j_1u_1+\cdots+j_ku_k)/K}$$

such that $\sup\limits_{u\in R^k} |f_K(u) - g_\varepsilon(u)| \le \varepsilon$. Hence

$$\left| \int f_K(u)\,d\mu(u) - \int g_\varepsilon(u)\,d\mu(u) \right| \le \varepsilon.$$

On the other hand, $\int g_\varepsilon(u)\,d\mu(u) = \sum c^\varepsilon_{j_1,\dots,j_k} \varphi\left(\frac{\pi j_1}{K}, \dots, \frac{\pi j_k}{K}\right)$, that is the above integral can be calculated by means of the characteristic function of the measure μ. This implies that this characteristic function determines the integrals of the form $\int f_K(u)\,d\mu(u)$. Hence it also determines the measure μ.

The proof can be generalized without any essential modification to arbitrary signed measures μ with bounded variation.

20.) First we show that for all numbers $a > 0$ there exists an even density function $f(u)$ whose Fourier transform $\varphi(t)$ is sufficiently smooth, e.g. it is twice differentiable, and it equals zero outside the interval $[-a, a]$.

Indeed, let us consider a continuously differentiable function $g(u)$ which is concentrated in the interval $\left[-\frac{a}{2}, \frac{a}{2}\right]$, $g^-(u) = g(-u)$. Then put $h(u) = g * g^-(u)$, $f(u) = \frac{2\pi}{M} \int e^{itu} h(u)\,du$, where $*$ denotes convolution, and $M = h(0) = \int |f(u)|^2\,du$. We claim that this function f is a density function, and its characteristic function is the function $\frac{h^-(u)}{M}$, $h^-(u) = h(-u)$, which vanishes outside the interval $[-a, a]$. Indeed, the function $h(\cdot)$ is twice differentiable, (see problem 17), hence its Fourier transform tends to zero in plus–minus infinity with order $|t|^{-2}$ (see e.g. problem 28 of this series of problems discussed later), hence the above defined Fourier transform $f(\cdot)$ of the function $\frac{h(u)}{M}$ is integrable, and we can apply the inverse Fourier transform for it. Since $f(\cdot)$ is en even function, this means that the function $\frac{h^-(u)}{M} = \int e^{itx} f(u)\,du$ is the Fourier transform of the function $f(u)$. In particular, $\frac{h(0)}{M} = 1 = \int f(u)\,du$. Finally, $f(t) \ge 0$ for all numbers $t \in R^1$, since the Fourier transform of the function $g * g^-(\cdot)$ equals $\int e^{itu} g * g^-(u)\,du = \int e^{itu} g(u)\,du \int e^{itu} g^-(u)\,du = \left|\int e^{itu} g(u)\,du\right|^2 \ge 0$. These properties mean that the function $f(\cdot)$ is a density function. (We shall return to the above problem in the second part of this series of problems where such a construction will be useful in a different context.)

Let us consider an even density function $f(u)$ whose characteristic function $\varphi(t)$ is twice differentiable, and vanishes outside of a finite interval $[-a, a]$. Consider a number $T > a$, and let us define the numbers $a_k = \frac{1}{4\pi T} \int e^{-i\pi tk/T} \varphi(t)\,dt$, $k =$

$0, \pm 1, \pm 2, \dots$. Let us put weights a_k in the points $\frac{\pi k}{T}$, $k = 0, \pm 1, \pm 2, \dots$. We claim that in such a way we constructed a probability distribution on the lattice $\frac{\pi k}{T}$, $k = 0, \pm 1, \pm 2, \cdots$, whose characteristic function to the interval $[-T, T]$ agrees with the restriction of the characteristic function $\varphi(t)$ to this interval. The characteristic function of this discrete distribution is the periodic extension of the restriction of $\varphi(t)$ to the interval $[-T, T]$ with period $2T$. This statement means in particular, that the above defined discrete distribution together with the distribution with density function $f(\cdot)$ yield an example satisfying the statement of problem 20.

The statement formulated for the discrete distribution with weights a_k holds, because for one hand a comparison of the definition of the number a_k with the inverse Fourier transformation formula expressing the function $f(\cdot)$ yields that $a_k = \frac{1}{2T} f\left(\frac{\pi k}{T}\right) \ge 0$. On the other hand, the trigonometrical sum $\sum\limits_{k=-\infty}^{\infty} a_k e^{\pi i k/T}$ is the Fourier series of the function $\varphi(t)$ restricted to the interval $[-T, T]$. In particular, $\varphi(0) = 1 = \sum\limits_{k=-\infty}^{\infty} a_k$. (As $\varphi(\cdot)$ is a twice differentiable function, hence it equals his Fourier series in all points.)

21.) Let us first show that if the sequence of the probability measures μ_n, $n = 1, 2, \dots$, is relatively compact, then it is also tight.

Let us assume indirectly this sequence of measures μ_n, $n = 1, 2, \dots$, is not tight. Then there exists a constant $\varepsilon > 0$, a subsequence μ_{n_k} of the sequence of probability measures μ_n and a sequence of positive numbers K_n, $n = 1, 2, \dots$, such that $K_n \to \infty$, and $\mu_{n_k}([-K_n, K_n] \times \cdots \times [-K_n, K_n]) < 1 - \varepsilon$. We shall show that this subsequence μ_{n_k} of the sequence of measures μ_n has no sub-subsequence convergent in distribution. This means that the above formulated indirect assumption leads to contradiction.

Indeed, let us assume indirectly that the sequence of measures μ_{n_k} has a subsequence $\mu_{n_{k_j}}$ which converges in distribution to a probability measure μ. Then, there exists a constant $K > 0$ such that $\mu([-K, K] \times \cdots \times [-K, K]) > 1 - \frac{\varepsilon}{2}$, and the hyperplanes $u_j = \pm K$, $j = 1, 2, \dots, k$, have μ measure zero. Then also the relation $\lim\limits_{j \to \infty} \mu_{n_{k_j}}([-K, K] \times \cdots \times [-K, K]) = \mu([-K, K] \times \cdots \times [-K, K])$ should hold. But this is not possible, since the lim sup of the probabilities at the left-hand side is smaller than $1 - \varepsilon$, while the right-hand side is greater than $1 - \frac{\varepsilon}{2}$.

Let us show that if the sequence of measures μ_n is tight then it is also relatively compact.

We have to show that an arbitrary subsequence of the sequence μ_n has a sub-subsequence convergent in distribution. For the sake of simpler notations let us denote by reindexing the elements of the subsequence again by μ_n. We have to show that this (also tight) sequence of probability measures μ_n has a subsequence convergent in distribution.

Let $F_n(u) = F_n(u_1, \dots, u_k)$ denote the distribution function $F_n(u_1, \dots, u_k) =$

$\mu_n(\{(v_1,\dots,v_k)\colon\ v_j<u_j,\ j=1,\dots,k\})$ determined by the measure μ_n. Let

$$u^{(p)}=\left(u_1^{(p)},\dots,u_k^{(p)}\right),\quad p=1,2,\dots,$$

denote the (countable) set of points $u^{(p)}\in R^k$ with rational coordinates with some indexing. First we show with the help of the so-called diagonal procedure that there exists an appropriate sequence of positive integers n_j, $j=1,2,\dots$, such that the limit

$$\lim_{j\to\infty}F_{n_j}\left(u_1^{(p)},\dots,u_k^{(p)}\right)=\tilde F\left(u_1^{(p)},\dots,u_k^{(p)}\right) \tag{2.4}$$

exists for all numbers $p=1,2,\dots$.

Indeed, as $0\le F_n(u)\le 1$, there is a subsequence $\bar n_j=(n_{j,1})$ of the integers such that the limit $\lim\limits_{j\to\infty}F_{n_{j,1}}(u^{(1)})=\tilde F(u^{(1)})$ exists. This sequence has a subsequence $n_{j,2}$ such that the limit $\lim\limits_{j\to\infty}F_{n_{j,2}}(u^{(2)})=\tilde F(u^{(2)})$ also exists. Following this procedure we can construct sequences $n_{j,p}$, $j=1,2,\dots$, $p=1,2,\dots$ in such a way that the $p+1$-th sequence is a subsequence of the p-th sequence, i.e. $\{n_{j,p+1},\ j=1,2,\dots\}\subset\{n_{j,p},\ j=1,2,\dots\}$, $p=1,2,\dots$, and the limit $\lim\limits_{j\to\infty}F_{n_{j,p}}(u^{(p)})=\tilde F(u^{(p)})$ exists for all numbers $p=1,2,\dots$. Then the sequence $n_j=n_{j,j}$ satisfies relation (2.4).

Let us introduce the function

$$F(u_1,\dots,u_k)=\sup_{\left\{u^{(p)}=(u_1^{(p)},\dots,u_k^{(p)})\colon\ u_s^{(p)}<u_s,\ s=1,\dots,k\right\}}\tilde F\left(u_1^{(p)},\dots,u_k^{(p)}\right), \tag{2.5}$$

where the function $\tilde F$ satisfies relation (2.4) with an appropriate (fixed) sequence of positive integers n_j, and the points $u^{(p)}$ are the points of the k-dimensional space with rational coordinates. We claim that the above defined function $F(u_1,\dots,u_k)$ is a distribution function, and the distribution functions $F_{n_j}(u_1,\dots,u_k)$ converge in distribution to it. If we prove this statement then we complete the solution of problem 21.

To show that the function $F(u_1,\dots,u_k)$ is a distribution function we recall an “internal” characterization result of distribution functions which describes the distribution functions only with the help of their property. The result we recall states that a function $F(u_1,\dots,u_k)$ is a distribution function if and only if it satisfies the following four properties:

(i) The function $F(u_1,\dots,u_k)$ is a function continuous from the left in all of its arguments.

(ii) $\displaystyle\lim_{\substack{u_j\to\infty\\ \text{for all indices } j=1,\dots,k}}F(u_1,\dots,u_k)=1.$

(iii) $\displaystyle\lim_{\substack{u_j\to-\infty\\ \text{for some index } 1\le j\le k}}F(u_1,\dots,u_k)=0.$

Finally, to formulate the last condition let us define for a function F on the space R^k and a rectangle $\mathbf{K} = [a_1, b_1) \times \cdots \times [a_k, b_k)$ the number

$$\mu(\mathbf{K}) = \mu_F(\mathbf{K}) = \sum_{\substack{u_j = a_j \text{ vagy } b_j \\ j=1,\dots,k}} (-1)^{\chi(u_1,\dots,u_k)} F(u_1, \dots, u_k),$$

where $\chi(u_1, \dots, u_k)$ denotes the quantity of the numbers a_j in the sequence $u_1, \dots, u_k$. Then

(iv) $\mu_F(\mathbf{K}) \ge 0$ for all rectangles $\mathbf{K}$.

Since the function $\tilde F(u_1, \dots, u_k)$ defined in points of rational coordinates is a monotone increasing function in all of its coordinates, the function $F(\cdot)$ defined in formula (2.5) also satisfies property (i). Furthermore, this monotonicity also implies that we can replace the sup by lim in formula (2.5) if we consider such sequences of indices $\left(u_1^{(p)}, \dots, u_k^{(p)}\right)$, $p = 1, 2, \dots$, in the limit for which $u_j^{(p)} < u_j$ for all numbers $1 \le j \le k$ and $p = 1, 2, \dots$, and $\lim\limits_{p\to\infty} u_j^{(p)} = u_j$ for all indices $j = 1, \dots, k$. Let us consider such rectangles $\mathbf{K}(p)$ whose edges have rational coordinates, and the edges of these coordinates converge in a monotone increasing way to the coordinates of the corresponding edges of the rectangle $\mathbf{K}$. Then $\mu_{\tilde F}(\mathbf{K}(p)) \ge 0$, since $\tilde F$ is the limit of distribution functions. Hence $\mu_F(\mathbf{K}) = \lim\limits_{p\to\infty} \mu_{\tilde F}(\mathbf{K}(p)) \ge 0$, i.e. the function F satisfies property (iv). Let us observe that properties (ii) and (iii) hold if the functions F_n are replaced by the function $\tilde F$, and the limit is taken only in rational points. (We applied at this point of the proof the tightness of the measures μ_n.) This property together with formula (2.5) imply that the function F satisfies properties (ii) and (iii). Property (i) also holds. This is a simple consequence of it definition in formula (2.5).

To show that the distribution functions F_{n_j} converge in distribution to the distribution function F let us consider a point of continuity $u = (u_1, \dots, u_k)$ of the function F, and let us then choose for all numbers $\varepsilon > 0$ such a constant number $\delta = \delta(\varepsilon) > 0$ for which $F(u) - \varepsilon \le F(u - \delta) \le F(u) \le F(u + \delta) \le F(u) + \varepsilon$, where $u \pm \delta = (u_1 \pm \delta, \dots, u_k \pm \delta)$. Let us then choose two points $r = (r_1, \dots, r_k) \in R^k$ and $\bar r = (\bar r_1, \dots, \bar r_k) \in R^k$ with rational coordinates such that $u_j - \delta < r_j < u_j < \bar r_j < u_j + \delta$ for all indices $j = 1, \dots, k$. Then by the monotonicity properties of the function $\tilde F(\cdot)$ and the definition of the function F

$$F(u) - \varepsilon \le F(u - \delta) < \tilde F(r) \le F(u) \le \tilde F(\bar r) \le F(u + \delta) \le F(u) + \varepsilon.$$

This relation together with the definition of the function $\tilde F$ and the monotonicity property of the functions F_{n_j} imply that

$$\begin{aligned} F(u) - \varepsilon &\le \lim_{j\to\infty} F_{n_j}(r) \le \liminf_{j\to\infty} F_{n_j}(u) \\ &\le \limsup_{j\to\infty} F_{n_j}(u) \le \lim_{j\to\infty} F_{n_j}(\bar r) \le F(u) + \varepsilon, \end{aligned}$$

hence

$$-\varepsilon \le \liminf_{j\to\infty} F_{n_j}(u) - F(u) \le \limsup_{j\to\infty} F_{n_j(u)} - F(u) \le \varepsilon.$$

Since this relation holds for all $\varepsilon > 0$, it implies that $\lim\limits_{j\to\infty} F_{n_j}(u) = F(u)$.

22.) Fix some $\delta > 0$ and write the following identity:

$$\begin{aligned}
\frac{1}{2\delta}\int_{-\delta}^{\delta} \operatorname{Re}\,[1-\varphi_n(t)]\,dt &= \int_{-\delta}^{\delta}\frac{1}{2\delta}\int_{-\infty}^{\infty}[1-\cos tx]\;dF_n(x)\,dt \\
&= \int_{-\infty}^{\infty}\frac{1}{2\delta}\int_{-\delta}^{\delta}[1-\cos tx]\,dt\,dF_n(x) = \int_{-\infty}^{\infty}\left[\frac{t}{2\delta}-\frac{\sin tx}{2\delta x}\right]_{t=-\delta}^{t=\delta} dF_n(x) \\
&= \int_{-\infty}^{\infty}\left(1-\frac{\sin\delta x}{\delta x}\right)\,dF_n(x) = \int_{-K}^{K}\left(1-\frac{\sin\delta x}{\delta x}\right)\,dF_n(x) \\
&\qquad + \int_{|x|>K}\left(1-\frac{\sin\delta x}{\delta x}\right)\,dF_n(x) = I_{1,n}^{\delta}(K) + I_{2,n}^{\delta}(K).
\end{aligned} \tag{2.6}$$

First we show with the help of relation (2.6) that the validity of formula (10) implies that the sequence of distribution functions F_n is tight. Since $\left(1-\frac{\sin\delta x}{\delta x}\right) \ge 0$ for all x and δ, hence the left-hand side of formula (2.6) yields an upper bound on the expression $I_{2,n}^{\delta}(K)$ for all numbers $\delta > 0$, $n \ge 1$ and $K > 0$. If formula (10) holds, then for all $\varepsilon > 0$ there exists a number $\delta = \delta(\varepsilon) > 0$ and threshold index $n_0 = n_0(\delta,\varepsilon)$ such that $\frac{\varepsilon}{2} \ge \int_{|x|>K}\left(1-\frac{\sin\delta x}{\delta x}\right)\,dF_n(x)$ for $n \ge n_0$. Put $K = \frac{2}{\delta}$. Then $1-\frac{\sin\delta x}{\delta x} \ge \frac{1}{2}$ for all $|x| \ge K$. Hence the previous estimate implies that $\frac{\varepsilon}{2} \ge \int_{|x|>K}\left(1-\frac{\sin\delta x}{\delta x}\right)\,dF_n(x) \ge \frac{1}{2}[(1-F_n(K))+F_n(-K)]$, i.e. $\varepsilon \ge [(1-F_n(K))+F_n(-K)]$ with this number K if $n \ge n_0$. By increasing the number $K > 0$ if it is necessary we can achieve that the above inequality holds for all indices $n \ge 1$. This means that the distribution functions F_n, $n = 1, 2, \dots$, are tight.

Let us prove with the help of formula (2.6) that the tightness of the distribution functions F_n implies formula (10) and even its slightly stronger version, formula (10′) where $\limsup\limits_{n\to\infty}$ is replaced by $\sup\limits_{n\ge 1}$. Since $\left|1-\frac{\sin\delta x}{\delta x}\right| \le 2$, the tightness of the distribution functions F_n makes possible to choose a number $K = K(\varepsilon) > 0$ for all $\varepsilon > 0$ such that $|I_{2,n}^{\delta}(K)| = \left|\int_{|x|>K}\left(1-\frac{\sin\delta x}{\delta x}\right)\,dF_n(x)\right| \le \frac{\varepsilon}{2}$ for all numbers $\delta > 0$ and $n = 1, 2, \dots$. After fixing the number $K = K(\varepsilon) > 0$ we can choose a number $\bar\delta = \bar\delta(\varepsilon, K) > 0$ such that the inequality $\varepsilon \ge 1-\frac{\sin\delta x}{\delta x} \ge 0$ holds for all numbers $|x| < K$ and $0 < \delta < \bar\delta$. Hence $|I_{1,n}^{\delta}(K)| = \left|\int_{-K}^{K}\left(1-\frac{\sin\delta x}{\delta x}\right)\,dF_n(x)\right| \le \frac{\varepsilon}{2}$. These estimates together with relations (2.6) imply that $\left|\frac{1}{2\delta}\int_{-\delta}^{\delta}\operatorname{Re}\,[1-\varphi_n(t)]\,dt\right| \le \varepsilon$ for all $n \ge 1$ if $\delta \le \bar\delta(\varepsilon)$. The statements of problem 22 are proved.

23.) Let us consider the j-th coordinate, $1 \le j \le k$, of the random vectors $\xi^{(n)}$, i.e. the random variables $\xi_j^{(n)}$ for all indices $n = 1, 2, \dots$. The characteristic function of the random variable $\xi_j^{(n)}$ is the function

$$\varphi_n^{(j)}(t) = \varphi_n(\underbrace{0,\cdots,0,}_{\substack{j-1 \\ \text{0 coordinates}}}\ t,\ \underbrace{0,\cdots,0}_{\substack{n-j-1 \\ \text{0 coordinates}}}).$$

By the conditions of the problem the functions $\varphi_n^{(j)}(t)$ converge to a function $\varphi^{(j)}(t)$ continuous in zero in a small neighbourhood of the origin. Let us remark that $\varphi^{(j)}(0) = \lim\limits_{n\to\infty} \varphi_n^{(j)}(0) = 1$. Hence the continuity of the function $\varphi^{(j)}(t)$ in the origin implies that for all numbers $\varepsilon > 0$ there exists a threshold index $\bar\delta = \bar\delta(\varepsilon) > 0$ such that for all numbers $0 < \delta < \bar\delta$

$$0 \le \frac{1}{2\delta}\int_{-\delta}^{\delta} \operatorname{Re}\,[1-\varphi_n^{(j)}(t)]\,dt < \varepsilon.$$

Furthermore, as $\lim\limits_{n\to\infty} \operatorname{Re}\,[1-\varphi_n^{(j)}(t)] = \operatorname{Re}\,[1-\varphi^{(j)}(t)]$ if $|t| < \delta < \bar\delta$ (we choose a smaller threshold $\bar\delta > 0$ if it is necessary), and $0 \le \operatorname{Re}\,[1-\varphi_n^{(j)}(t)] \le 2$, it follows from Lebesgue's dominated convergence theorem that $0 \le \limsup\limits_{n\to\infty} \frac{1}{2\delta}\int_{-\delta}^{\delta} \operatorname{Re}\,[1-\varphi^{(j)}(t)]\,dt < \varepsilon$. Hence the result of problem 22 shows that the distribution functions of the random variables $\xi_n^{(j)}$, $n = 1, 2, \dots$, are tight, i.e. for all $\varepsilon > 0$ there exists a constant $K = K(\varepsilon) > 0$ such that the inequality $P\left(\left|\xi_n^{(j)}\right| > K\right) < \frac{\varepsilon}{k}$ holds. Since this statement holds for all numbers $j = 1, \dots, k$, it implies that the distributions of the random vectors $\bar\xi_n = (\xi_1^{(n)}, \dots, \dots, \xi_k^{(n)})$ are tight.

24.) It follows from the results of problems 21 and 23 that the sequence of distribution functions $F_n(u_1, \dots, u_k)$, $n = 1, 2, \dots$, is relatively compact, i.e. an arbitrary subsequence of the sequence of the distribution functions F_n has a sub-subsequence convergent in distribution if the characteristic functions of the distribution functions F_n converge to a function continuous in the origin. (Moreover, it is enough to assume that this property holds for the restriction of the characteristic functions to the coordinate axes.) To see that under the conditions of the first part of problem 24 the distribution functions F_n converge in distribution it is enough to show that in this case all convergent subsequences of this sequence of distribution functions have the same limit. To justify this reduction of the problem let us choose a convergent subsequence F_{n_l} which converges to some distribution function $F(u_1, \dots, u_k)$. If this distribution function $F(\cdot)$ were not the limit of the distribution functions F_n, then the distribution function $F(u_1, \dots, u_k)$ would have a point of continuity $u = (u_1, \dots, u_k)$ together with a constant $\varepsilon > 0$ and a sequence of indices n_j, $j = 1, 2, \dots$, such that $|F_{n_j}(u_1, \dots, u_k) - F(u_1, \dots, u_k)| > \varepsilon$ for all $j = 1, 2, \dots$. But this would mean that a convergent subsequence of the sequence of distribution functions F_{n_j}, $j = 1, 2, \dots$, would have a limit different of the distribution function $F(u_1, \dots, u_k)$.

The statement that all convergent subsequences of the distribution functions F_n have the same limit follows from the results of Theorem A and problem 19. Indeed, it follows from Theorem A that the characteristic function of the limit distribution function of a convergent subsequence of the distribution functions F_n is the limit of the characteristic functions of the distribution functions in this subsequence. The limit of these characteristic functions does not depend on which convergent subsequence we have considered. But by the result of problem 19 a distribution function is determined by its distribution function. Hence the condition that the characteristic functions of a sequence of distribution functions converge to a function continuous in the origin implies that these distribution functions converge in distribution to a distribution function, and besides the characteristic function of the limit distribution function equals the limit of the characteristic functions of the distribution functions we have considered.

If a sequence of distribution functions $F_n(u_1,\dots,u_k)$ converges in distribution to a distribution function $F_0(u_1,\dots,u_k)$, then it follows from Theorem A that the characteristic functions $\varphi_n(t_1,\dots,t_k)$ of these distribution functions converge to the characteristic function $\varphi_0(t_1,\dots,t_k)$ of the distribution function F_0 in all points $(t_1,\dots,t_k)\in R^k$. To complete the proof of the Fundamental Theorem we have still to show that this convergence is uniform in all compact subset of the Euclidean space R^k.

To prove this statement let us observe that since the distribution functions F_n converge in distribution they are tight. Hence for all $\varepsilon>0$ there exists a constant $K=K(\varepsilon)$ such that a sequence of random vectors $\xi_n=(\xi_n^{(1)},\dots,\xi_n^{(k)})$, $n=1,2,\dots$, with distribution functions F_n satisfy the inequality $P(|\xi_n|>K)<\frac{\varepsilon}{3}$ for all indices $n=0,1,2,\dots$. (In the further part of the proof $\xi(\omega)$, $t\in R^k$, $u\in R^k$ denote points of the k-dimensional space, and (u,t), $u\in R^k$, $t\in R^k$, denotes the scalar product of the vectors u and t.) Let us choose a finite set of points $\mathbf{T}=\left\{t^{(1)},\dots,t^{(s)}\right\}\subset\mathbf{K}$, $s=s(\mathbf{K},\delta)$, in a compact set $\mathbf{K}\subset R^k$ which is δ-dense in the set $\mathbf{K}$, i.e. for all $t\in\mathbf{K}$ there is a point $t^{(j)}\in\mathbf{T}$ such that $\rho(t,t^{(j)})<\delta$. Then

$$\begin{aligned}\left|\varphi_n(t)-\varphi_n(t^{(j)})\right|&=\left|Ee^{i(t,\xi_n)}-e^{i(t^{(j)},\xi_n)}\right|\\&\le E\left|e^{i(t-t^{(j)},\xi_n)}-1\right|I\left(|\xi_n|\le K\right)+P(|\xi_n|>K)\le\frac{2\varepsilon}{3}\end{aligned}$$

for all numbers $n=1,2,\dots$. Further we can choose a threshold index $n_0=n_0(\varepsilon)$, such the inequality $\sup\limits_{n\ge n_0}\sup\limits_{t^{(j)}\in\mathbf{T}}\left|\varphi_n(t^{(j)})-\varphi_0(t^{(j)})\right|<\frac{\varepsilon}{3}$ holds. It follows from the last inequalities that $\sup\limits_{t\in\mathbf{K}}|\varphi_n(t)-\varphi_0(t)|<\varepsilon$ if $n\ge n_0$. This means that the convergence $\varphi_n(t)\to\varphi_0(t)$ is uniform on all compact sets $\mathbf{K}$.

25.) Let $\varphi_0(t)$ denote the characteristic function of the uniform distribution in the interval $[-1,1]$, i.e. $\varphi_0(t)=\int_{-1}^1\frac12e^{itu}\,du=\frac{e^{it}-e^{-it}}{2it}$. Let us define the characteristic functions $\varphi_n(t)$ as the characteristic functions of the following discretizations μ_n, $n=1,2,\dots$, of the uniform distribution on the interval $[-1,1]$: $\mu_n\left(\frac kn\right)=\frac{1}{2n+1}$, $-n\le k\le n$. Then

$$\varphi_n(t)=\frac{1}{2n+1}\sum_{k=-n}^n e^{ikt/n}=\frac{e^{i(n+1)t/n}-e^{i(-n+1)t/n}}{(2n+1)(e^{it/n}-1)}.$$

Simple calculation shows that $\varphi_n(t) \to \varphi_0(t)$ for all points $t \in R^1$, and the convergence is uniform in all finite intervals. On the other hand, the convergence is not uniform on the whole real line, since $\lim\limits_{t\to\infty} \varphi_0(t) = 0$, while $\varphi_n(t) = 1$ in the points of the form $t = 2\pi kn$, $k = 0, \pm 1, \pm 2, \dots$. (The background of this construction: We have approximated the characteristic function of a distribution function having a density function with the characteristic function of more and more closer discretization of this distribution function. These discretized approximations of the original distribution functions had lattice distributions. The characteristic functions of such approximating distributions converge to the characteristic function of the limit distribution by the Fundamental Theorem. Besides, the characteristic function of a distribution function with a density function tends to zero in the infinity by the Riemann lemma. On the other hand the characteristic function of a lattice valued distribution is a periodic function which has absolute value 1 in certain points.)

26.) An example for case a): Let the measures μ_n have uniform distribution in the interval $[-n, n]$. Then $\varphi_n(t) = \frac{1}{2n}\int_{-n}^{n} e^{itu}\,du = \frac{e^{itn}-e^{-itn}}{2int}$. Hence $\lim\limits_{n\to\infty} \varphi_n(t) = 0$ if $t \neq 0$, and $\lim\limits_{n\to\infty} \varphi_n(0) = 1$.

An example for case b): Let $\mu_{2n}(\{n\}) = \mu_{2n}(\{-n\}) = \frac{1}{2}$, and μ_{2n+1} be the probability measure μ_n defined in case a). Then $\varphi_{2n}(t) = \frac{1}{2}(e^{itn} + e^{-itn})$, and it equals 1 in the points of the form $\frac{2k\pi}{n}$. This means that in the points of the form $t = \varphi\left(\frac{2k\pi}{l}\right)$ $\varphi_{n_k}(t) = 0$ for a certain subsequence n_k, and $\lim\limits_{k\to\infty} \varphi_{\bar n_k}(t) = 1$ for a certain subsequence $\bar n_k$.

27.) Let $F(x)$ denote the distribution function of the random variable ξ. Then $\varphi(t) = \int e^{itu} dF(u)$. By successive differentiation we get that $\frac{d^k\varphi(t)}{dt^k} = i^k \int u^k e^{itu}\,dF(u)$, in particular $\left.\frac{d^k\varphi(t)}{dt^k}\right|_{t=0} = i^k \int u^k\,dF(u) = i^k E\xi^k$ if the order of derivation and integration can be changed in the above calculation. The above calculation is legitim if the distribution F is concentrated in the interval $[-K, K]$, because the integrand in the integral expressing the fraction $\frac{\varphi(t+h)-\varphi(t)}{h}$ satisfies the relation $\frac{e^{i(t+h)u}-e^{itu}}{h} = iue^{itu} + O(h)$, and for a fixed number t the order $O(h)$ is uniform if $u \in [-K, K]$.

If $E|\xi| = \int |u|\,dF(u) < \infty$, then to prove the statement of problem 27 for the first derivative we introduce the functions

$$G_n(t) = \int_{-n}^{n} e^{iut}\,dF(u), \quad H_n(t) = i\int_{-n}^{n} ue^{iut}\,dF(u), \quad n = 1, 2, \dots,$$

and $G(t) = \int_{-\infty}^{\infty} e^{iut}\,dF(u)$ és $H(t) = i\int_{-\infty}^{\infty} ue^{iut}\,dF(u)$. Then $\lim\limits_{n\to\infty} G_n(t) = G(t)$, and $\lim\limits_{n\to\infty} \frac{dG_n(t)}{dt} = H(t)$ with $H(t) = i\int_{-\infty}^{\infty} ue^{iut}\,dF(u)$. We show with the help of the above statements that the function $G(t)$ is differentiable, and $\frac{dG(t)}{dt} = H(t)$. Indeed, $G(t) = \lim\limits_{n\to\infty} G_n(t) = \lim\limits_{n\to\infty} \left[G_n(0) + \int_0^t H_n(s)\,ds\right]$, hence the relations $\lim\limits_{n\to\infty} G_n(0) = G(0)$, $\lim\limits_{n\to\infty} H_n(s) = H(s)$, and the validity of the inequality

$|H_n(s)| \le E|\xi|$ for all numbers n and s together with Lebesgue dominated convergence theorem imply that $G(t) = G(0) + \int_0^t H(s)\,ds$. Hence $\frac{dG(t)}{dt} = H(t)$, and this is what we had to prove.

The statement about the k-th derivative of the Fourier transform under the condition $E|\xi|^k < \infty$ can be proved similarly by induction with respect to the parameter k with the help of the identity $\frac{dG^k(t)}{dt^k} = \frac{d}{dt}(\frac{dG^{k-1}(t)}{dt^{k-1}})$. Only in this case we have to work with the functions $G_n(t) = i^{k-1}\int_{-\infty}^{\infty} u^{k-1}e^{iut}\,dF(u)$, $H_n(t) = i^k\int_{-n}^{n} u^k e^{iut}\,dF(u)$, $n = 1, 2, \dots$, and $G(t) = i^{k-1}\int_{-\infty}^{\infty} u^{k-1}e^{iut}\,dF(u)$, $H(t) = i^k\int_{-\infty}^{\infty} u^k e^{iut}\,dF(u)$. We prove the identity $G(t) = G(0) + \int_0^t H(s)\,ds$ also in this case.

If $Ee^{t\xi} < \infty$ with some number $t > 0$, then $P(\xi > x) = P(e^{t\xi} > e^{tx}) \le e^{-tx}Ee^{t\xi} \le \text{const.}\, e^{-tx}$ for all numbers $x \ge 0$. Similarly, $P(\xi < -x) \le \text{const.}\, e^{-tx}$, if $Ee^{-t\xi} < \infty$. Hence $P(|\xi| > x) \le \text{const.}\, e^{-tx}$ if $Ee^{ux} < \infty$ for $|u| \le t$. Conversely, if $G(u) = P(|\xi| > x) \le \text{const.}\, e^{-\alpha x}$, then we get by partial differentiation that $Ee^{t|\xi|} = \int_0^\infty e^{tu}\,dG(u) = [e^{tu}G(u)]_0^\infty - \int_0^\infty te^{tu}G(u)\,du < \infty$ in the case $0 < t < \alpha$. Hence $Ee^{\pm t\xi} \le 1 + Ee^{t|\xi|} < \infty$.

Finally, if $P(|\xi| > x) \le \text{const.}\, e^{-\alpha x}$, then the function $G(z) = \int e^{izx}\,dF(x)$ is analytic in the domain $\{z\colon |\mathrm{Im}\, z| < \alpha\}$, because in an arbitrary compact set in the interior of this domain the function $G(z)$ can be represented as the uniform limit of analytic functions (finite sums approximating this integral). This function $G(z)$ is the analytic continuation of the function $\varphi(t)$ to the above domain.

28.) Let us first prove Riemann's lemma. If $g(u) = I([a, b])$ is the indicator function of an interval $[a, b]$, then $\int e^{itu}g(u)\,du = \frac{e^{ibt}-e^{iat}}{it} \to 0$ if $t \to \infty$ or $t \to -\infty$. This relation also holds if $g(u) = \sum\limits_{j=1}^{k} c_j I([a_j, b_j])$, i.e. $g(\cdot)$ is a finite linear combination of indicator functions of intervals. Such functions constitute an everywhere dense subset of the integrable functions in L_1 norm, that is for all numbers $\varepsilon > 0$ and integrable function $f(\cdot)$ there exists a function $g(\cdot)$ or the above form such that $\int |f(u) - g(u)|\,du < \varepsilon$. This relation implies that $\left|\int e^{itu}f(u)\,du - \int e^{itu}g(u)\,du\right| < \varepsilon$ for all numbers $t \in R^1$. The Riemann lemma is a consequence of the above relations.

If the function $f(t)$ is k times differentiable, and the first k derivatives are integrable functions on the real line, then we get by successive partial differentiation that $\varphi(t) = i^k t^{-k}\int e^{itu} \left.\frac{df^k(s)}{ds^k}\right|_{s=u} du$. This relation together with Riemann's lemma imply that in this case $\varphi(t) = o(t^{-k})$ if $t \to \pm\infty$.

If the function $f(\cdot)$ is analytic in a strip $\{z\colon \mathrm{Re}\, z \in [-A, A]\}$ around the real line and the function $f(-ia + \cdot)$ is integrable for $a < A$, then we can write, because of the inequality $|e^{i(u-ia)t}| < e^{-at}$ that

$$|\varphi(t)| = \left|\int_{-ia-\infty}^{-ia+\infty} e^{itu}f(u)\,du\right| \le e^{-at}\int_{-\infty}^{\infty} |f(u - ia)|\,du \le \text{const.}\, e^{-at}$$

for $t > 0$. The case $t < -0$ can be handled similarly, only the integral has to be replaced to the line $[-\infty+ia,\infty+ia]$ on the positive half-space.

29.) The relation $|\varphi(t)|=1$ holds if and only if $\varphi(t)=e^{ita}$, that is $Ee^{it(\xi-a)}=1$ with some real number a. This identity holds if and only if $P(t(\xi-a)\in 2\pi\{0,\pm1,\pm2,\dots,\})=1$. This means that $|\varphi(t)|=1$ with some $t\neq0$ if the values of the random variable ξ are concentrated to a lattice $\left\{\frac{2\pi k}{t}+a,\ k=0,\pm1,\pm2,\dots\right\}$ of width $\frac{2\pi}{t}$. In particular, the relation $|\varphi(t)|=1$ for all $t\in R^1$ can hold if and only if the random variable takes a constant value with probability 1. If ξ is not a deterministic constant, and it is lattice distributed, then there is a largest $h>0$ such that the distribution of ξ has period h. Indeed, in this case there are two numbers a and b, $a\neq b$, such that $P(\xi=a)>0$, and $P(\xi=b)>0$. Then the distribution of ξ may have a period only with $\frac{b-a}{k}$, where k is a positive integer. Since the distribution of ξ has a period $h>0$, the above statement holds.

To finish the solution of the problem it is enough to observe that the characteristic function is continuous on the real line, and the characteristic function $\varphi(\cdot)$ of a non-lattice valued random variable satisfies the inequality $|\varphi(t)|<1$ if $t\neq0$. Hence $\sup\limits_{A\le|t|\le B}|\varphi(t)|<1$ in this case.

30.) The characteristic function of the random variable ξ considered in this problem equals $\varphi(t)=\frac12\left(\cos t+\cos(\sqrt2 t)\right)$. There exist pairs of integers (p_n,q_n), $n=1,2,\dots$, such that $q_n\to\infty$, and $|\sqrt2 q_n-p_n|\le\frac1{q_n}$. (Such pairs of integers (p_n,q_n) can be found as the nominator and denominator of the continued fraction of the number $\sqrt2$.) Choose $t_n=2\pi q_n$. Then $\cos t_n=1$, and $\lim\limits_{n\to\infty}\cos(\sqrt2 t_n)=1$. Hence $t_n\to\infty$, and $|\varphi(t_n)|\to1$, if $n\to\infty$. On the other hand, $\varphi(t)\neq1$ if $t\neq0$.

If ξ is a random variable whose values are concentrated in a subset of the real line consisting of finitely or countably many points, then for all numbers $\varepsilon>0$ there exists an integer $s=s(\varepsilon)<\infty$ and points $u_1,\dots,u_s$ on the real line such that $P(\xi\in\{u_1,\dots,u_s\})\ge1-\frac\varepsilon3$. Further, by a classical (and simple) result of the number theory, by a result of Dirichlet, for all numbers $N\ge1$ there exists an integer $1\le q_N\le N$ and a set of s integers $p_1,\dots,p_s$ such that $|q_Nu_k-p_k|\le N^{-1/s}$, for all indices $k=1,\dots,s$. Hence, by choosing the number N sufficiently large we can achieve with the choice $t=2\pi q_N$ that $\operatorname{Re}e^{itu_k}\ge1-\frac\varepsilon3$ for all indices $1\le k\le s$. Then $\operatorname{Re}Ee^{it\xi}\ge\sum\limits_{k=1}^s P(\xi=u_k)(1-\frac\varepsilon3)-\frac\varepsilon3\ge1-\varepsilon$. Since we can make this construction for all $\varepsilon>0$, there exists a sequence of positive integers q_N, $N=1,2,\dots$, such that the sequence $t_N=2\pi q_N$ satisfies the relation $\lim\limits_{N\to\infty}\varphi(t_N)=1$. Since the random variable ξ is not lattice distributed, hence $\sup\limits_{2\pi\le t\le B}|\varphi_n(t)|\le q<1$ for any $B>2\pi$ with an appropriate constant $q=q(B)<1$. This implies that the above constructed sequence of positive real numbers t_N satisfies the relation $t_N\to\infty$ if $N\to\infty$.

31.) Let us denote the distribution function of the random variable ξ by $F(u)$, its characteristic function by $\varphi(t)$ and put $\operatorname{Re}\varphi(t)=u(t)$. Then for all $h\ge 0$

$$\frac{1-u(h)}{h^2}=\int_{-\infty}^{\infty}\frac{1-\cos hu}{h^2u^2}u^2\,dF(u),$$

and since $\lim\limits_{h\to0}\frac{1-\cos hu}{h^2u^2}=\frac12$, $\frac{1-\cos hu}{h^2u^2}\ge0$ for all numbers $u\in R^1$ and $h\in R^1$, hence $\liminf\limits_{h\to0}\int_{-\infty}^{\infty}\frac{1-\cos hu}{h^2u^2}u^2\,dF(u)\ge\frac12\int_{-\infty}^{\infty}u^2\,dF(u)=\frac12E\xi^2$ by Fatou's lemma. Hence, to solve the problem it is enough to show that $\limsup\limits_{h\to0}\frac{1-u(h)}{h^2}<\infty$ if the function $\varphi(t)$ is twice differentiable in the origin. If the function $\varphi(t)$ is twice differentiable in the origin, then the same relation holds for the function $u(t)$. Then the derivative $u'(t)=\frac{du(t)}{dt}$ exists in a small neighbourhood of the origin, and $u'(0)=0$, as $u(\cdot)$ is an even function. Further, $u(0)=1$, $u(t)\le1$ for all numbers $t\in R^1$, hence $0\le\frac{1-u(h)}{h^2}=\frac{u(0)-u(h)}{h^2}=-\frac{u'(\vartheta h)}{h}\le\sup\limits_{0\le s\le h}\frac{u'(0)-u'(s)}{s}<\infty$ for small number $h>0$, if $u(\cdot)$ is twice differentiable in the origin, where $0\le\vartheta\le1$ is an appropriate number, and $u'(\cdot)$ denotes the derivative of the function $u(\cdot)$. Hence $\limsup\limits_{h\to0}\frac{1-u(h)}{h^2}<\infty$ in this case.

By induction with respect to the parameter k we can see that if ξ is a random variable with distribution function F, and the $2k$-th derivative of the characteristic function in the origin is finite, then the random variable ξ has finite $2k$-th moment. Indeed, by the induction hypothesis there exists a distribution function $F^{(k-1)}(du)=\frac{u^{(2k-2)}F(du)}{m_{2k-2}}$ where $m_{2k-2}=\int u^{2k-2}\,dF(u)$. Further, the characteristic function of the distribution function $F^{(k-1)}$ has finite second moment in the origin, since the characteristic function of this distribution function equals the $2k-2$-th derivative of the characteristic function of the distribution function F multiplied by $(-1)^{k-1}m_{2k-2}^{-1}$, and the characteristic function of a random variable with distribution function $F^{(k-1)}(u)$ has finite second derivative in the origin. Hence by the already proven part of this problem a random variable with distribution function $F^{(k-1)}$ has finite second moment. This is equivalent to the statement that a random variable with distribution function F has finite $2k$-th moment.

32.) The solution of this problem applies similar idea as the proof of problem 22. In problem 22 we deduced from some kind of continuity of the characteristic function of a distribution function in the origin some sort of estimate about the tail behaviour of the distribution function. In this problem we exploit that if we know more continuity about the distribution function in the neighbourhood of the origin, then we get sharper estimates about the tail behaviour of the distribution function.

Let $F(x)$ denote the distribution function of the random variable ξ, and put $u(t)=\operatorname{Re}\varphi(t)$. The estimate $\left|\frac{1-u(h)}{h}\right|=|u'(\vartheta h)|\le\text{const.}\,h^\alpha$ holds with an appropriate constant $0<\vartheta<1$ under the conditions of the problem. Then the following analog of formula (2.6) holds.

$$\text{const.}\, h^{1+\alpha} > \int_{-h}^{h} \frac{1-u(t)}{h}\,dt = \int_{-1/2h}^{1/2h} \left(1-\frac{\sin hx}{hx}\right)\,dF(x)$$
$$+\int_{|x|>\frac{1}{2h}} \left(1-\frac{\sin hx}{hx}\right)\,dF(x) \ge \int_{|x|>\frac{1}{2h}} \frac12\,dF(x)$$
$$=\frac12 P\left(|\xi|>\frac{1}{2h}\right).$$

This implies that $P(|\xi| > u|) \le \text{const.}\, u^{-1-\alpha}$ for all numbers $u > 0$. Let us introduce the function $G(u) = P(|\xi| > u|)$. Integration by parts yields that

$$E|\xi| = \int_0^\infty |u|\,dG(u) = [uG(u)]_0^\infty - \int G(u)\,du < \infty.$$

If the characteristic function $\varphi(t)$ is $2k+1$-times differentiable in a small neighbourhood of the origin, and the $2k+1$-th differential is a Lipschitz α function, $\alpha > 0$, in this small neighbourhood, then let us introduce the distribution function $F^{(k)}(du) = \frac{u^{2k}F(du)}{m_k}$, where $F(\cdot)$ denotes the distribution function of the random variable ξ, and $m_k = \int u^{2k}\,dF(u)$. The argument applied in the solution of the previous problem can be adapted to the present case. The already proven part of this problem can be applied to a random variable with distribution function $F^{(k)}$, and it yields that $E|\xi|^{2k+1} < \infty$.

33.) The relation $(-1)^k E\xi^{2k} = \left.\frac{d^{2k}\varphi(t)}{dt^{2k}}\right|_{t=0} = \frac{(2k)!}{2\pi i}\oint_{z=R}\frac{\varphi(z)}{z^{2k+1}}\,dz$ holds for all integers k by the result of problem 31 and the Cauchy integral formula if the circle with center in the origin and radius R is in the domain of analiticity of the function $\varphi(z)$. Since $\sup\limits_{z=R}|\varphi(z)| < \infty$ the above relation implies that $E\xi^{2k} \le (ak)^{2k}$ with some constant $a > 0$, and $P(|\xi| > x) \le \left(\frac{ak}{x}\right)^{2k}$ for all positive integers $k \ge 1$. If $x \ge C_0$ with some number $C_0 > 0$, then let us fix the constant $k = \left[\frac{x}{2a}\right]$, where $[u]$ is the greatest integer smaller than u. This implies that the inequality $P(|\xi| > x) < \text{const.}\, e^{-\alpha x}$ holds for all numbers $x > 0$ with an appropriate constant $\alpha > 0$.

Let us remark that we had to apply the above relatively complicated argument, because at the beginning of the proof we did not know that the analytic continuation of the characteristic function of a random variable ξ is always the function $\varphi(z) = Ee^{iz\xi}$.

34.) If the characteristic function of a random variable ξ is integrable, then the result about the inverse Fourier transform can be applied. It implies that the random variable ξ also has a density function $f(x)$, and the identity $f(x) = \frac{1}{2\pi}\int e^{-itx}\varphi(t)\,dt$ holds. Also the formula $\frac{d^k f(x)}{dx^k} = \frac{(-i)^k}{2\pi}\int t^k e^{-itx}\varphi(t)\,dt$ holds if the order of integration and differentiation can be changed in the inverse Fourier transform formula. In the solution of problem 27 we have proved that this change of order can be carried out if the functions $|t|^j|\varphi(t)|$ are integrable for all indices $0 \le j \le k$. Hence the statement of this problem holds if $|\varphi(u)| < \text{const.}\,|u|^{-(k+1+\varepsilon)}$ with some number $\varepsilon > 0$. If $|\varphi(u)| < \text{const.}\, e^{-\alpha|u|}$ with some constant $\alpha > 0$, then the function $f(z) = \frac{1}{2\pi}\int e^{-itz}\varphi(t)\,dt$ is an analytic continuation of the density function $f(x)$ of the random variable ξ.

35.) Let $\varphi(t) = E^{it\xi}$ denote the characteristic function of the random variable ξ_1. The normalized partial sum $\frac{S_n}{\sqrt{n}}$ has the characteristic function $\varphi^n\left(\frac{t}{\sqrt{n}}\right)$. Hence by the Fundamental Theorem about convergence in distribution and the result of problem 13a) it is enough to show that $\varphi^n\left(\frac{t}{\sqrt{n}}\right) \to e^{-t^2/2}$ for all numbers $t \in R^1$ if $n \to \infty$. On the other hand, a Taylor expansion of the function $\varphi(t)$ around the point $t = 0$ and the result of problem 27 yield that $\varphi(t) = 1 - \frac{t^2}{2} + o(t^2)$ if $t = o(1)$, and $\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{n} + o\left(\frac{1}{n}\right) = e^{-t^2/2n + o(n^{-1})}$ if $t = O(1)$. The last relation implies that $\varphi^n\left(\frac{t}{\sqrt{n}}\right) = e^{-t^2/2+o(1)} \to e^{-t^2/2}$ for all fixed number t, if $n \to \infty$.

36.) Let us introduce the function $F(t) = e^{it} - \left(1 + \frac{it}{1!} + \cdots + \frac{(it)^k}{k!}\right)$, and let us consider its derivatives $F^{(j)}(t)$, $j = 1, 2, \dots, k$. Observe that $F^{(j)}(0) = 0$ of all $0 \le j \le k$, and $|F^{(k+1)}(t)| = |e^{ikt}| = 1$ for all $t \in R^1$. We get by induction with respect to the parameter j that $|F^{(j)}(t)| \le \int_0^t |F^{(j+1)}(s)|\,ds \le \int_0^t \frac{|s|^{k-j}\,ds}{(k-j)!} = \frac{|t|^{k+1-j}}{(k+1-j)!}$ for all indices $j = k+1, k, \dots, 0$. In particular, $|F(t)| \le \frac{|t|^{k+1}}{(k+1)!}$, and this is the statement of the problem.

37.a) By applying formula (11) with $k = 1$ we get that $|e^{it\xi} - 1 - it\xi| \le \frac{t^2\xi^2}{2}$. Taking the expected value of the expression at the left-hand side between the absolute value sign we get that $|\varphi(t) - 1| \le \frac{t^2}{2}E\xi^2$. If $E\xi^2 < \varepsilon$ with a sufficiently small number $\varepsilon = \varepsilon(t) > 0$, then $|1 - \varphi(t)| \le \frac{1}{4}$, and $|\log\varphi(t) + (1 - \varphi(t))| = |\log(1 - (1 - \varphi(t))) - (1 - \varphi(t))| \le |1 - \varphi(t)|^2 \le t^4\left(E\xi^2\right)^2$.

b.) The sequence of random variables S_k converge in distribution to a normal random variable with expected value m and variance σ^2 if and only if

$$\lim_{k\to\infty} \prod_{j=1}^{n_k} \varphi_{k,j}(t) = e^{-\sigma^2t^2/2+imt} \quad \text{for all } t \in R^1. \tag{2.7}$$

Let us take the logarithm in the relation (2.7). We claim that formula (2.7) is equivalent to formula (2.7′) formulated below.

$$\lim_{k\to\infty} \sum_{j=1}^{n_k} \log\varphi_{k,j}(t) = -\frac{\sigma^2t^2}{2} + imt \quad \text{for all } t \in R^1. \tag{2.7′}$$

The equivalence of relations (2.7) and (2.7′) is less obvious than it may seem at first sight. Some difficulty arises because in the space of complex numbers where we have to work the equation $e^{z_1} = e^{z_2}$ only implies that $z_1 = z_2 + i2k\pi$ with some integer k, but the numbers z_1 and z_2 may be different. Hence although the implication $(2.7') \Rightarrow (2.7)$ needed in the proof of the central limit theorem is straightforward, the proof of the implication $(2.7) \Rightarrow (2.7')$ needed in the proof of the converse of the central limit theorem requires a more careful argument.

Before the proof of the implication $(2.7) \Rightarrow (2.7')$ let us remark that because of the uniform smallness condition for a fixed value $t \in R^1$ the characteristic function $\varphi_{k,j}(t)$ we consider is in a small neighbourhood of the number 1 for all $1 \le j \le n_k$ if the index k is sufficiently large. Hence for a sufficiently large $k \ge k_0(t) > 0$ there is a version $\log \varphi_{k,j}(t)$ of the functions $\varphi_{k,j}(t)$ which is in a small neighbourhood of the origin, say $|\log \varphi_{k,j}(t)| < \frac{1}{2}$. We take this version of the logarithm in formula $(2.7')$.

To prove the implication $(2.7) \Rightarrow (2.7')$ let us first make the following observation. We have seen in the proof of problem 24 (in the proof of the Fundamental Theorem about convergence of distribution functions) that the convergence in formula (2.7) is uniform in all finite intervals. Besides, the right-hand side of formula (2.7) is separated both from zero and infinity in a finite interval. Hence we get by taking logarithm in formula (2.7) that for all $\varepsilon > 0$ and $T > 0$ there exists a $k_0 = k_0(\varepsilon, T)$ and $\delta = \delta(\varepsilon, k, T)$ such that

$$\left| \sum_{j=1}^{n_k} \log \varphi_{k,j}(t) - \left(-\frac{\sigma^2 t^2}{2} + imt \right) + i2\pi l_k(t) \right| < \varepsilon \quad \text{if } |t| \le T \text{ and } k \ge k_0 \tag{2.7''}$$

with some integer $l_k(t)$ which may depend both on k and t. We have to show that $l_k(t) \equiv 0$ in formula $(2.7'')$. First we prove the weaker statement that $l_k(t) = l_k$, i.e. this constant in formula $(2.7'')$ does not depend on t. Indeed, otherwise for all $\delta = \delta_k > 0$ a pair of constant $-T \le s, t \le T$, $|t - s| < \delta$ could be found such that $l_k(t) \ne l_k(s)$, i.e. $|2\pi i(l_k(t) - l_k(s)| \ge 2\pi$. But this is not possible, because the function $g_k(t) = \sum\limits_{j=1}^{n_k} \log \varphi_{k,j}(t)$ is uniformly continuous in the interval $[-T, T]$. Hence fixing a small $\varepsilon > 0$ we can write $|g_k(t) - g_k(s)| < \varepsilon$ for sufficiently small $\delta = \delta(k, \varepsilon, T) > 0$. Besides, also the inequality $\left| \left(-\frac{\sigma^2 t^2}{2} + imt \right) - \left(-\frac{\sigma^2 s^2}{2} + ims \right) \right| < \varepsilon$ holds if $\delta > 0$ is sufficiently small. Let $\varepsilon < \frac{1}{3}$. The above relations together with formula $(2.7'')$ would contradict to the assumption $|2\pi i(l_k(t) - l_k(s)| \ge 2\pi$. Hence $l_k(t) = l_k$. Finally, it is easy to see that $l_k = l_k(0) = 1$. Hence relation $(2.7'')$ holds for all $\varepsilon > 0$ with $l_k(t) \equiv 0$, and this implies relation $(2.7')$.

Because of the uniform smallness condition of problem 37 and the already proved part a) of this problem we can write for $k \ge k_0(t)$

$$\left| \sum_{j=1}^{n_k} \log \varphi_{k,j}(t) - \sum_{j=1}^{n_k} (\varphi_{k,j}(t) - 1) \right| \le t^4 \sum_{j=1}^{n_k} \left(E\xi_{k,j}^2 \right)^2 \le \text{const.}\, t^4 \max_{1 \le j \le n_k} E\xi_{k,j}^2,$$

since $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k} E\xi_{j,k}^2 = 1$. Because of this relation and the uniform smallness condition

$$\lim_{k\to\infty}\left|\sum_{j=1}^{n_k}\log\varphi_{k,j}(t)-\sum_{j=1}^{n_k}(\varphi_{k,j}(t)-1)\right|=0.$$

These relations also imply part b) of problem 37.

38.) Let us fix a number $\varepsilon>0$. Then

$$E\xi_{k,j}^2=E\xi_{k,j}^2 I(\{(|\xi_{k,j}|<\varepsilon\})+E\xi_{k,j}^2 I(\{|\xi_{k,j}|\ge\varepsilon\})\le\varepsilon^2+\sum_{j=1}^{n_k}E\xi_{k,j}^2 I(\{|\xi_{k,j}|\ge\varepsilon\}),$$

hence by the Lindeberg condition $\limsup\limits_{k\to\infty}\sup\limits_{1\le j\le n_k} E\xi_{k,j}^2\le\varepsilon^2$. Since this formula holds for all numbers $\varepsilon>0$, it implies the uniform smallness property.

By formula (12) (with the choice $m=0$ and $\sigma=1$) to prove the central limit theorem it is enough to show that under the conditions of this problem

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}(\varphi_{k,j}(t)-1)=\lim_{k\to\infty}\sum_{k=1}^{n_k}E\left(e^{it\xi_{k,j}}-1-it\xi_{k,j}\right)\to-\frac{t^2}{2},$$

or since $\lim\limits_{k\to\infty}\sum\limits_{j=1}^{n_k}E\xi_{k,j}^2=1$ it is enough to show that

$$\lim_{k\to\infty}\sum_{j=1}^{n_k}E\left(e^{it\xi_{k,j}}-1-it\xi_{k,j}+\frac{t^2}{2}\xi_{k,j}^2\right)\to 0.$$

By applying formula (11) for $k=2$ if $|tx|\le\varepsilon$ and for $k=1$ if $|tx|\ge\varepsilon$ together with the Lindeberg condition we get that

$$\begin{aligned}\left|\sum_{j=1}^{n_k}E\left(e^{it\xi_{k,j}}-1-it\xi_{k,j}+\frac{t^2}{2}\xi_{k,j}^2\right)I(\{|\xi_{k,j}|\le\varepsilon\})\right|&\le\sum_{j=1}^{n_k}E\frac{|t\xi_{k,j}|^3}{6}I(\{|\xi_{k,j}|\le\varepsilon\})\\&\le\varepsilon|t|^3\sum_{j=1}^{n_k}E\frac{\xi_{k,j}^2}{6}\le\text{const.}\,\varepsilon,\end{aligned}$$

and

$$\begin{aligned}&\lim_{k\to\infty}\left|\sum_{j=1}^{n_k}E\left(e^{it\xi_{k,j}}-1-it\xi_{k,j}+\frac{t^2}{2}\xi_{k,j}^2\right))I(\{|\xi_{k,j}|>\varepsilon\})\right|\\&\qquad\le\lim_{k\to\infty}\sum_{j=1}^{n_k}Et^2\xi_{k,j}^2 I(\{|\xi_{k,j}|>\varepsilon\})=0.\end{aligned}$$

Since these relations hold for all $\varepsilon > 0$, it implies formula (12) with the choice $m = 0$ and $\sigma^2 = 1$. In such a way we have solved problem 38.

39.) If the conditions of the problem are satisfied, then $\lim\limits_{k\to\infty} \sum\limits_{j=1}^{n_k} \operatorname{Re}(\varphi_{k,j}(t)-1) = -\frac{t^2}{2}$. Further, since the sum of the random variables in the k-th row is almost 1 for large index k, hence

$$\lim_{k\to\infty} \sum_{j=1}^{n_k} E\left(\cos(t\xi_{k,j}) - 1 + \frac{t^2\xi_{k,j}^2}{2}\right) = 0 \quad \text{for all numbers } t \in R^1.$$

Let us observe that $\cos u - 1 + \frac{u^2}{2} \ge 0$ for all numbers $u \in R^1$, since we have for the function $F(u) = \cos u - 1 + \frac{u^2}{2} \ge 0$, $F(0) = 0$, $F'(0) = 0$ and $F''(u) = 1 - \cos u \ge 0$ for all numbers $u \in\in R^1$. Besides, $\cos u - 1 + \frac{u^2}{2} \ge \frac{u^2}{4}$ if $|u| > 3$. The above inequalities imply that

$$\lim_{k\to\infty} \sum_{j=1}^{n_k} \frac{t^2}{4} E\xi_{k,j}^2 I\left(\left\{|\xi_{k,j}| \ge \frac{3}{t}\right\}\right) = 0.$$

We get the solution of problem 39 from this relation with the choice $t = \frac{3}{\varepsilon}$.

40.) By the Schwarz inequality

$$(E\xi_{k,j} I(|\xi_{k,j}| \le \varepsilon)^2 = (E\xi_{k,j} I(|\xi_{k,j}| > \varepsilon)^2 \le E\xi_{k,j}^2 I(|\xi_{k,j}| > \varepsilon).$$

Hence the Lindeberg condition implies that

$$\lim_{k\to\infty} \sum_{j=1}^{n_k} (E\xi_{k,j} I(|\xi_{k,j}| \le \varepsilon))^2 \le \lim_{k\to\infty} \sum_{j=1}^{n_k} E\xi_{k,j}^2 I(|\xi_{k,j}| > \varepsilon) = 0,$$

and

$$\lim_{k\to\infty} \sum_{j=1}^{n_k} E\xi_{k,j}^2 I(|\xi_{k,j}| \le \varepsilon) = 1.$$

These two relations imply the first statement of the problem.

To prove the second statement let us first observe that $E(\bar\xi_{k,j} - \xi_{k,j}) = E\bar\xi_{k,j} - E\xi_{k,j} = 0$, $1 \le j \le n_k$. Hence the Chebishev inequality and Lindeberg condition imply that for all $\varepsilon > 0$

$$P(|S_k - \bar S_k| > \varepsilon) \le \frac{\operatorname{Var}(S_k - \bar S_k)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{j=1}^{n_k} \operatorname{Var}(\xi_{k,j} - \bar\xi_{k,j})$$

$$\le \frac{1}{\varepsilon^2} \sum_{j=1}^{n_k} E\xi_{k,j}^2 I(|\xi_{k,j}| > 0) \to 0.$$

Since this relation holds for all $\varepsilon > 0$ it implies the second statement of the problem.

41.) By the Hölder inequality

$$\sum_{k=1}^n E\xi_k^2 I(|\xi_k| > \varepsilon s_n) \le \sum_{k=1}^n \left(E|\xi_k|^{(2+\alpha)}\right)^{(2/\alpha+2)} P(|\xi_k| > \varepsilon s_n)^{\alpha/(2+\alpha)}$$
$$\le \left(\sum_{k=1}^n E|\xi_k|^{(2+\alpha)}\right)^{(2/\alpha+2)} \left(\sum_{k=1}^n P(|\xi_k| > \varepsilon s_n)\right)^{\alpha/(2+\alpha)}.$$

On the other hand, by the Chebishev inequality $\sum\limits_{k=1}^n P(|\xi_k| > \varepsilon s_n) \le \sum\limits_{k=1}^n \frac{E\xi_k^2}{\varepsilon^2 s_n^2} = \frac{1}{\varepsilon^2}$. Hence it the conditions of part a) of problem 41 hold, then

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^n E\xi_k^2 I(|\xi_k| > \varepsilon s_n) = 0,$$

i.e. these conditions imply the Lindeberg condition.

In the case considered at the end of part a.) $s_n^2 \ge \text{const.}\, n$, and $\sum\limits_{k=1}^n E|\xi_k|^{2+\alpha} = o\left(n^{(\alpha+2)/2)}\right)$ if $n \to \infty$, hence in this case the condition formulated in part a) of problem 41 is satisfied.

If $\xi_1, \xi_2, \dots,$ is a sequence of independent and identically distributed random variables, $E\xi_1 = 0$, $0 < E\xi_1^2 < \infty$, then

$$\frac{1}{s_n^2} \sum_{k=1}^n E\xi_k^2 I(|\xi_k| > \varepsilon s_n) = \frac{1}{E\xi_1^2} E\xi_1^2 I\left(|\xi_1| > \varepsilon\sqrt{nE\xi_1^2}\right) \to 0$$

if $n \to \infty$. This means that the Lindeberg condition holds also in this case.

42.) If the point x is a point of continuity of the limit distribution function $F(\cdot)$, then for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $F(x) - \frac{\varepsilon}{2} < F(x-\delta) < F(x) < F(x+\delta) < F(x+\delta) + \frac{\varepsilon}{2}$. Since the monotone increasing function $F(\cdot)$ has at most countably infinite points of discontinuity, hence we may assume without violating the generality that we choose the point $\delta > 0$ in such a way that the points $x \pm \delta$ are also points of continuity of the function $F(\cdot)$. Then there exists an index $n_0 = n_0(\delta, \varepsilon)$ such that $P(S_n < x+\delta) < F(x+\delta) + \frac{\varepsilon}{4}$, $P(S_n > x-\delta) < 1 - F(x-\delta) + \frac{\varepsilon}{4}$, and $P(|T_n| \ge \delta) < \frac{\varepsilon}{4}$ if $n \ge n_0$. Then $P(S_n + T_n < x) \le P(S_n < x+\delta) + P(|T_n| > \delta) < F(x+\delta) + \frac{\varepsilon}{2} < F(x) + \varepsilon$ if $n \ge n_0(\varepsilon, \delta)$. We get in a similar way that $P(S_n + T_n > x) < 1 - F(x) + \varepsilon$ if $n \ge n_0(\varepsilon, \delta)$. Since the above statements hold for all $\varepsilon > 0$, they imply the statement of the problem.

43.) Let the independent random variables ξ_n, $n = 1, 2, \dots,$ have the following distribution: $P(\xi_n = n) = P(\xi_n = -n) = \frac{1}{4n^2}$, $P(\xi_n = 2) = P(\xi_n = -2) = \frac{1}{4}$, and $P(\xi_n = 0) = \frac{1}{2} - \frac{1}{2n^2}$, $n = 1, 2, \dots$. Then $E\xi_n = 0$, $E\xi_n^2 = 1$. Put

$X_k=\xi_kI(|\xi_k|\le 2)$, $Y_n=\xi_kI(|\xi_k|>2)$ for all $k=1,2,\dots$. Consider the partial sums $S_n=\sum\limits_{k=n}^{n}X_k$ and $T_n=\sum\limits_{k=1}^{n}Y_k$, $n=1,2,\dots$. Then the normalized partial sums $\sqrt{\frac{2}{n}}S_n$ converge in distribution to the standard normal distribution. Indeed, the partial sums of the random variables X_k, $k=1,2,\dots$, satisfy the conditions of the central limit theorem, and $EX_k^2=\frac12$, $k=1,2,\dots$. On the other hand, the expressions $\sqrt{\frac{2}{n}}T_n$ converge stochastically to zero if $n\to\infty$. Indeed, $\sum\limits_{k=1}^{\infty}P(Y_k\neq 0)<\infty$, hence with probability 1 only finitely many terms $Y_k(\omega)$ do not equal zero, and $\sum\limits_{k=1}^{\infty}|Y_k(\omega)|\le K(\omega)$. Since $\sqrt{\frac{2}{n}}\sum\limits_{k=1}^{n}\xi_k=\sqrt{\frac{2}{n}}S_n+\sqrt{\frac{2}{n}}T_n$, the above calculation and the result of problem 42 imply that the above construction yields an example for the statement of part a) of problem 43.

Let us make some slight modifications in the construction of the above random variables ξ_n. Let us put similarly to the previous construction $P(\xi_n=2)=P(\xi_n=-2)=\frac14$ and $P(\xi_n=n)=\frac{1}{4n^2}$. Let us define further

$$P\left(\xi_n=\frac{1}{\sqrt{n}}\right)=\frac12-\frac{1}{2n^2},\quad\text{and}\quad P\left(\xi_n=-n-2n^{3/2}\left(1-\frac{1}{n^2}\right)\right)=\frac{1}{4n^2},$$

$n=1,2,\dots$. Then $E\xi_n=0$, $n=1,2,\dots$. By applying the truncation technique of part a) and carrying out a natural modification of the calculation following it we get that these random variable ξ_n, $n=1,2,\dots$, yield an example for part b) of problem 43.

44.) Let us choose an arbitrary number L such that $\int u^2F_0(\,du)>L$. It is enough to show that $\liminf\limits_{n\to\infty}\int u^2F_n(\,du)\ge L$. There exists such a bounded and continuous function $g(u)=g_L(u)$ for which $g(u)\le u^2$, and $\int g(u)F_0(\,du)\ge L$. Indeed, the function $g(u)=g_L(u)=\min(u^2,K)$ satisfies this property if we choose a sufficiently large constant $K=K(L)>0$. Then the characterization of the convergence in distribution given in Theorem A implies that $\liminf\limits_{n\to\infty}\int u^2F_n(\,du)\ge\lim\limits_{n\to\infty}\int g(u)F_n(\,du)=\int g(u)F_0(\,du)\ge L$.

45.) The distribution of the random vector $\mathbf{Z}=(Z_1,\dots,Z_m)$ is determined by its characteristic function. (See the result of problem 19.) On the other hand, the characteristic function $E^{i(t_1Z_1+\cdots+t_mZ_m)}$ of the random vector $(Z_1,\dots,Z_m)$ in the point $(t_1,\dots,t_m)$ agrees with the characteristic function of the random variable $Z(t_1,\dots,t_m)$ in the point 1. Hence the characteristic function and distribution function of the random vector $\mathbf{Z}$ is determined by the distribution of the above considered one dimensional distributions.

46.) First we show with the help of the Fundamental Theorem about the convergence of distribution functions that the random vectors $\mathbf{Z}_n=(Z_{1,n},\dots,Z_{m,n})$, $n=1,2,\dots$, converge in distribution to some m-dimensional distribution as $n\to\infty$ if the one-dimensional random variables $Z_n=Z_n(a_1,\dots,a_m)$, $n=1,2,\dots$, converge in distribution for all real numbers $a_1,\dots,a_m$ as $n\to\infty$. Indeed, if the one-dimensional random variables interested in this problem convergence in distribution, then the characteristic functions of the random vectors $\mathbf{Z}_n$, $n = 1, 2, \dots$, converge to a function $\varphi(t_1, \dots, t_m)$ in all points $(t_1, \dots, t_m) \in R^m$, and the restriction of this limit function to the coordinates axes is continuous. Hence by the Fundamental Theorem the random vectors $\mathbf{Z}_n$, $n = 1, 2, \dots$, also converge in distribution, and the characteristic function of the limit distribution is the above limit function. The Fundamental Theorem also implies that the convergence of the random vectors $\mathbf{Z}_n$ implies the convergence of the random variables of the random variables $Z_n(a_1, \dots, a_m)$ in distribution.

If the random vectors $\mathbf{Z}_n$ converge in distribution, then the limit distribution is determined by its characteristic function which is the limit of the characteristic functions of these random variables. Similarly, the characteristic function of the limit of the one-dimensional random variables we have considered equals the limit of the characteristic function of these random variables. These facts imply the characterization of the limit distribution μ given in this problem together with the statement that the above characterization determines the limit distribution in a unique way.

47.) Let $\Sigma = (D_{j,k})$, $1 \le j, k \le m$, denote the covariance matrix of an m-dimensional random vector $(Z_1, \dots, Z_m)$, i.e. let $D_{j,k} = E(Z_j - EZ_j)(Z_k - EZ_k)$, $1 \le j, k \le m$. Then the matrix Σ is symmetrical. Besides, $\mathbf{x}\Sigma\mathbf{x}^* = \sum\limits_{j=1}^m \sum\limits_{k=1}^m x_j E(Z_j - EZ_j)(Z_k - EZ_k)x_k = E\left(\sum\limits_{j=1}^k x_j(Z_j - EZ_j)\right)^2 \ge 0$ for all vectors $x = (x_1, \dots, x_m) \in R^m$, and this means that the matrix Σ is positive semi-definite.

On the other hand, if Σ is an arbitrary $m \times m$ positive semi-definite matrix, then the results of linear algebra imply that there exists a matrix B such that $\Sigma = B^*B$. (The matrix B satisfying this relation is not determined in a unique way. A possible construction of a matrix B satisfying the above relation can be given in the following way: It is known from linear algebra that a symmetric matrix Σ can be represented in the form $\Sigma = U\Lambda U^*$ where the matrix U is unitary and the matrix Λ is diagonal with some elements $\lambda_1, \dots, \lambda_m$ in the diagonal. The matrix Σ is positive semi-definite if and only if all elements λ_j, $1 \le j \le m$, in the above representation are non-negative. If $\Sigma = U\Lambda U^*$ is a positive semi-definite matrix, then let us define the symmetric matrix $B = U\sqrt{\Lambda}U^*$, where $\sqrt{\Lambda}$ is the diagonal matrix with elements $\sqrt{\lambda_j}$, $j = 1, \dots, m$, in the diagonal. Then $\Sigma = B^2 = B^*B$.)

Let $\Sigma = (D_{j,k})$, $1 \le j, k \le m$, be an arbitrary $m \times m$ positive semi-definite matrix, and $\mathbf{M} = (M_1, \dots, M_m) \in R^m$ a vector in the Euclidean space R^m. Let $\xi = (\xi_1, \dots, \xi_m)$ be an m-dimensional random variable with standard normal distribution, $B = (b_{j,k})$, $1 \le j, k \le m$, an $m \times m$ matrix such that $B^*B = \Sigma$. Let us define the m-dimensional random vector $\eta = (\eta_1, \dots, \eta_m) = \xi B + \mathbf{M}$. Then η has normal distribution, and we claim that it has expected value $\mathbf{M}$ and covariance matrix $B^*B = \Sigma$. This implies that for all vectors $\mathbf{M} \in R^m$ and $m \times m$ positive semi-definite matrices Σ there exists an m-dimensional normally distributed random vector with expectation zero and covariance matrix Σ.

Indeed, $E\eta = (E\eta_1, \dots, E\eta_m) = (M_1, \dots, M_m) = \mathbf{M}$, and the elements of the covariance matrix of η can be calculated in the following way.

$$E(\eta_j - E\eta_j)(\eta_k - E\eta_k) = E\left(\sum_{l=1}^{m} b_{j,l}\xi_l\right)\left(\sum_{p=1}^{m} b_{k,p}\xi_k\right) = \sum_{l=1}^{p} b_{j,l}b_{k,l} = D_{j,k}$$

for all indices $1 \le j,k \le m$, because $E\xi_l\xi_p = 0$, if $l \ne p$, $E\xi_l^2 = 1$. The last identity means the identity $B^*B = \Sigma$ in coordinate form.

Let us consider an m-dimensional random vector $\eta = (\eta_1, \dots, \eta_m) = \xi B + \mathbf{M}$ with normal distribution function, where the random vector $\xi = (\xi_1, \dots, \xi_m)$ has standard normal distribution, B is an $m \times m$ matrix, $\mathbf{M} = (M_1, \dots, M_m) \in R^m$. Put $B^*B = \Sigma = (D_{j,k})$, $1 \le j,k \le m$. Let us calculate the characteristic function $\varphi(t_1, \dots, t_m) = Ee^{i(t_1\eta_1 + \dots + t_m\eta_m)}$ η of the random vector η. To calculate it let us introduce the random variable $\zeta = t_1\eta_1 + \dots + t_m\eta_m$. Then ζ is a normally distributed random variable with expected value $\bar{M} = \bar{M}(t_1, \dots, t_m) = \sum_{k=1}^{m} t_k M_k = (\mathbf{M}, \mathbf{t})$ and variance $\sigma^2 = \sum_{j=1}^{m}\sum_{k=1}^{m} t_j t_k E\eta_j\eta_k = \sum_{j=1}^{m}\sum_{k=1}^{m} t_j t_k D_{j,k} = \mathbf{t}\Sigma\mathbf{t}^*$, where $\mathbf{t} = (t_1, \dots, t_m)$. Hence the characteristic function of the random variable ζ with normal distribution equals $\psi(u) = Ee^{iu\zeta} = e^{-u^2\mathbf{t}\Sigma t/2 + i(\mathbf{M},t)u}$. This implies that $\varphi(t_1, \dots, t_m) = Ee^{i(t_1\eta_1 + \dots + t_m\eta_m)} = \psi(1) = e^{-\mathbf{t}\Sigma\mathbf{t}^*/2 + i(\mathbf{M},\mathbf{t})}$, that is relation (13) holds.

It follows from formula (13) that if η is an m-dimensional random variable with normal distribution, then the characteristic function and as a consequence, the distribution function of this random vector is determined by its expectation and covariance matrix. Let us remark that there can be given two different $m \times m$ matrices B_1 and B_2 such that $B_1^*B_1 = B_2^*B_2$. Let ξ be an m-dimensional random vector with standard normal distribution, B_1 and B_2 two $m \times m$ matrices such that $B_1^*B_1 = B_2^*B_2$, an $\mathbf{M} \in R^m$ an arbitrary vector. Let us define the random vectors $\eta_1 = \xi B_1 + \mathbf{M}$ and $\eta_2 = \xi B_2 + \mathbf{M}$. Then the expectation and covariance matrix, hence the distribution function of the random vectors η_1 and η_2 agree, although this statement is not self-evident because of the relation $B_1 \ne B_2$.

48.) We get similarly to the proof of formula (13) in problem 47., that the characteristic function of the random vector $\eta = (\eta_1, \dots, \eta_l)$ equals $Ee^{i(\mathbf{t},\eta)} = e^{-\mathbf{t}\Sigma\mathbf{t}^*/2 + i(\mathbf{M},\mathbf{t})}$, where $\mathbf{t} = (t_1, \dots, t_l)$, and $\Sigma = B^*B$. It follows from this formula that η is a normally distributed random vector with covariance matrix Σ and expectation vector $\mathbf{M}$.

Given a normally distributed random vector η of dimension m, let us write it in the form $\eta = \xi B + \mathbf{M}$, where ξ is a random vector of dimension m with standard normal distribution. (It can be proved that such a representation is always possible. Actually it would be enough for us such a representation of a random vector with the same distribution as η. The possibility of such a representation follows from the definition of the normally distributed random vectors.) If we omit some of the coordinates of the random vector η and we preserve only l coordinates, then the random vector η' we obtain after this delition of coordinates can be presented in the following way. Let us omit those rows of the vector B which have the same indices as the elements of η we omitted. Similarly let us omit the coordinates of the vector $\mathbf{M}$ with the same indices as the coordines omitted from η. Let us denote the matrix and vector obtained in such a way by $\mathbf{M}'$ and B'. Then we have $\eta' = \xi B' + \mathbf{M}'$, hence η' is normally distributed by the already proved part of the problem.

49.) I shall give two different solutions of the problem, because, this may be instructive.

First soulution: We can write up the characteristic function of the random vector η, as $Ee^{i(\mathbf{t},\eta)} = e^{-\mathbf{t}\Sigma\mathbf{t}^*/2+i(\mathbf{M},\mathbf{t})}$ where $\mathbf{t} = (t_1, \dots, t_m)$, and $\mathbf{M} = (M_1, \dots, M_m)$ is the expected value of η, with the introduction of some notations in the following way. Let $\mathbf{t}_j$ denote the restriction of the vector $\mathbf{t}$, and let $\mathbf{M}_j$ denote the restriction of vector $\mathbf{M}$ to the coordinates $p \in L_j$, $1 \le j \le k$. Let us define similarly the matrix Σ_j as the restriction of the matrix Σ to the coordinates $\sigma_{p,q}$, $p \in L_j$ and $q \in L_j$, $1 \le j \le k$. With such a notation $Ee^{i(\mathbf{t},\eta)} = \prod\limits_{j=1}^k e^{-\mathbf{t}_j\Sigma_j\mathbf{t}_j^*/2+i(\mathbf{M}_j,\mathbf{t}_j)}$. (We exploited the properties of the matix Σ at this point.) Let $\eta'_1, \dots, \eta'_k$ independent, normally distributed random vectors with covariance matrix Σ_j and expected value $\mathbf{M}_j$, $1 \le j \le k$. The characteristic function of $\mathrm{E}\eta'_j$ equals $Ee^{i(\mathbf{t}_j,\eta'_j)} = e^{-\mathbf{t}_j\Sigma_j\mathbf{t}_j^*/2+i(\mathbf{M}_j,t_j)}$ for all indices $1 \le j \le k$. Hence the characteristic function of the random vectors $\eta' = (\eta'_1, \dots, \eta'_k)$ and η agree. Hence the distribution of η and η' also agree, and as a consequence the random vectors $\bar\eta_j$, $1 \le j \le k$, are independent of each other, similarly to the random vectors η'_j.

Second solution: Let us apply the notations introduced in the previous solution. The random vector η' defined there is normally distributed with covariance matrix Σ and expected value $\mathbf{M}$. As the distribution of a normal random vector is determined by its covariance matrix and expected value, hence the distribution of η és η' agree. Therefore the random vectors $\eta'_1, \dots, \eta'_k$ are independent, similarly to the random vectors $\bar\eta_1, \dots, \bar\eta_k$.

50.) Because of the result of problem 46 it is enough to show that for all vectors $\mathbf{a} = (a_1, \dots, a_m) \in R^m$ the normalized partial sums $\frac{1}{A_n}S_n = \frac{1}{A_n}S_n(a_1, \dots, a_m) = \frac{1}{A_n}\sum\limits_{p=1}^m a_pS_{p,n}$, $n = 1, 2, \dots$ converge in distribution to the normal distribution with expectation zero and variance $\sigma^2 = \mathbf{a}\Sigma\mathbf{a}^*$ if $n \to \infty$. Let us observe that $\frac{1}{A_n}S_n = \frac{1}{A_n}\sum\limits_{k=1}^n \eta_k$, $n = 1, 2, \dots$, where $\eta_k = \sum\limits_{p=1}^m a_p\xi_{p,k}$, $k = 1, 2, \dots$. Besides, the random variables η_k, $k = 1, 2, \dots$ are independent, $E\eta_k = 0$, $E\eta_k^2 = \mathbf{a}\Sigma_k\mathbf{a}^*$.

Let us consider separately the cases $\mathbf{a}\Sigma\mathbf{a}^* = 0$ and $\mathbf{a}\Sigma\mathbf{a}^* > 0$. If $\mathbf{a}\Sigma\mathbf{a}^* = 0$, then $\frac{1}{A_n^2}ES_n^2 = \frac{1}{A_k^2}\sum\limits_{j=1}^k \mathbf{a}\Sigma_j\mathbf{a}^* \to 0$ if $k \to 0$, hence $\frac{S_n}{A_n}$ converges to zero stochastically if $n \to \infty$. That is, in this case the random variables $\frac{1}{A_n}S_n$ converge in distribution to the (degenerated) normal variable with expectation zero and variance zero. In the case $\mathbf{a}\Sigma\mathbf{a}^*>0$ let us define the triangular array $\eta_{k,n}=\frac{\eta_k}{A_n\sqrt{\mathbf{a}\Sigma\mathbf{a}^*}}$, $1\le k\le n$, $n=1,2,\dots$. Then $\sum\limits_{k=1}^n E\eta_{k,n}^2=\frac{1}{A_n^2}\sum\limits_{k=1}^n\frac{\mathbf{a}\Sigma_k\mathbf{a}^*}{\mathbf{a}\Sigma\mathbf{a}^*}\to1$ if $n\to\infty$. Hence in the case $\mathbf{a}\Sigma\mathbf{a}^*>0$ the convergence of the normalized partial sums $\frac{1}{A_n}S_n$, $n=1,2,\dots$, to the normal distribution with expectation zero and variance $\mathbf{a}\Sigma\mathbf{a}^*$ follows from the central limit theorem for triangular arrays formulated in problem 38 if we show that the above defined triangular array $\eta_{k,n}$, $1\le k\le n$, $n=1,2,\dots$, satisfies the Lindeberg condition. This implies the statement of the problem. To prove the Lindeberg condition first we show the following inequality (2.8). Here we shall apply the notation $K=\max\limits_{1\le p\le m}|a_p|$.

$$E(\eta_{k,n}^2I(|\eta_{k,n}|>\varepsilon)\le\frac{K^2m^2}{A_n^2\mathbf{a}\Sigma\mathbf{a}^*}\sum_{p=1}^m E\xi_{p,k}^2I(|\xi_{p,k}|>\bar\varepsilon A_n) \tag{2.8}$$

for all indices $1\le k\le n$ and $n=1,2,\dots$, where $\bar\varepsilon=\bar\varepsilon(k)=\frac{\varepsilon}{m\sup\limits_{1\le p\le m}|a_p|}\cdot\frac{1}{\sqrt{\mathbf{a}\Sigma\mathbf{a}^*}}$. By summing up these inequalities for all indices $1\le k\le n$, and applying formula (14) for all $1\le p\le m$, we get that the triangular array $\eta_{k,n}$, $1\le k\le n$, $n=1,2,\dots$ satisfies the Lindeberg condition.

To prove formula (2.8) let us introduce the random index $\bar p(k)=\bar p(k,\omega)$ which is the (smallest) number p such that $|\xi_{\bar p(k),k}(\omega)|=\max\limits_{1\le p\le m}|\xi_{p,k}(\omega)|$. Let us observe that

$$\{\omega\colon\ |\eta_{k,n}(\omega)|>\varepsilon\}\subset\bigcup_{p=1}^M\{\omega\colon\ |\xi_{p,k}(\omega)|>\bar\varepsilon A_n\},$$

hence $I(|\eta_k(\omega)|>\varepsilon)\le I(|\xi_{\bar p(k,\omega),k}(\omega)|>\bar\varepsilon A_n)$. Besides, $|\eta_k(\omega)|^2\le\frac{m^2K^2}{A_n^2\mathbf{a}\Sigma\mathbf{a}^*}\xi_{\bar p(k),k}^2$. This implies that

$$\begin{aligned}\eta_k^2I(|\eta_k|>\varepsilon)&\le\frac{m^2K^2}{A_n^2\mathbf{a}\Sigma\mathbf{a}^*}\xi_{\bar p(k),k}^2I(|\xi_{\bar p(k,\omega),k}(\omega)|>\bar\varepsilon A_n)\\&\le\frac{m^2K^2}{A_n^2\mathbf{a}\Sigma\mathbf{a}^*}\sum_{p=1}^m\xi_{p,k}^2I(|\xi_{p,k}|>\bar\varepsilon A_n).\end{aligned}$$

By taking expection in this inequality we get formula (2.8).

Let us finally observe that if $\xi_k=(\xi_{1,k},\dots,\xi_{m,k})$, $k=1,2,\dots$, is a sequence of independent and identically distributed m-dimensional random vectors with expectation zero and finite covariance matrix Σ, then this sequence of random vectors satisfies relation (14) with the choice $A_n^2=n$. This statement was proved in part b) of problem 41 for those coordinates p for which $E\xi_{p,1}^2>0$. For those coordinates p for which $E\xi_{p,1}^2=0$, $\xi_{p,1}\equiv0$. Hence these coordinates can be omitted.

51.) Let us consider first those numbers p, $1\le p\le m$ for which the p-th element of the diagonal of the (semi-definite) matrix Σ $D_{p,p}$ satisfies the inequality $D_{p,p}>0$. Let us define the triangular array $\eta_{k,j}=\eta_{k,j}(p)=\frac{\xi_{p,j}}{A_kD_{p,p}}$, $1\le j\le k$, $k=1,2,\dots$. Then $\lim\limits_{k\to\infty}E\eta^2_{k,j}=0$, and the triangular array $\eta_{k,j}$, $1\le j\le k$, $k=1,2,\dots$ satisfies the condition of uniform smallness and the central limit theorem. Hence by the result of problem 39 the Lindeberg condition formulated in formula (14) holds for all such indices p.

Since the matrix Σ is positive semi-definite, hence $D_{p,p}\ge 0$ for all $1\le p\le m$. Therefore we have still consider those indices p for which $D_{p,p}=0$. In this case $\lim\limits_{n\to\infty}\frac{1}{A_n^2}\sum\limits_{k=1}^n E\xi^2_{p,k}=0$. Since $E\xi^2_{p,k}\ge E\xi^2_{p,k}I(|\xi_{p,k}|>\varepsilon A_n)$, relation (14) also holds for such indices p.

Appendix

The proof of the inversion formula for Fourier transforms.

Let us introduce the function $\hat f(u)=\int e^{-itu}\tilde f(u)\,du$. To prove formula (6) we have to show that $\hat f(u)=f(u)$ for almost all numbers $u\in R^1$. This statement is equivalent to the identity $\int_0^t f(u)\,du=\int_0^t\hat f(u)\,du$ for all numbers $t\in R^1$. Since

$$\int_0^t\hat f(u)\,du=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_0^t e^{-ius}\tilde f(u)\,ds\,du=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itu}-1}{-itu}\tilde f(u)\,du, \tag{A1}$$

we have to show the identity

$$\int_{-\infty}^{\infty}I_{[0,t]}(u)f(u)\,du=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itu}-1}{-itu}\tilde f(u)\,du, \tag{A2}$$

where $I_{[0,t]}(\cdot)$ is the indicator function of the interval $[0,t]$. The identity (A2) is a special case of an important identity of the Fourier analysis, of the Parseval formula. Let us formulate it.

Parseval formula.

$$\int f(u)\bar g(u)\,du=\frac{1}{2\pi}\int\tilde f(u)\bar{\tilde g}(u)\,du, \tag{A3}$$

where $\tilde f(\cdot)$ denotes the Fourier transform and $\bar f(\cdot)$ the conjugate of a function $f(\cdot)$. Formula (A3) holds if one of the following conditions is satisfied:

a.) Both functions f and g are square integrable.

b.) Both functions $\tilde f$ and $\tilde g$ are square integrable.

If one of the conditions a.) and b.) is satisfied, then also the other condition holds. In this case the identity $\int|f(u)|^2\,du=\frac{1}{2\pi}\int|\tilde f(u)|^2,du$ holds (because of the Parseval formula). The transformation $\mathbf{T}f=f\to\frac{1}{\sqrt{2\pi}}\tilde f$ is an automorphism in the space of square integrable function. This statement means not only the validity of the identity

$\int \mathbf{T}f(u)\overline{\mathbf{T}g}(u)\,du = \int f(u)\bar{g}(u)\,du$. *It also states that all square integrable functions f can be represented in the form $f = \mathbf{T}h$ with a square integrable function h.)*

Let us finally remark that in a complete formulation of the Parseval formula the notion of the Fourier transform has to be defined for all square integrable but not necessarily integrable functions $f(\cdot)$. This definition can be given by means of the above mentioned L_2 isomorphism. For all square integrable functions $f(\cdot)$ there exists a sequence of integrable and square integrable functions $f_n(\cdot)$ which converges to the function $f(\cdot)$ in the L_2 norm of square integrable functions, i.e. $\int |f_n(u) - f(u)|^2\,du \to 0$ if $n \to \infty$. Then the Fourier transform $\tilde{f}(\cdot)$ of the function $f(\cdot)$ is the limit of the functions $\tilde{f}_n(\cdot)$ in the L_2 norm. This limit always exists, and it does not depend on the choice of the sequence of functions $f_n(\cdot)$ converging to the function $f(\cdot)$.

In the Parseval formula formulated in this text a norming factor $\frac{1}{2\pi}$ is present which does not appear in its formulation in text books. The reason of this difference is that we have chosen a different normalization in the definition of the Fourier transform. (We have omitted the factor $\frac{1}{\sqrt{2\pi}}$ from the definition.)

If the Fourier transform of an integrable function f is integrable, then it is also square integrable, since it is a bounded function. Further, the Fourier transform of the function $g(u) = I_{[0,t]}(u)$ is the square integrable function $\tilde{g}(v) = \int_0^t e^{iuv}\,du = \frac{e^{iv}-1}{iv}$. Hence the formula (A2) (and therefore also formula (6)) is a consequence of the Parseval formula with the choice of the above functions $f(\cdot)$ and $g(\cdot)$.

We can prove that a finite measure μ with an integrable Fourier transform $\tilde{f}(u) = \int e^{itu}\mu(\,dt)$ has a density function $f(\cdot)$ defined by formula (6) with the help of the following smoothing argument. Let us consider for all numbers $\varepsilon > 0$ the Gaussian measure ν_ε with expectation zero and variance ε. This measure has density function $\varphi_\varepsilon(u) = \frac{1}{\sqrt{2\pi\varepsilon}} e^{-u^2/2\varepsilon}$ and Fourier transform $e^{-\varepsilon u^2/2}$. Let us introduce the convolution $\mu_\varepsilon = \mu * \nu_\varepsilon$, i.e. $\mu_\varepsilon(A) = \mu * \nu_\varepsilon(A) = \int \mu(A-u)\varphi_\varepsilon(u)\,du$.

The measure μ_ε has a density function $f_\varepsilon(u) = \int \varphi_\varepsilon(u-v)\mu(\,dv)$, and the Fourier transform of this measure is the integrable function $\tilde{f}_\varepsilon(u) = e^{-\varepsilon u^2/2}\tilde{f}(u)$. Hence the function $f_\varepsilon(u)$ can be expressed as the inverse Fourier transform of the function $\tilde{f}_\varepsilon(u)$ defined in formula (6). If $\varepsilon \to 0$, then $f_\varepsilon(u) \to f(u)$, where $f(u)$ is the function defined in formula (6), and this convergence is uniform in the variable u. On the other hand, the measure μ is the weak limit of the measures μ_ε if $\varepsilon \to 0$, that is the probability measures $\frac{\mu_\varepsilon}{\mu(R^1)}$ converge weakly to the probability measure $\frac{\mu}{\mu(R^1)}$. (Let us remark that $\mu(R^1) = \mu_\varepsilon(R^1)$.) Hence we get by taking limit $\varepsilon \to 0$ that $\mu((a,b]) = \int_a^b f(u)\,du$ if $\mu(\{a\}) = \mu(\{b\}) = 0$, i.e. if the points a and b are points of continuity of the measure μ. This implies that the function f is the density function of the measure μ.

By some slight modification of the above argument we can prove the above statement also in the case if μ is a signed measure with bounded variation and integrable Fourier transform. Let us remark that by refining the argument of the above limit procedure and exploiting the L_2 isomorphism property of the Fourier transform, the above result about the density function of a measure μ and its Fourier transform can be strengthened. It is enough to assume that the Fourier transform of the (signed) measure μ is square integrable. But in this case the inverse Fourier transform defined in formula (6) has to be defined by means of the L_2 extension of the original integral, and we cannot claim that the density function of the measure μ is continuous.

The proof of the Parseval formula. Let us first prove the Parseval formula for such simple pairs of functions (f,g) which disappear outside of a finite interval $[-A,A]$, and which are sufficiently smooth, say they have two continuous derivatives. Then by considering the restriction of these functions to some interval $[-\pi T,\pi T]\supset[-A,A]$, $T\pi\ge A$, and the discrete version of the Parseval formula we can write that

$$\int f(u)g(u)\,du = 2\pi T\sum_{k=-\infty}^{\infty} a_k(T)\bar b_k(T),$$

where $a_k(T)=\frac{1}{2\pi T}\int e^{iku/T}f(u)\,du=\frac{1}{2\pi T}\tilde f\left(\frac{k}{T}\right)$, and $b_k(T)=\frac{1}{2\pi T}\tilde g\left(\frac{k}{T}\right)$. But the above expression $2\pi T\sum\limits_{k=-\infty}^{\infty} a_k(T)\bar b_k(T)$ is an approximating sum of the integral $\int \tilde f(u)\bar{\tilde g}(u)\,du$, and the Fourier transforms $\tilde f(u)$ and $\tilde g(u)$ tend to zero fast as $|u|\to\infty$ because of the smoothness of the functions f and g. (See for instance the result of problem 28.) Hence the limit procedure $T\to\infty$ yields formula (A3) in this special case.

The Parseval formula yields with the choice $f=g$ the identity $\int|f(u)|^2\,du=\frac{1}{2\pi}\int|\tilde f(u)|^2\,du$, further the functions f for which we have proved these identity are everywhere dense in the space of square integrable functions. Hence we get proof of the Parseval formula by extending the isometry $\mathbf{T}$: $f\to\mathbf{T}f=\frac{1}{\sqrt{2\pi}}\tilde f$ in the L_2 norm to the space of all square integrable functions. To complete the proof we still have to show that this extension of the transformation $\mathbf{T}$ maps to the whole space of square integrable functions.

To prove this missing part let us consider those functions f which are sufficiently smooth (say they are 10-times differentiable) and to zero sufficiently fast in plus minus infinity (say $|f(u)|\le\text{const.}\left(1+|u|^{100}\right)$). Since such functions constitute an everywhere dense set in the space of square integrable functions it is enough to show that they are in the image space of the operator $\mathbf{T}$. We will show that the identity $\mathbf{T}\sqrt{2\pi}\tilde{f^-}=f$, where $f^-(u)=f(-u)$, follows from the already proved statements.

Also the function $\tilde f$ is smooth, and it tends to zero fast. (This also follows from the statements of problems 27 and 28. Actually the statement of problem 27 deals only with the Fourier transform of probability measures, but it is not difficult to see that this statement also holds for the Fourier transform of all signed measures with bounded variance. We want to exploit that the measures $\mu^\pm$, $\mu^\pm(A)=\int_A f^\pm(u)\,du$, $f^+(u)=\max(f(u),0)$, $f^-(u)=-\min(f(u),0)$ have at least 8 moments.) Since both the functions $f(u)$ and the indicator function $I_{[0,t]}(u)$ of the interval $[0,t]$ are square integrable, formula (A2) holds by the already proved part of the Parseval formula. Since the faction $\hat f$ defined with the help of the function $\tilde f(u)$ at the start of this Appendix is integrable, also formula (A1) holds. Formulas (A1) and (A2) together imply that the pair of functions $(f,\tilde f)$ satisfy relation (6). But this relation is equivalent to the statement that the function f is the Fourier transform of the function $2\pi\tilde{f^-}$, where $f^-(u)=f(-u)$, and this is what we wanted to prove.

The proof of Weierstrass second approximation theorem.

The functions $\frac{1}{(2\pi)^{k/2}}e^{i(j_1t_1+\cdots+j_kt_k)}$ constitute a complete orthonormal system in the space of square integrable functions which are periodic by 2π in all their arguments. This important result of the theory of Fourier series implies that all sufficiently smooth and in all their arguments periodic functions are the uniform limits of their Fourier series. Indeed, in this case the Fourier coefficients tend to zero fast, and this implies the uniform convergence. Since such functions are everywhere dense in the supremum norm in the space of continuous functions, this statement implies Weierstrass second approximation theorem. Nevertheless, instead of this argument we present a direct proof of Weierstrass second approximation theorem which does not apply the completeness of the trigonometrical functions in the L_2 space. We shall prove Fejér's theorem, more precisely its multi-dimensional version. Weierstrass second approximation theorem is a direct consequence of this result.

Fejér's theorem. *Let $f(x_1,\dots,x_k)$ be a continuous function of k arguments which is periodic by 2π in all of its arguments. Let us define for all k-dimensional vectors $(n_1,\dots,n_k)$ with non-negative integers the trigonometrical sum*

$$s_{n_1,\dots,n_k}(f)(t_1,\dots,t_k)=\sum_{j_1=-n_1}^{n_1}\cdots\sum_{j_k=-n_k}^{n_k}A_{j_1,\dots,j_k}e^{i(j_1t_1+\cdots+j_kt_k)},$$

where

$$A_{j_1,\dots,j_k}=\frac{1}{(2\pi)^k}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}e^{-i(j_1u_1+\cdots+j_ku_k)}f(u_1,\dots,u_k)\,du_1\dots du_k.$$

Let us also consider the following Cesaro means $A_n(f)$, $n=1,2,\dots,$ of the above trigonometrical sums:

$$A_n(f)(t_1,\dots,t_k)=\frac{1}{(n+1)^k}\sum_{\substack{0\le n_j\le n\\ \text{for all indices } 1\le j\le k}}s_{n_1,\dots,n_k}(f)(t_1,\dots,t_k).$$

Then $\lim\limits_{n\to\infty}A_n(f)(t_1,\dots,t_k)=f(t_1,\dots,t_k)$, and the above convergence is uniform in all of its arguments $t_1,\dots,t_k$.

The proof of Fejér's theorem. The proof of Fejér's theorem is based on the following formula:

$$A_n(f)(t_1,\dots,t_k)=\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}f(u_1,\dots,u_k)\bar K_n(t_1-u_1,\dots,t_k-u_k)\,du_1\dots du_k, \tag{A4}$$

where
$$\bar{K}_n(u_1,\dots,u_k)=K_n(u_1)\cdots K_n(u_k), \tag{A5}$$
and
$$K_n(u)=\frac{1}{2\pi(n+1)}\sum_{k=-n}^{n}(n+1-|k|)e^{iuk}=\frac{\sin^2\left(\frac{n+1}{2}u\right)}{2\pi(n+1)\sin^2\left(\frac{u}{2}\right)}. \tag{A5$'$}$$

These relations hold, since by writing into the definitions of the expressions $A_n(f)$ and $s_{n_1,\dots,n_k}(f)$ the definition of the Fourier coefficient $A_{j_1,\dots,j_k}$ we get relation (A4) together with formula (A5), where

$$\begin{aligned}\bar{K}_n(u_1,\dots,u_k)&=\frac{1}{(2\pi(n+1))^k}\sum_{\substack{0\le n_j\le n\\ \text{for all indices } 1\le j\le k}}\;\sum_{\substack{|m_j|\le n_j\\ \text{for all indices } 1\le j\le k}}e^{i(m_1u_1+\cdots+m_ku_k)}\\ &=\frac{1}{(2\pi(n+1))^k}\prod_{j=1}^{k}\left(\sum_{n_j=0}^{n}\sum_{m_j=-n_j}^{n_j}e^{im_ju_j}\right)=K_n(u_1)\dots K_n(u_k),\end{aligned}$$

and the function $K_n(u)$ is defined in the middle term of formula (A5$'$). This sum can be written in a closed form with the help of the following calculation.

$$\begin{aligned}\frac{1}{2\pi(n+1)}\sum_{k=-n}^{n}(n+1-|k|)e^{iuk}&=\frac{1}{2\pi(n+1)}\left(\sum_{k=0}^{n}e^{iuk}\right)\left(\sum_{k=0}^{n}e^{-iuk}\right)\\ &=\frac{1}{2\pi(n+1)}\left|\frac{e^{i(n+1)u}-1}{e^{iu}-1}\right|^2=\frac{1}{2\pi(n+1)}\frac{\left|e^{i(n+1)u/2}-e^{-i(n+1)u/2}\right|^2}{\left|e^{iu/2}-e^{-iu/2}\right|^2}\\ &=\frac{\sin^2\left(\frac{n+1}{2}u\right)}{2\pi(n+1)\sin^2\left(\frac{u}{2}\right)}.\end{aligned}$$

The function $K_n(u)$ defined in formula (A5$'$) have the following properties important for us:

(i) $\int_{-\pi}^{\pi}K_n(u)\,du=1$. This statement follows from the representation of the functions $K_n(\cdot)$ in the form of a sum.

(ii) $K_n(u)\ge 0$ for all numbers $u\in R^1$.

(iii) $\lim\limits_{n\to\infty}\sup\limits_{\varepsilon\le|u|\le\pi}K_n(u)=0$ for all numbers $\varepsilon>0$.

Statements (ii) and (iii) follow from the representation of the functions $K_n(\cdot)$ given in a closed form.

Since a function continuous on a compact set is uniformly continuous, there exists a constant $\delta=\delta(\varepsilon,f)$ for all $\varepsilon>0$ such that the continuous and in its coordinates periodic function f satisfies the inequality $|f(x_1,\dots,x_k)-f(y_1,\dots,y_k)|<\varepsilon$ if $|x_j-y_j|<\delta$ for all indices $j=1,\dots,k$. (In this relation we identify the points $x_j+2\pi l$, $l=0\pm1,\pm2,\dots$, and the inequality $|x_j-y_j|<\delta$ means that $|x_j-y_j+2\pi l_l|<\delta$ with an appropriate integer l_j.) Let us introduce the notation $\mathbf{B}(\delta,(t_1,\dots,t_k))=\{(u_1,\dots,u_k)\colon |u_j-t_j|<\delta,\ -\pi\le u_j<\pi,\ j=1,\dots,k\}$. Because of property (i)

$$\begin{aligned}
&A_n(f)(t_1,\dots,t_k)-f(t_1,\dots,t_k)\\
&\qquad=\int_{[-\pi,\pi)^k}(f(u_1,\dots,u_k)-f(t_1,\dots,t_k))\,K_n(t_1-u_1)\cdots K_n(t_k-u_k)\,du_1\dots du_k\\
&\qquad=\int_{\mathbf{B}(\delta,(t_1,\dots,t_k))}[\cdots]\,du_1\dots du_k+\int_{[-\pi,\pi)^k\setminus\mathbf{B}(\delta,(t_1,\dots,t_k))}[\cdots]\,du_1\dots du_k\\
&\qquad=I_{1,n}(t_1,\dots,t_k)+I_{2,n}(t_1,\dots,t_k).
\end{aligned}$$

It follows from the definition of the set $B(\delta,(t_1,\dots,u_t))$, the number δ and properties (i) and (ii) that

$$|I_{1,n}(t_1,\dots,t_k)|\le\varepsilon\int_{[-\pi,\pi)^k}K_n(t_1-u_1)\cdots K_n(t_k-u_k)\,du_1\dots du_k\le\varepsilon$$

for all indices $n=1,2,\dots$ and points $(t_1,\dots,t_k)$. On the other hand, by applying the notation $\sup\limits_{(u_1,\dots,u_k)}|f(u_1,\dots,u_k)|=L$ and carrying out the substitutions $t_j-u_j=\bar u_j$ we can show with the help of relations (i), (ii) and (iii) that

$$|I_{2,n}(t_1,\dots,t_k)|\le 2L\int_{[-\pi,\pi)^k\setminus\mathbf{B}(\delta,(0,\dots,0))}K_n(\bar u_1)\cdots K_n(\bar u_k)\,d\bar u_1\dots d\bar u_k\to 0,$$

if $n\to\infty$, since

$$\begin{aligned}
\int\limits_{\substack{\delta<|\bar u_j|<\pi\\ -\pi\le\bar u_l<\pi,\,l\ne j,\,1\le l\le k}}K_n(\bar u_1)\cdots K_n(\bar u_k)\,d\bar u_1\dots d\bar u_k&=\int_{\delta<|u|<\pi}K_n(u)\,du\\
&\le 2\pi\sup_{\delta<|u|<\pi}K_n(u)\to 0,\quad \text{ha } n\to\infty.
\end{aligned}$$

for all numbers $1\le j\le k$. Since the above estimates hold for all constants $\varepsilon>0$ (together with an appropriate number $\delta=\delta(\varepsilon,f)$), they imply Fejér's theorem.