ARKIV FÖR MATEMATIK Band 5 nr 28

1.64 12 3 Communicated 14 October 1964 by HARALD CRAMÉR and LENNART CARLESON

Limit theorems for sampling from finite populations

By BENGT ROSÉN

Introduction and summary

Let $\pi=\{a_1, a_2,\dots, a_N\}$ be a finite population, i.e. a finite set of real numbers. $X_1, X_2, \dots, X_n$ is a sample from π drawn with or without replacement. In the first case we speak of independent observations from π and in the latter simply of a random sample from π. A major problem in sampling theory can be formulated thus: Compute, in terms of π, the probability distribution of the random variable $\varphi(X_1, X_2, \dots, X_n)$, where φ is a given function of n variables. It is seldom that a neat closed form solution of the above problem can be found. However, one can often give an approximate solution by considering a limit procedure.

For independent observations a natural limit procedure is obtained by letting the number of observations tend to infinity. There is a wealth of limit results for different functions φ.

For samples drawn without replacement the above limit procedure looses meaning as the population will be exhausted after a finite number of drawings. A fruitful limit procedure, treated by many authors, is obtained by considering a double sequence of random variables:

$$\begin{array}{llll} X_{11}, & X_{12}, & \dots, X_{1n_1} & \text{is a random sample from } \pi_1 \\ \vdots & \vdots & \vdots & \qquad\qquad\qquad\qquad\quad\ \vdots \\ X_{k1}, & X_{k2}, & \dots, X_{kn_k} & \text{is a random sample from } \pi_k \\ \vdots & \vdots & \vdots & \qquad\qquad\qquad\qquad\quad\ \vdots \,. \end{array}$$

One can now consider the limiting behavior of various sample functions $\varphi_k(X_{k1}, \dots, X_{kn_k})$, when $k\to\infty$. The present paper is devoted to a study of limit problems associated with the above scheme. The problems and the results are in many respects analogous to those in the case of independent observations.

In what follows we give a brief summary of the contents of the paper and we describe the problems considered by referring to analogue problems for independent observations.

Chapter 1 contains definitions and some basic elementary results concerning sampling from finite populations.

Chapter 2 treats strong and weak laws of large numbers. There is given an analogue of the law of the iterated logarithm.

Chapter 3 is expository and in it we have collected results about stochastic processes, that will be used in the later chapters.

In Chapter 4 we consider central limit problems. Especially we give an analogue of the so-called invariance principle (cf. Donsker [3]).

Chapter 5 is concerned with the behavior of empirical distributions and empirical fractiles.

Chapter 1. Definitions and fundamentals about sampling from finite populations

1. *Definitions and some elementary results*

By a *d-dimensional finite population* π we mean a finite set of d-tuples of real numbers.

$$\pi = \{a_1, a_2, \dots, a_N\}, \quad a_\nu = (a_\nu^{(1)}, \dots, a_\nu^{(d)}),$$

where $a_\nu^{(i)}$ are real numbers

$$i = 1, 2, \dots, d, \quad \nu = 1, 2, \dots, N.$$

We will often omit the word finite and say simply a population π. The letter N will throughout be used to denote *population size*. The *population mean vector* μ_π and *population covariance matrix* $[\sigma_{ij}]$ $i, j, = 1, 2, \dots, d$, are defined as follows

$$\mu_\pi = \frac{1}{N}\sum_{\nu=1}^{N} a_\nu \quad \text{and} \quad [\sigma_{ij}] = \frac{1}{N-1}\sum_{\nu=1}^{N} (a_\nu - \mu_\pi)'(a_\nu - \mu_\pi)$$

($'$stands for matrix transposition). For 1-dimensional populations we write σ_π^2 instead of σ_{11} for the *population variance*, and we also define the quantity

$$D_\pi^2 = \sum_{\nu=1}^{N} (a_\nu - \mu_\pi)^2 = (N-1)\sigma_\pi^2. \tag{1.1}$$

The *population distribution* is the probability measure on R^d (d-dim. Euclidean space) obtained by giving each element in π the mass N^{-1}. We denote the corresponding distribution function, which we assume to be right continuous, by $F_\pi(x)$. The *centered distribution function* $F_\pi^c(x)$ is defined as $F_\pi^c(x) = F_\pi(x - \mu_\pi)$.

Next we define a random sample $X_1, X_2, \dots, X_n$ from π. Let Ω_N be the set of all permutations $\omega = (i_1, i_2, \dots, i_N)$ of the numbers $1, 2, \dots, N$. The probability P on Ω_N is defined by $P(\omega) = (N!)^{-1}$ for all $\omega \in \Omega_N$. Let T_π be the mapping $T_\pi : \Omega_N \to \times_{\nu=1}^{N} R_\nu^d$ ($\times$ stands for Cartesian product) given by:

$$\text{for } \omega = (i_1, i_2, \dots, i_N) \text{ is } T_\pi(\omega) = (a_{i_1}, a_{i_2}, \dots, a_{i_N}).$$

This vector valued function T_π on Ω_N we call a *random permutation* (r.p.) *of the elements in* π and we denote its components by $T_\pi = (X_1, X_2, \dots, X_N)$. By a *random sample of size* n, ($n \leqslant N$), from π we mean the random vector $(X_1, X_2, \dots, X_n)$. In the sequel we will denote the sample space simply by Ω instead of Ω_N. For future reference we list some simple properties of random samples.

Exchangeability: $X_1, X_2, \ldots, X_N$ are exchangeable, i.e. their joint distribution is invariant under permutation of the X's.

Duality principle: We define $S_0 = 0$ and $S_\nu = X_1 + X_2 + \ldots + X_\nu$, $\nu = 1, 2, \ldots, N$. Let $i_1, i_2, \ldots, i_n$ be chosen among $1, 2, \ldots, N$. If $\mu_\pi = 0$, then the two random vectors $(S_{i_1}, S_{i_2}, \ldots, S_{i_n})$ and $(-S_{N-i_1}, -S_{N-i_2}, \ldots, -S_{N-i_n})$ have identical distributions.

$X_1, X_2, \ldots, X_N$ are dependent and it will be essential for us to be able to handle the dependence. When dealing with conditioning concepts, we will follow Loève's notations, [14] Chapter VII. The following result is intuitively obvious and easily proved.

Conditioning principle: The mixed conditional distribution of $X_{n+1}, X_{n+2}, \ldots, X_{n+m}$, $n + m \leqslant N$, given $X_1, X_2, \ldots, X_n$, is at the point ω identical with the distribution of a random sample of size m from the population $\pi'(\omega) = \pi$ with the elements $X_1(\omega), \ldots, X_n(\omega)$ removed.

We will use standard probability notations, particularly EX and $\sigma^2(X)$ for respectively expectation and variance of the random variable X. We will also use standard abbreviations as r.v. = random variable, pr. = probability, d.f. = distribution function, c.f. = characteristic function, i.d. = in distribution. Some additional abbreviations are introduced in connection with definitions. Concerning the notations for well-known distributions such as normal, binomial, etc. we follow Wilks [20]. Finally, we will use $[a]$ to denote the integral part of a, and A' for the complement of the set A.

The following result is well-known, see e.g. Wilks [20] p. 222.

Theorem 1.1. *$X_1, X_2, \ldots, X_N$ is a r.p. of the elements in π, which has covariance matrix $[\sigma_{ij}]$. $S_n = X_1 + \ldots + X_n$, $n = 1, 2, \ldots, N$. Then*

$$E(S_n - n\mu_\pi)'(S_n - n\mu_\pi) = n(1 - n/N)[\sigma_{ij}]. \tag{1.2}$$

Particularly, when π is 1-dimensional,

$$\sigma^2(S_n) = n(1 - n/N)\,\sigma_\pi^2 \tag{1.3}$$

We shall mainly be concerned with 1-dimensional populations, and from now on π stands for a 1-dimensional population, unless otherwise stated.

2. *Limit procedures for sampling from finite populations*

As stated in the introduction, we shall consider a limit procedure based on a sequence $\{\pi_k\}_1^\infty$ of populations. We assume once and for all that a population sequence has the property that $N_{\pi_k} \to \infty$ when $k \to \infty$. When we are dealing with a sequence $\{\pi_k\}_1^\infty$, we will usually write $\mu_k, \sigma_k^2, N_k, F_k$ etc., instead of μ_{π_k}, $\sigma_{\pi_k}^2$, N_{π_k}, F_{π_k} etc.

A sequence $\{\pi_k\}_1^\infty$ is said to be *degenerate* if it contains a subsequence $\{\pi_{k_\nu}\}_{\nu=1}^\infty$ for which $F_{k_\nu}^c(x)$ converges i.d. to the distribution with pr. 1 in $x = 0$.

Let $\{\pi_k\}_1^\infty$ be given. Then $X_{k1}, X_{k2}, \ldots, X_{kN_k}$ will throughout the paper stand for a r.p. of the elements in π_k. Let $\{\varphi_{kn}\}$, $n=1,2,\dots,N_k$, $k=1,2,\dots$ be a double sequence of r.v.'s, where φ_{kn} is a function of $X_{k1},\dots,X_{kn}$, i.e. $\varphi_{kn}=\varphi_{kn}(X_{k1},X_{k2},\dots,X_{kn})$. We say that $\{\varphi_{kn}\}$ *converges in pr.* to the constant c for the *sample size sequence* (sss.) $\{n_k\}_1^\infty$ if for every $\varepsilon>0$ it holds that $\lim_k P(|\varphi_{kn_k}-c|\geqslant\varepsilon)=0$, and we say that $\{\varphi_{kn}\}$ *converges strongly in pr.* to c for the sss. $\{n_k\}_1^\infty$ if for every $\varepsilon>0$ it holds that $\lim_k P(\max_{n_k\leqslant n\leqslant N_k}|\varphi_{kn}-c|\geqslant\varepsilon)=0$. Obviously, strong convergence in pr. for the sss. $\{n_k\}_1^\infty$ implies convergence in pr. for the same sss.

3. *Relations between sampling with and without replacement*

We shall here consider two types of relations between sampling with and without replacement, which can be roughly described as follows.

1. If the proportion sampled from π is small, then sampling with and without replacement are almost equivalent.

2. When a sample is drawn without replacement there is a tendency that the sample sum lies closer to its mean than it does when the sample is drawn with replacement.

We shall give some exact formulations of these heuristic statements.

Lemma 3.1. *$\{\pi_k\}_1^\infty$ is a population sequence for which $F_k\to F$ i.d. when $k\to\infty$ and $X_{k1},\dots,X_{kN_k}$ is a r.p. of the elements in π_k. Then it holds for every fixed m that the distribution of $X_{k1},\dots,X_{km}$ converges i.d. when $k\to\infty$ to the distribution of m independent r.v.'s each having d.f. F.*

Proof. Let $x_1\leqslant x_2\leqslant\dots\leqslant x_m$. Then

$$P(X_{k1}\leqslant x_1,\dots,X_{km}\leqslant x_m)$$

$$=\frac{N_kF_k(x_1)}{N_k}\cdot\frac{N_kF_k(x_2)-1}{N_k-1}\cdot\dots\cdot\frac{N_kF_k(x_m)-(m-1)}{N_k-(m-1)}\to F(x_1)F(x_2)\dots F(x_m)\ \text{i.d.}\qquad(3.1)$$

as according to our general assumption $\lim_k N_k=\infty$. Because of the exchangeability of the X's (3.1) holds without the above restriction on the x_ν's, and the lemma is proved.

The following theorem due to Hoeffding [12] is a precise formulation of 2.

Theorem 3.1. *$X_1,X_2,\dots,X_n$ and $X_1',X_2',\dots,X_n'$ are samples from π, drawn respectively without and with replacement, $S_n=X_1+\dots+X_n$ and $S_n'=X_1'+\dots+X_n'$. If ψ is continuous and convex it holds that $E\psi(S_n)\leqslant E\psi(S_n')$.*

As pointed out by Hoeffding, the theorem can be used to get inequalities for probabilities concerning sums of samples from finite populations from inequalities concerning sums of independent r.v.'s.

Lemma 3.2. *Let $\psi(x)$ be positive, nondecreasing and convex for $x>0$. S_n and S_n' are defined in Theorem 3.1. Then it holds for $\alpha>0$ that*

$$P(|S_n - ES_n| \geq \alpha) \leq [\psi(\alpha)]^{-1} E\,\psi(|S_n' - ES_n'|).$$

The lemma is an immediate consequence of Tchebychev's inequality and Theorem 3.1.

Chapter 2. On laws of large numbers

4. *Inequalities concerning sample sums*

Let $X_1, X_2, \ldots, X_N$ be a r.p. of the elements in the population π, $S_0 = 0$ and $S_n = X_1 + \ldots + X_n$, $n = 1, 2, \ldots, N$. We denote sample mean S_n/n by $\bar{X}_n$. Our aim is to prove analogues of the strong and weak law of large numbers. We start by deriving two estimates of $P(\max_{\nu \geq n} |b_\nu S_\nu| \geq \varepsilon)$ where $b_0, b_1, \ldots, b_N$ are real numbers. The method used to obtain these estimates is similar to that used by Hájek and Rényi in [10].

Lemma 4.1. *π is a population with mean 0 and variance σ^2, and $\{b_\nu\}_0^N$ a sequence of real numbers. Then it holds for every $\varepsilon > 0$ and for $n = 1, 2, \ldots, N$*

$$P(\max_{\nu \geq n} |b_\nu S_\nu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}\left[\left(1 - \frac{1}{N}\right)\sum_{\nu=n}^{N-1} b_\nu^2 + \sum_{\nu=n+1}^{N-1} \nu\left(1 - \frac{\nu}{N}\right)\left(b_{\nu-1}^2\left(1 - \frac{2}{\nu}\right) - b_\nu^2\right)^+\right]$$

where $a^+ = \max\ (0, a)$.

Proof. We consider instead $P(\max_{\nu \leq n} |c_\nu S_\nu| \geq \varepsilon)$ for $c_\nu = b_{N-\nu}$. Let $\varepsilon > 0$ be given. We define the following r.v.'s. For $\nu = 0, 1, 2, \ldots, N$, I_ν is the indicator function of the event $(\max_{t \leq \nu} c_t^2 S_t^2 < \varepsilon^2)$ and $H_\nu = 1 - I_\nu$. L is the first (if any) index ν for which $c_\nu^2 S_\nu^2 \geq \varepsilon^2$. Then

$$\varepsilon^2 H_n \leq \sum_{\nu=0}^{n-1} I_\nu (c_{\nu+1}^2 S_{\nu+1}^2 - c_\nu^2 S_\nu^2). \tag{4.1}$$

To prove (4.1) we first assume that $\max_{\nu \leq n} c_\nu^2 S_\nu^2 < \varepsilon^2$. Then the left-hand side of (4.1) is 0, while the right-hand side is $c_n^2 S_n^2 \geq 0$, and (4.1) holds. If $\max_{\nu \leq n} c_\nu^2 S_\nu^2 \geq \varepsilon^2$, the left-hand side is ε^2 and the right-hand side equals $c_L^2 S_L^2 \geq \varepsilon^2$ and (4.1) is proved. By integrating (4.1) over Ω we obtain

$$\varepsilon^2 P(\max_{\nu \leq n} c_\nu^2 S_\nu^2 \geq \varepsilon^2) = \varepsilon^2 EH_n \leq \sum_{\nu=0}^{n-1} \int_{(I\nu=1)} (c_{\nu+1}^2 S_{\nu+1}^2 - c_\nu^2 S_\nu^2)\, P(d\omega). \tag{4.2}$$

Let for $\nu = 0, 1, 2, \ldots, N$ $\mathcal{B}_\nu$ be the algebra of subsets of Ω induced by $S_0, S_1, \ldots, S_\nu$. From the conditioning principle (p. 385) it readily follows that $E^{\mathcal{B}_\nu} X_{\nu+1} = -S_\nu/(N-\nu)$. Using this result, and some well-known computational rules for conditioning (see e.g. Loève [14]) we get

$$\begin{aligned} E^{\mathcal{B}_\nu} S_{\nu+1}^2 &= E^{\mathcal{B}_\nu} S_\nu^2 + 2\,E^{\mathcal{B}_\nu} S_\nu X_{\nu+1} + E^{\mathcal{B}_\nu} X_{\nu+1}^2 \\ &= S_\nu^2 + 2\,S_\nu(-S_\nu/(N-\nu)) + E^{\mathcal{B}_\nu} X_{\nu+1}^2 = S_\nu^2(1 - 2/(N-\nu)) + E^{\mathcal{B}_\nu} X_{\nu+1}^2. \end{aligned} \tag{4.3}$$

From (4.2), (4.3) and the obvious inclusion $(I_\nu = 1) \in \mathcal{B}_\nu$ we get

$$\varepsilon^2 P(\max_{\nu\leqslant n} c_\nu^2 S_\nu^2 \geqslant \varepsilon^2) \leqslant \sum_{\nu=0}^{n-1} \int_{(I_\nu=1)} E^{\mathcal{B}_\nu}(c_{\nu+1}^2 S_{\nu+1}^2 - c_\nu^2 S_\nu^2)\, P(d\omega)$$

$$= \sum_{\nu=0}^{n-1} \int_{(I_\nu=1)} S_\nu^2 \left(c_{\nu+1}^2 \left(1 - \frac{2}{N-\nu}\right) - c_\nu^2\right) P(d\omega) + \sum_{\nu=0}^{n-1} c_{\nu+1}^2 \int_{(I_\nu=1)} E^{\mathcal{B}_\nu} X_{\nu+1}^2.$$

The above inequality is only unsharpened if we first change

$$\varrho_\nu = (c_{\nu+1}^2(1 - 2/(N-\nu)) - c_\nu^2) \text{ into } \varrho_\nu^+$$

and then enlarge the domain of integration to Ω in the integrals in both sums. This observation together with Lemma 1.1 yields

$$\varepsilon^2 P(\max_{\nu\leqslant n} c_\nu^2 S_\nu^2 \geqslant \varepsilon^2) \leqslant \sum_{\nu=0}^{n-1} ES_\nu^2 \left(c_{\nu+1}^2 \left(1 - \frac{2}{N-\nu}\right) - c_\nu^2\right)^+ + \sum_{\nu=0}^{n-1} c_{\nu+1}^2 EX_{\nu+1}^2$$

$$= \sigma^2 \left[\left(1 - \frac{1}{N}\right) \sum_{\nu=1}^{n} c_\nu^2 + \sum_{\nu=1}^{n-1} \nu \left(1 - \frac{\nu}{N}\right) \left(c_{\nu+1}^2 \left(1 - \frac{2}{N-\nu}\right) - c_\nu^2\right)^+\right]. \tag{4.4}$$

The inequality in Lemma 2.1 now follows from (4.4) and the formula

$$P(\max_{\nu\geqslant n} |b_\nu S_\nu| \geqslant \varepsilon) = P(\max_{\nu\leqslant N-n} |b_{N-\nu} S_\nu| \geqslant \varepsilon), \tag{4.5}$$

which is an immediate consequence of the duality principle. Thus the lemma is completely proved. The following result is a special case.

Lemma 4.2. *π is a population with mean 0 and variance σ^2. Then it holds for every $\varepsilon > 0$ and for $n = 1, 2, \ldots, N$.*

$$P(\max_{\nu\geqslant n} |\overline{X}_\nu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}\left(1 - \frac{1}{N}\right) \sum_{\nu=n}^{N-1} \frac{1}{\nu^2} = \frac{\sigma^2}{\varepsilon^2}\left(\frac{1}{n} - \frac{1}{N}\right) C(n, N)$$

where $C(n, N) \leqslant 1 + n^{-1}$.

Proof. The inequality follows readily from Lemma 4.1 by putting $b_\nu = \nu^{-1}$. It is easily checked that $(b_{\nu-1}^2(1 - 2\nu^{-1}) - b_\nu^2)^+ = 0$ for $b_\nu = \nu^{-1}$. The upper bound for $C(n, N)$ can be derived thus. When $n = N$ we can put $C(n, N) = 0$. For $n = 1, 2, \ldots, N-1$ we have

$$C(n, N) = n(N-1)(N-n)^{-1} \sum_{\nu=n}^{N-1} \nu^{-2}. \tag{4.6}$$

Now $$\sum_{\nu=n}^{N-1} \nu^{-2} \leqslant 1/n^2 + \sum_{\nu=n}^{N-2} 1/\nu(\nu+1) = 1/n^2 + 1/n - 1/(N-1).$$

We insert the last estimate into (4.6) and get $C(n, N) \leqslant (N-n-1)/(N-n) + (N-1)/n(N-n) \leqslant 1+n^{-1}$ and Lemma 4.2 is proved.

Next we give an estimate of $P(\max_{\nu \geqslant n} |b_\nu S_\nu| \geqslant \varepsilon)$ in terms of $E|S_n|$.

Lemma 4.3. *π is a population with mean 0 and $\{b_\nu\}_0^N$ is a sequence of non-negative numbers such that $\{\nu b_\nu\}_1^{N-1}$ is non-increasing. Then it holds for every $\varepsilon > 0$ and for $n = 1, 2, \ldots, N$*

$$P(\max_{\nu \geqslant n} |b_\nu S_\nu| \geqslant \varepsilon) \leqslant \varepsilon^{-1} b_n E|S_n|. \tag{4.7}$$

Proof. The proof will run almost parallel to that of Lemma 4.1 and again we first prove the dual result. Let $c_\nu = b_{N-\nu}$. Thus $c_\nu(N-\nu)$ is non-decreasing for $\nu = 1, 2, \ldots, N-1$. For fixed $\varepsilon > 0$ we define the following r.v.'s. For $\nu = 0, 1, \ldots, N$, I_ν is the indicator of the event $(\max_{t \leqslant \nu} c_t S_t < \varepsilon)$ and $H_\nu = 1 - I_\nu$. L is the first (if any) index ν for which $c_\nu S_\nu \geqslant \varepsilon$. Then the following inequality, where $a^- = \min(0, a)$, holds

$$\varepsilon H_n \leqslant -c_n S_n^- + \sum_{\nu=0}^{n-1} I_\nu (c_{\nu+1} S_{\nu+1} - c_\nu S_\nu). \tag{4.8}$$

If $\max_{\nu \leqslant n} c_\nu S_\nu < \varepsilon$ the left-hand side in (4.8) is 0, while the right-hand side is $-c_n S_n^- + c_n S_n \geqslant 0$. If $\max_{\nu \leqslant n} c_\nu S_\nu \geqslant \varepsilon$ the left-hand side is ε and the right-hand side $-c_n S_n^- + c_L S_L \geqslant \varepsilon$. Thus (4.8) is proved. Let, as before, $\mathcal{B}_\nu$ be the algebra of events which are determined by conditions on $S_0, S_1, \ldots, S_\nu$. Then $(I_\nu = 1) \in \mathcal{B}_\nu$. By integrating (4.8) over Ω we get

$$\begin{aligned}
\varepsilon P(\max_{\nu \leqslant n} c_\nu S_\nu \geqslant \varepsilon) = \varepsilon E H_n &\leqslant -c_n E S_n^- + \sum_{\nu=0}^{n-1} \int_{(I_\nu=1)} (c_{\nu+1} S_{\nu+1} - c_\nu S_\nu) P(d\omega) \\
&= -c_n E S_n^- + \sum_{\nu=0}^{n-1} \int_{(I_\nu=1)} (c_{\nu+1} E^{\mathcal{B}_\nu} S_{\nu+1} - c_\nu S_\nu) P(d\omega) \\
&= -c_n E S_n^- + \sum_{\nu=0}^{n-1} \left(c_{\nu+1} \frac{N-(\nu+1)}{N-\nu} - c_\nu \right) \int_{(I_\nu=1)} S_\nu P(d\omega).
\end{aligned} \tag{4.9}$$

In the last step we used the formula $E^{\mathcal{B}_\nu} S_{\nu+1} = [(N-(\nu+1))/(N-\nu)] \cdot S_\nu$, which follows from the conditioning principle. From the assumption that $c_\nu(N-\nu)$ is non-decreasing, it follows that $c_{\nu+1}[(N-(\nu+1))/(N-\nu)] - c_\nu \geqslant 0$. If we show that the integrals in the last sum in (4.9) are non-positive we can cancel the sum and end up with the inequality

$$P(\max_{\nu \leqslant n} c_\nu S_\nu \geqslant \varepsilon) \leqslant \varepsilon^{-1} c_n E S_n^- = (2\varepsilon)^{-1} c_n E|S_n|. \tag{4.10}$$

The equality in (4.10) is a consequence of the assumption that π has mean 0.

For fixed $\varepsilon > 0$ we define for $t = 0, 1, 2, \ldots, N$ the events A_t as follows: $A_t = (S_t \geqslant \varepsilon$ and $S_r < \varepsilon$ for $r < t)$. Then it holds that $A_t \in \mathcal{B}_t$ and further

$$(I_\nu = 0) = \bigcup_{t=0}^{\nu} A_t \quad \text{where } A_t \text{ and } A_s \text{ are disjoint for } t \neq s \tag{4.11}$$

Remembering that π is assumed to have mean 0, we get from (4.11)

$$-\int_{(I_\nu=1)} S_\nu P(d\omega) = \int_{(I_\nu=0)} S_\nu P(d\omega) = \sum_{t=0}^{\nu} \int_{A_t} S_\nu P(d\omega) = \sum_{t=0}^{\nu} \int_{A_t} E^{B_t} S_\nu P(d\omega)$$

$$= \sum_{t=0}^{\nu} \frac{N-\nu}{N-t} \int_{A_t} S_t P(d\omega) \geqslant \varepsilon \sum_{t=0}^{\nu} \frac{N-\nu}{N-t} P(A_t) \geqslant 0$$

and (4.10) is completely proved. By considering $-S_0, -S_1, \ldots, -S_N$, we obtain from (4.10)

$$P(\min_{\nu \leqslant n} c_\nu S_\nu \leqslant -\varepsilon) \leqslant (2\varepsilon)^{-1} c_n E\,|S_n|. \tag{4.12}$$

Addition of (4.10) and (4.12) yields

$$P(\max_{\nu \leqslant n} |c_\nu S_\nu| \geqslant \varepsilon) \leqslant \varepsilon^{-1} c_n E\,|S_n|. \tag{4.13}$$

By use of the duality principle we get from (4.13)

$$P(\max_{\nu \geqslant n} |b_\nu S_\nu| \geqslant \varepsilon) = P(\max_{\nu \leqslant n} |b_\nu S_{N-\nu}| \geqslant \varepsilon)$$

$$= P(\max_{\nu \leqslant N-n} |c_{N-\nu} S_\nu| \geqslant \varepsilon) \leqslant \varepsilon^{-1} b_n E\,|S_{N-n}| = \varepsilon^{-1} b_n E\,|S_n|$$

and the proof of Lemma 4.3 is concluded. When applying the lemma we will need an estimate of $E\,|S_n|$. The next lemma gives such an estimate.

Lemma 4.4. *$\lambda(x)$ is an even function such that for $x \geqslant 0$ is $\lambda(x)/x$ positive and non-decreasing, while $\lambda(x)/x^2$ is non-increasing. For a population with mean 0 it then holds that*

$$E\,|S_n| \leqslant \lambda^{-1}(n) \left[\left(\frac{N-n}{N-1} E\lambda(X_1) + (E\lambda(X_1))^2 \right)^{\frac{1}{2}} + E\lambda(X_1) \right],$$

where $\lambda^{-1}(y)$ is the inverse of $\lambda(x)$ for $x \geqslant 0$.

Proof. We fix n and define for $\nu = 1, 2, \ldots, N$, $X'_\nu = X_\nu$ if $|X_\nu| \leqslant \lambda^{-1}(n)$ and $X'_\nu = 0$ otherwise, $Y_\nu = X_\nu - X'_\nu$ and $S'_\nu = X'_1 + \ldots + X'_\nu$. Evidently $X'_1, X'_2, \ldots, X'_N$ is a r.p. of the elements in the population π', which is obtained from π by replacing all elements in π with absolute value $> \lambda^{-1}(n)$ by zeros. From Schwarz's inequality we get

$$E\,|S_n| \leqslant E\,|S'_n| + E\,|\sum_{\nu=1}^{n} Y_\nu| \leqslant \sqrt{ES'^2_n} + nE\,|Y_1|. \tag{4.14}$$

Now

$$ES_n'^2 = \sigma^2(S_n') + (ES_n')^2 = n(1-n/N)\,(1+1/(N-1))\,(EX_1'^2-(EX_1')^2)$$
$$+ n^2(EX_1')^2 = n(N-n)\,(N-1)^{-1}EX_1'^2 + nN(n-1)\,(N-1)^{-1}(EX_1')^2. \quad (4.15)$$

As $\lambda(x)/x^2$ is non-increasing for $x \geqslant 0$ we have

$$EX_1'^2 = \int_{|x|\leqslant\lambda^{-1}(n)} \frac{x^2}{\lambda(x)}\cdot\lambda(x)\,dF_\pi(x) \leqslant \frac{\lambda^{-1}(n)^2}{n}E\lambda(X_1). \quad (4.16)$$

From $\mu_\pi = 0$ and the assumption that $\lambda(x)/x$ is non-decreasing for $x \geqslant 0$ it follows that

$$|EX_1'| = |EY_1| = \left|\int_{|x|>\lambda^{-1}(n)} \frac{x}{\lambda(x)}\cdot\lambda(x)\,dF_\pi(x)\right| \leqslant \frac{\lambda^{-1}(n)}{n}E\lambda(X_1). \quad (4.17)$$

We now obtain the desired inequality by inserting the estimates (4.16) and (4.17) into (4.15) and (4.14). Thus the lemma is proved.

5. *Convergence of sample means*

Thus prepared we shall prove some results about convergence of sample means. We shall consider population sequences $\{\pi_k\}_{k=1}^\infty$. As usual we denote a r.p. of the elements in π_k by $X_{k1}, X_{k2}, \ldots, X_{kN_k}$. Furthermore, we will use the notations $S_n^{(k)} = X_{k1} + \ldots + X_{kn}$, $n = 1, 2, \ldots, N_k$, $S_0^{(k)} = 0$ and $\bar{X}_n^{(k)} = S_n^{(k)}/n$.

Theorem 5.1. *$\{\pi_k\}_1^\infty$ is a sequence of populations. A sufficient condition for $\bar{X}_n^{(k)} - \mu_k$ to converge strongly to 0 in pr. for the sss. $\{n_k\}_1^\infty$ is that*

$$\lim_{k\to\infty} \sigma_k^2\,(n_k^{-1} - N_k^{-1}) = 0. \quad (5.1)$$

If $\{\pi_k\}_1^\infty$ satisfies

$$\lim_{A\to\infty}\,\sup_k \int_{|x|>A} x^2\,dF_k^c(x) = 0, \quad (5.2)$$

then condition (5.1) *is necessary already for $\bar{X}_n^{(k)} - \mu_k$ to converge in pr. to 0 for the sss. $\{n_k\}_1^\infty$.*

Proof. It is no loss of generality to assume that $\mu_k = 0$ for all k and we will do so. The sufficiency part of the theorem is an immediate consequence of Lemma 4.2. In the proof of the necessity part we will use the following simple lemma.

Lemma 5.1. *If for some integer l it holds that $S_l^{(k)} \to 0$ in pr. when $k \to \infty$, then $X_{k1} \to 0$ in pr. when $k \to \infty$.*

Proof. Assume the lemma to be false, i.e. there are positive numbers ε and ϱ and a subsequence $\{k_\nu\}_{\nu=1}^\infty$ such that $P(|X_{k_\nu 1}| \geqslant \varepsilon) \geqslant 2\varrho$, while

$$P(|S_l^{(k_\nu)}| \geqslant \varepsilon) \to 0 \text{ when } \nu \to \infty. \tag{5.3}$$

Then at least one of the inequalities $P(X_{k_\nu 1} \geqslant \varepsilon) \geqslant \varrho$ and $P(X_{k_\nu 1} \leqslant -\varepsilon) \geqslant \varrho$ holds for an infinity of ν's. We assume the first inequality. In other words, there is a sequence $\{\pi_{k_\tau}\}_{\tau=1}^{\infty}$ so that a portion $\geqslant \varrho$ of the elements in π_{k_τ} are $\geqslant \varepsilon$. Thus,

$$P(S_l^{(k_\tau)} \geqslant l\,\varepsilon) \geqslant \binom{N_{k_\tau}\varrho}{l}\binom{N_{k_\tau}}{l}^{-1} \to \varrho^l > 0 \text{ when } \tau \to \infty. \tag{5.4}$$

Now (5.3) and (5.4) contradict each other and the lemma is proved.

We also prove the necessity part of Theorem 5.1 by contradiction. We assume its negation to hold, i.e.

$$\bar{X}_{n_k}^{(k)} \to 0 \text{ in pr. when } k \to \infty \tag{5.5}$$

$$\sigma_k^2 (n_k^{-1} - N_k^{-1}) \geqslant \varrho > 0, \; k = 1, 2, \dots . \tag{5.6}$$

To guarantee (5.6) we may have to restrict to a subsequence of $\{\pi_k\}_1^\infty$. However, to simplify writing we use the same indices. From (5.2) it follows that $\sigma_k^2 \leqslant C < \infty$ and thus from (5.6) that $\sup_k n_k < \infty$. Thus there is a subsequence $\{k_\nu\}_{\nu=1}^{\infty}$ such that $n_{k_\nu} = l$. From (5.5) and Lemma 5.1 we conclude that $X_{k_\nu 1} \to 0$ in probability when $\nu \to \infty$. This is, however, incompatible with (5.6) when (5.2) holds, because

$$\begin{aligned} P(|X_{k_\nu 1}| \geqslant \varepsilon) &\geqslant P(\varepsilon \leqslant |X_{k_\nu 1}| \leqslant A) \geqslant A^{-2} \int_{\varepsilon \leqslant |x| \leqslant A} x^2 \, dF_{k_\nu}(x) \\ &\geqslant A^{-2}\left[(1 - N_{k_\nu}^{-1})\,\sigma_{k_\nu}^2 - \int_{|x|>A} x^2\, dF_{k_\nu}(x) - \varepsilon^2\right]. \end{aligned} \tag{5.7}$$

By choosing A so large that $\int_{|x|>A} x^2\, dF_{k_\nu}(x) < \varrho/6$ and $\varepsilon^2 < \varrho/6$ we get from (5.6) and (5.7) as soon as $N_{k_\nu} \geqslant 2$

$$P(|X_{k_\nu 1}| \geqslant \varepsilon) \geqslant A^{-2}(\varrho/2 - \varrho/6 - \varrho/6) > 0$$

and we have obtained a contradiction. Thus Theorem 5.1 is completely proved.

Corollary. $\{\pi_k\}_1^\infty$ is a population sequence which is non-degenerate and satisfies condition (5.2). Then a necessary and sufficient condition for $\bar{X}_n^{(k)} - \mu_k$ to converge strongly in pr. to 0 for the sss. $\{n_k\}_1^\infty$ is that

$$\lim_k n_k = +\infty. \tag{5.8}$$

Proof. From (5.2) we conclude that $\sigma_k^2 \leqslant C < \infty$. From the non-degenerateness of $\{\pi_k\}_1^\infty$ there follows the existence of $\varrho > 0$, $\varepsilon > 0$ and a k_0 so that $P(|X_{k1}| \geqslant \varepsilon) \geqslant \varrho > 0$ for $k \geqslant k_0$. Then $\sigma_k^2 \geqslant \varrho\varepsilon^2$ for $k \geqslant k_0$ and the corollary is now an immediate consequence of condition (5.1).

We shall consider the following proposition, which we denote (A).

$(A): \overline{X}_n^{(k)} - \mu_k$ converges strongly in pr. to 0 for every sss. $\{n_k\}_1^\infty$, which satisfies $n_k \to \infty$ when $k \to \infty$.

The above corollary says that (5.2) is a sufficient condition for (A) to hold. In the next theorem we weaken this condition.

Theorem 5.2. *Necessary and sufficient for (A) to hold is that $\{\pi_k\}_1^\infty$ satisfies*

$$\lim_{A\to\infty} \sup_k \int_{|x|>A} |x|\, dF_k^c(x) = 0. \tag{5.9}$$

Proof. Without loss of generality we assume that $\mu_k = 0$ for all k. First we prove the sufficiency part. Lemmata 4.3 and 4.4 yield

$$P(\max_{\nu \geqslant n_k} |\overline{X}_\nu^{(k)}| \geqslant \varepsilon) \leqslant \varepsilon^{-1} \cdot \frac{\lambda^{-1}(n_k)}{n_k} \left[\left(\frac{N_k - n_k}{N_k - 1} E\lambda(X_{k1}) + (E\lambda(X_{k1}))^2 \right)^{\frac{1}{2}} + E\lambda(X_{k1}) \right]. \tag{5.10}$$

The sufficiency of condition (5.9) now follows from (5.10) if we show that (5.9) guarantees the existence of a function $\lambda(x)$ which fulfils the conditions of Lemma 4.4 and in addition

$$\lambda^{-1}(x)/x \to 0 \text{ when } x \to \infty \tag{5.11}$$

$$\sup_k E\lambda(X_{k1}) < \infty. \tag{5.12}$$

From (5.9) it readily follows that $\sup_k E|X_{k1}| < \infty$ and that there exists a sequence $\{A_t\}_{t=1}^\infty$, $A_t \nearrow \infty$ such that

$$\sup_k \int_{|x|>A_t} |x|\, dF_k(x) \leqslant 3^{-t} \sup_k E|X_{k1}|, \quad t = 1, 2, \ldots .$$

Let $\{\tau_t\}_{t=1}^\infty$ satisfy 1) $\tau_t \nearrow \infty$ when $t \to \infty$. 2) $\tau_t \leqslant 2^t$, $t = 1, 2, \ldots$. 3) $(\tau_{t+1} - \tau_t)/(A_{t+1} - A_t) \leqslant \tau_t A_t^{-1}$, $t = 1, 2, \ldots$. Define for $x \geqslant 0$ $\varrho(x)$ to be the function the graph of which is the linear interpolation between the points $(0,0)$, (A_1, τ_1), $(A_2, \tau_2), \ldots$. It is easily checked that $\lambda(x) = |x| \varrho(|x|)$ satisfies the conditions in Lemma 4.4 and in (5.11), (5.12). Thus the sufficiency of (5.9) is proved.

Next we prove the necessity of (5.9). Remember the assumption that $\mu_k = 0$ for all k. First we show that if (A) is fulfilled, then it holds for every $\varepsilon > 0$ that

$$P(\max_{n \geqslant n_k} |X_{kn}/n| \geqslant \varepsilon) \to 0 \text{ if } n_k \to \infty \text{ when } k \to \infty. \tag{5.13}$$

(5.13) follows from the inequality below:

$$P(\max_{n \geqslant n_k} |X_{kn}/n| \geqslant \varepsilon) = P(\max_{n \geqslant n_k} |(S_n^{(k)} - S_{n-1}^{(k)})/n| \geqslant \varepsilon)$$

$$\leqslant P\left(\max_{n \geqslant n_k} |\overline{X}_n^{(k)}| \geqslant \frac{\varepsilon}{2}\right) + P\left(\max_{n \geqslant n_k} (1 - n^{-1}) |\overline{X}_{n-1}^{(k)}| \geqslant \frac{\varepsilon}{2}\right).$$

We shall give an indirect proof and we assume that (A) and thus (5.13) are fulfilled, but not (5.9). Then, by restricting to a subsequence of $\{\pi_k\}_1^\infty$ without introducing a new notation, there is a sequence of integers $\{A_k\}_1^\infty$, $A_k \nearrow \infty$, such that

$$\int_{|x|\geqslant A_k} |x|\, dF_k(x) \geqslant \varrho > 0, \quad k=1,2,\dots. \tag{5.14}$$

We can also assume that

$$P(|X_{k1}|\geqslant A_k) \to 0 \text{ when } k\to\infty \tag{5.15}$$

$$P(|X_{k1}|\geqslant N_k) = 0 \text{ for } \quad k\geqslant k_0, \tag{5.16}$$

because if (5.15) does not hold, then (5.13) cannot be fulfilled for $n_k = A_k$, and if (5.16) does not hold (5.13) cannot hold for $n_k=[N_k/2]$.

We have for $k \geqslant k_0$

$$\begin{aligned}
P(\max_{n\geqslant n_k} |X_{kn}/n|\geqslant 1) &= 1 - P(|X_{kn}|<n,\, n=n_k,\dots,N_k)\\
&= 1 - \prod_{n=n_k}^{N_k} P(|X_{kn}|<n \mid |X_{k\nu}|<\nu,\, \nu=n_k,\dots,n-1)\\
&= 1 - \prod_{n=n_k}^{N_k} \frac{N_k P(|X_{k1}|<n)-(n-n_k)}{N_k-(n-n_k)} \geqslant 1 - \prod_{n=n_k}^{N_k} P(|X_{k1}|<n)\\
&= 1 - \prod_{n=n_k}^{N_k}\left(1-\int_{|x|\geqslant n} dF_k(x)\right) \geqslant 1-\exp\left\{-\sum_{n=n_k}^{N_k}\int_{|x|\geqslant n} dF_k(x)\right\}\\
&\geqslant 1-\exp\left\{-\sum_{s=n_k}^{\infty}(s+1-n_k)\,P(s\leqslant |X_{k1}|<s+1)\right\}.
\end{aligned} \tag{5.17}$$

In the last step we used (5.16). By virtue of (5.15), we can choose $\{n_k\}_1^\infty$ such that $n_k\leqslant A_k$, $k=1,2,\dots$, $n_k\to\infty$, and $n_k P(|X_{k1}|\geqslant A_k)\to 0$ when $k\to\infty$. Then we get from (5.17) and (5.14)

$$\begin{aligned}
\underline{\lim}_k\, P(\max_{n=n_k} |X_{kn}/n|\geqslant 1) &\geqslant 1-\overline{\lim_k}\,\exp\left\{-\sum_{s=A_k}^{\infty}(s+1-n_k)\cdot P(s\leqslant |X_{k1}|<s+1)\right\}\\
&\geqslant 1-\overline{\lim_k}\,\exp\left\{-\int_{|x|\geqslant A_k}|x|\,dF_k(x)+n_k P(|X_{k1}|\geqslant A_k)\right\} \geqslant 1-e^{-\varrho}>0.
\end{aligned}$$

This result contradicts (5.13) and thus the necessity of (5.9) follows. Hence Theorem 5.2 is completely proved.

6. *The law of the iterated logarithm*

The sharpest result (under suitable conditions) about the asymptotic fluctuations of sample sums in the case of independent r.v.'s is the law of the iterated logarithm. We shall prove an analogue of this law and we shall first prove some lemmata.

Lemma 6.1. *π is a population of size N and π' is obtained from π by removal of t elements. Then*

$$\sigma_{\pi'}^2 \leqslant \sigma_\pi^2 (N-1)/(N-t-1), \quad t=1,2,\ldots,N-2.$$

Proof. Without loss of generality we assume that $\pi = \{a_1, a_2, \ldots, a_N\}$ has mean 0. Let $\pi' = \{a_1', a_2', \ldots, a_{N-t}'\}$

$$\sigma_{\pi'}^2 \leqslant (N-t-1)^{-1} \sum_{\nu=1}^{N-t} a_\nu'^2 \leqslant (N-t-1)^{-1} \sum_{\nu=1}^{N} a_\nu^2 = \sigma_\pi^2 \frac{N-1}{N-t-1}$$

and the lemma is proved. The next lemma is the counterpart of a well-known result for independent r.v.'s (see e.g. Loève [14] p. 248).

Lemma 6.2. *π is a population with mean* 0 *and variance σ^2, and x is a positive number. Then*

$$P(\max_{1\leqslant \nu \leqslant n} |S_\nu| \geqslant x) \leqslant 2P(|S_n| \geqslant x(1-n/N) - \sigma\sqrt{2n}).$$

Proof. For $t=1,2,\ldots,n$ A_t is the event $(\max_{\nu<t}|S_\nu| < x$ and $|S_t| \geqslant x)$. B is the event $(|S_n| \geqslant x(1-n/N) - \sigma\sqrt{2n})$. Let, as before, $\mathcal{B}_t$ be the algebra of events defined by conditions on $X_1, X_2, \ldots, X_t$. From the disjointness of $A_1, \ldots, A_n$ and the fact $A_t \in \mathcal{B}_t$ it follows that

$$P(B) \geqslant \sum_{t=1}^{n} P(BA_t) = \sum_{t=1}^{n} \int_{A_t} P^{\mathcal{B}_t}(|S_n| \geqslant x(1-n/N) - \sigma\sqrt{2n})\, P(d\omega). \qquad (6.1)$$

According to the conditioning principle (p. 385), the mixed conditional distribution of S_n, given $\mathcal{B}_t$, is at the point ω, equal to the distribution of $S_t(\omega) + S_{n-t}'$, where $S_{n-t}' = X_1' + \ldots = X_{n-t}'$ and $X_1', \ldots, X_{n-t}'$ is a random sample from the population $\pi'(\omega) = \pi$ with the elements $X_1(\omega), \ldots, X_t(\omega)$ removed. Thus

$$E^{\mathcal{B}_t} S_n = S_t\left(1 - \frac{n-t}{N-t}\right) \text{ and } \sigma^2(S_n \mid \mathcal{B}_t)(\omega) = \sigma_{\pi'(\omega)}^2 (n-t)(1-(n-t)/(N-t)).$$

Lemma 6.1 yields

$$\sigma^2(S_n \mid \mathcal{B}_t) \leqslant \sigma^2 (n-t)(1-(n-t)/(N-t))(N-1)/(N-t-1) \leqslant n\sigma^2 \qquad (6.2)$$

From Tchebychev's inequality we get

$$P^{\mathcal{B}_t}\left(|S_n| \geqslant x\left(1-\frac{n}{N}\right)-\sigma\sqrt{2n}\right) \geqslant P^{\mathcal{B}_t}\left(|S_n - E^{\mathcal{B}_t} S_n| \leqslant |E^{\mathcal{B}_t} S_n| - x\left(1-\frac{n}{N}\right)+\sigma\sqrt{2n}\right)$$

$$\geqslant 1-\sigma^2(S_n \mid \mathcal{B}_t)\left(\sigma\sqrt{2n}+|E^{\mathcal{B}_t} S_n| - x\left(1-\frac{n}{N}\right)\right)^{-2}. \tag{6.3}$$

When $\omega \in A_t$ it holds that $|S_t| \geqslant x$ and thus $|E^{\mathcal{B}_t} S_n| \geqslant x(1-(n-t)/(N-t)) \geqslant x(1-n/N)$. Combining this with (6.2) we get from (6.3) that

$$\text{when } \omega \in A_t \text{ then } P^{\mathcal{B}_t}(|S_n| \geqslant x(1-n/N)-\sigma\sqrt{2n}) \geqslant \tfrac{1}{2}. \tag{6.4}$$

We insert (6.4) into (6.1) and obtain

$$P(B) \geqslant \tfrac{1}{2}\sum_{t=1}^{n} P(A_t) \text{ i.e. } \sum_{t=1}^{n} P(A_t) \leqslant 2P(B),$$

which is the desired inequality, and Lemma 6.2 is proved.

Lemma 6.3. *$\pi = \{a_1, a_2, \ldots, a_N\}$ is a population with mean 0, variance σ^2 and $|a_\nu| \leqslant M$, $\nu = 1, 2, \ldots, N$. Then*

$$P\left(\frac{S_n}{\sigma\sqrt{n}} \geqslant \delta\right) \leqslant \exp\left\{-\frac{\delta^2}{2}\left(1-\frac{\delta M}{2\sigma\sqrt{n}}\right)\right\}, \quad n = 1, 2, \ldots, N.$$

Proof. This is a well-known inequality for independent r.v.'s (if σ^2 stands for ordinary variance), see e.g. Loève [14] p. 254. A scrutiny of the proof shows that it is an inequality of the Tchebychev type based on a function which satisfies the conditions of Lemma 3.2. Thus the inequality carries over to sampling without replacement. The relation $\sigma_\pi^2 = N(N-1)^{-1} EX_1^2$ only weakens the inequality.

Theorem 6.1. *$\{\pi_k\}_1^\infty$ is a sequence of populations, all having mean 0 and variance 1 and all elements on an interval $[-M, M]$. Then it holds for every $\varepsilon > 0$ and every sequence $\{n_k\}_1^\infty$ for which $\lim_k n_k = \infty$, that*

$$\lim_{k\to\infty} P\left(\max_{\nu \geqslant n_k} \frac{S_\nu^{(k)}}{\sqrt{2\nu \log\log \nu}} \leqslant 1+\varepsilon\right) = 1. \tag{6.5}$$

Conversely, to a given sequence $\{\pi_k\}_1^\infty$, which satisfies the above conditions, there exists a sequence $\{n_k\}_1^\infty$ with $\lim_k n_k = \infty$, such that for every $\varepsilon > 0$ it holds tha

$$\lim_{k\to\infty} P\left(\max_{\nu \geqslant n_k} \frac{S_\nu^{(k)}}{\sqrt{2\nu \log\log \nu}} \geqslant 1-\varepsilon\right) = 1. \tag{6.6}$$

Proof. First we prove the direct part (6.5). After the preparation with the above lemmata the rest of the proof is an almost verbatim repetition of the proof of the analogue result for independent r.v.'s. Let $0<\varepsilon<\frac{1}{4}$ and $\lambda>1$. We define $T_t=[\lambda^t]$, $t=1,2,\ldots$ and the events

$$A_{kt}=(\max_{1\leqslant\nu<T_{t+1}} S_\nu^{(k)}\geqslant(1+\varepsilon)\sqrt{2\,T_t\log\log T_t}),\quad k,t=1,2,\ldots.$$

First we only consider t-values for which $T_{t+1}\leqslant 2^{-1}\varepsilon N_k$. (As $N_k\to\infty$ when $k\to\infty$ such t-values exist when k is large enough.) From Lemma 6.2 we get

$$\alpha_{kt}=P(A_{kt})\leqslant 2P\left(|S^{(k)}_{T_{t+1}}|\geqslant(1+\varepsilon)\left(1-\frac{\varepsilon}{2}\right)\sqrt{2\,T_t\log\log T_t}-\sqrt{2\,T_{t+1}}\right).$$

As $T_{t+1}(T_t\log\log T_t)^{-1}\to 0$ when $t\to\infty$, we have for sufficiently large t's

$$\alpha_{kt}\leqslant 2P\left(|S^{(k)}_{T_{t+1}}|\geqslant\left(1+\frac{\varepsilon}{4}\right)\sqrt{2\,T_t\log\log T_t}\right)$$

and from Lemma 6.3 we conclude

$$\alpha_{kt}\leqslant 4\exp\left\{-\left(1+\frac{\varepsilon}{4}\right)^2\frac{T_t}{T_{t+1}}\log\log T_t\left(1-\frac{M\sqrt{2\,T_t\log\log T_t}}{2\,T_{t+1}}\right)\right\}.$$

We assume that λ was chosen $<1+\varepsilon/4$. Then it holds if t is sufficiently large

$$\alpha_{kt}\leqslant 4\exp\left\{-\left(1+\frac{\varepsilon}{4}\right)\log\log T_t\right\}\leqslant C(\lambda)\,t^{-(1+\varepsilon/4)},\tag{6.7}$$

where $C(\lambda)$ is a constant depending on λ. We define $t_k'=\max(t\,|\,T_t\leqslant n_k)$ and $t_k''=\min(t\,|\,T_t\geqslant 3^{-1}\varepsilon N_k)$. Then

$$P\left(\max_{n_k\leqslant\nu\leqslant 3^{-1}\varepsilon N_k}\frac{S_\nu^{(k)}}{\sqrt{2\,\nu\log\log\nu}}\geqslant 1+\varepsilon\right)\leqslant P\left(\bigcup_{t=t'}^{t''}A_{kt}\right)\leqslant\sum_{t=t_k'}^{t_k''}\alpha_{kt}.\tag{6.8}$$

As t_k' tends to infinity with k it follows from (6.7) that the sum in (6.8) tends to 0 when $k\to\infty$. Thus, we have proved: If $0<\varepsilon<\frac{1}{4}$, $n_k<3^{-1}\varepsilon N_k$, and $n_k\to\infty$ when $k\to\infty$, then

$$\lim_{k\to\infty}P\left(\max_{n_k\leqslant\nu\leqslant 3^{-1}\varepsilon N_k}\frac{S_\nu^{(k)}}{\sqrt{2\,\nu\log\log\nu}}\geqslant 1+\varepsilon\right)=0.\tag{6.9}$$

From (6.9) one easily deduces that

$$\lim_{k\to\infty}P\left(\max_{s6^{-1}\varepsilon N_k\leqslant\nu\leqslant(s+1)6^{-1}\varepsilon N_k}\frac{S_\nu^{(k)}}{\sqrt{2\,\nu\log\log\nu}}\geqslant 1+\varepsilon\right)=0,\quad s=1,2,\ldots,\left[\frac{6}{\varepsilon}\right].\tag{6.10}$$

The direct part of Theorem 6.1 now follows from (6.9) and (6.10). We shall prove the converse part by a reduction to the corresponding result for independent r.v.'s. We introduce a notation; $X_1, X_2, \ldots$ are r.v.'s and $S_\nu = X_1 + \ldots + X_\nu$. Then $A(n, m, \varepsilon)$ stands for the event $A(n, m, \varepsilon) = (\max_{n \leqslant \nu \leqslant m} S_\nu (2\nu \log \log \nu)^{-\frac{1}{2}} > 1 - \varepsilon)$. The pr. of $A(n, m, \varepsilon)$ depends on the distribution of the X's. When the X's are assumed independent with the same d.f. F we indicate the dependence on F by writing $P_F(A)$ and when the X's are sampled from the population π we write $P_\pi(A)$. The following proposition (B) is a consequence of the law of the iterated logarithm for independent r.v.'s

(B): Let $X_1, X_2, \ldots$ be independent copies of a r.v. X with d.f. F, which satisfies $EX = 0$, $\sigma(X) = 1$ and $|X| \leqslant M$ (with pr. 1). Then, for all positive numbers ε and δ and any natural number m, there is an N, depending only on ε, δ, m and M, such that

$$P_F(A(m, N, \varepsilon)) \geqslant 1 - \delta. \tag{6.11}$$

Let $\{\varepsilon_t\}_1^\infty$ and $\{\delta_t\}_1^\infty$ be sequences such that $\varepsilon_t \searrow 0$ and $\delta_t \searrow 0$ when $t \to \infty$. From (B) there follows the existence of a sequence $\{m_t\}_1^\infty$ of integers, $m_t \nearrow \infty$, such that $P_F(A(m_t, m_{t+1}, \varepsilon_t)) \geqslant 1 - \delta_t, t = 1, 2, \ldots$ if F satisfies the conditions in (B). We say that a population π has property E_t if $P_\pi(A(m_t, m_{t+1}, \varepsilon_t)) \geqslant 1 - 2\delta_t$. Now let $\{\pi_k\}_1^\infty$ be the given population sequence. We claim that for every fixed t only a finite number of the π_k's lack property E_t. Assume the contrary. Then, by selecting a subsequence (without introducing a new notation) we can assume that $F_k \to F_0$ i.d. when $k \to \infty$ and that $P_{\pi_k}(A(m_t, m_{t+1}, \varepsilon_t)) < 1 - 2\delta_t$. The event $A(m_t, m_{t+1}, \varepsilon_t)$ depends only on the finitely many variables $X_1, \ldots, X_{m_{t+1}}$. Thus from Lemma 3.1 we conclude that $P_{F_0}(A(m_t, m_{t+1}, \varepsilon_t)) = \lim_k P_{\pi_k}(A(m_t, m_{t+1}, \varepsilon_t)) \leqslant 1 - 2\delta_t$, which yields a contradiction. Thus only finitely many π_k's lack property E_t. Let for $t = 1, 2, \ldots, k(t) = \min\ (k \mid \pi_\nu$ has property $E_t, \nu \geqslant k)$ and let $n_k = m_t$ when $\max_{s \leqslant t} k(s) \leqslant k < \max_{s \leqslant t+1} k(s)$. It is easily checked that (6.6) holds for this sequence $\{n_k\}_1^\infty$ and that $n_k \to \infty$ when $k \to \infty$. This concludes the proof.

Chapter 3. On stochastic processes with continuous sample paths

7. *Generalities*

We shall later study certain stochastic processes related to samples from finite populations, and especially convergence of such processes. We shall then make use of the general convergence theory for stochastic processes, worked out especially by Prokhorov [18]. We give a brief exposition of some fundamental concepts and results, and we follow closely Prokhorov's ideas.

$C[0, 1]$ is the metric space of all real-valued continuous functions on [0, 1], with the uniform metric. Points in $C[0, 1]$ will be denoted x or $x(t)$. By *a stochastic process on* [0, 1] *with continuous sample paths* we mean a comple probability measure P, defined on a σ-algebra of sets in $C[0, 1]$ including all closed sets, and which is inner regular with respect to closed sets, see def. on p. 162 in [18]. When f is a measurable mapping of $C[0, 1]$ into another metric space S, we write P^f for the measure on S, which is the forward transportation of P

by f, see [18], p. 163. When f is real-valued we call it a functional of the process or a r.v. defined on the process. For a finite set M; $(t_1, t_2, \ldots, t_m)$, $t_\nu < t_{\nu+1}$, $\nu = 1, 2, \ldots, m-1$, of points in $[0, 1]$ we define the *marginal distribution* P^M as the measure in R^m which is induced by P and the mapping $x \to (x(t_1), x(t_2), \ldots, x(t_m))$, $x \in C[0, 1]$. A process is uniquely determined by its marginal distributions.

It is a nontrivial problem to determine if a given set of "marginal" distributions actually are the marginals of a process with continuous sample paths, see Loève [14], Theorem B, p. 517, and also Prokhorov [18].

8. *Wiener processes*

We will be much concerned with the stochastic processes known as Wiener processes and tied-down (or conditional) Wiener processes. These processes are real Gaussian processes and can be characterized by their mean value function $M(t) = EX(t)$ and their covariance function $R(s,t) = \text{Cov}\,(X(t), X(s))$. The ordinary *Wiener process* (with parameter σ^2) is defined by $M(t) = 0$ and $R(s,t) = \sigma^2 \min\,(s,t)$. It is well known that this process has continuous sample paths, see e.g. Loève [14], p. 547, and we can identify its restriction to $0 \leqslant t \leqslant 1$ with a measure on $C[0, 1]$. This measure we denote $W(\sigma^2)$.

Let $X(t)$ be a Wiener process with parameter σ^2. *The tied-down Wiener process* with tying point T can be intuitively described as the process $X(t)$ under the condition $X(T) = 0$. Formally we define it by $M(t) = 0$, $R(s,t) = \sigma^2 s(1 - t/T)$, $0 \leqslant s \leqslant t \leqslant T$. These processes also have continuous sample paths and we denote the measure on $C[0, 1]$ corresponding to the part $0 \leqslant t \leqslant 1$ of the process by $W(\sigma^2, T)$. We make the notational convention that $W(\sigma^2, \infty) = W(\sigma^2)$. A treatment of these processes can be found in Doob [5].

9. *Convergence of stochastic processes*

P_k, $k = 1, 2, \ldots$ and P are stochastic processes with continuous sample paths, i.e. pr. measures on $C[0, 1]$. We say that P_k *converges* (*weakly*) *to* P when $k \to \infty$, denoted $P_k \Rightarrow P$ when $k \to \infty$, if

$$\int_{C[0,1]} f(x)\, P_k(dx) \to \int_{C[0,1]} f(x)\, P(dx) \quad \text{when } k \to \infty$$

for every continuous, bounded functional f on $C[0, 1]$. See definition on p. 164 in [18].

The suitability of this convergence concept follows from the next lemma, which is contained in Theorem 1.8 in [18].

Lemma 9.1. *If $P_k \Rightarrow P$ and if f is a functional which is continuous almost everywhere (P), then $P_k^f \to P^f$ i.d. when $k \to \infty$.*

The following concept is crucial for characterization of weak compactness of families of pr. measures on $C[0, 1]$. The family $\mathcal{D} = \{P\}$ is said to be *tight* (to satisfy condition $(\varkappa)$ in [18]) if for every $\varepsilon > 0$ there is a compact set $K_\varepsilon \subset C[0, 1]$

such that $P(K_\varepsilon) > 1-\varepsilon$ holds for all $P \in \mathcal{D}$ (see p. 167 in [18]). The following result is contained in [18].

Theorem 9.1. *Necessary and sufficient for $P_k \Rightarrow P$ when $k \to \infty$ is that*

1. $P_k^M \to P^M$ *i.d. for every marginal* M.
2. $\{P_k\}_1^\infty$ *is tight.*

In order to apply the above theorem, one needs a manageable criterion for verification of tightness. The following criterion, usually referred to as Dynkin's criterion, will satisfy in the cases we shall consider.

Theorem 9.2. $\{P_k\}_1^\infty$ *is a sequence of pr. measures on* $C[0,1]$. *For* $\delta > 0$ and $\Delta > 0$ *we define*

$$\psi(\Delta, \delta) = \overline{\lim_{k\to\infty}} \sup_{0 \leqslant T \leqslant 1-\Delta} P_k\big(\max_{T \leqslant t \leqslant T+\Delta} |x(t) - x(T)| \geqslant \delta\big). \tag{9.1}$$

Then, sufficient for $\{P_k\}_1^\infty$ to be tight is that

1. *For every $\varepsilon > 0$ there is a constant C_ε such that*

$$P_k(|x(0)| \leqslant C_\varepsilon) \geqslant 1-\varepsilon, \quad k = 1, 2, \ldots.$$

2. *For every fixed δ it holds that $\Delta^{-1}\psi(\Delta, \delta) \to 0$ when $\Delta \to 0$.*

We indicate a proof (cf. Lemma 2.3 in [18], and Particular case 1^0 of Theorem A in 35.3, [14]). For $\delta > 0$ and $\Delta > 0$ we define the following subsets of $C[0,1]$.

$$B(\Delta, \delta) = (x \,\|\, x(t') - x(t'')| < \delta \text{ if } |t' - t''| \leqslant \Delta)$$

$$A_s(\Delta, \delta) = (x \mid \max_{s\Delta \leqslant t \leqslant (s+1)\Delta} |x(t) - x(s\Delta)| < \delta), \quad s = 0, 1, \ldots, [\Delta^{-1}].$$

For any positive numbers ε and δ there is a Δ_0 such that $P_k(B(\Delta_0, \delta)) > 1-\varepsilon, k = 1, 2, \ldots$. To prove this we chose Δ_0' such that

$$\Delta_0'^{-1}\psi\left(\Delta_0', \frac{\delta}{3}\right) < \frac{\varepsilon}{4}$$

and k_0 so that

$$\sup_T P_k\left(\max_{T \leqslant t \leqslant T+\Delta_0'} |x(t) - x(T)| \geqslant \frac{\delta}{3}\right) \leqslant 2\psi\left(\Delta_0', \frac{\delta}{3}\right)$$

for $k \geqslant k_0$. Then it holds for $k \geqslant k_0$

$$P_k(B(\Delta_0', \delta)) \geqslant 1 - \sum_{s=0}^{[\Delta_0'^{-1}]} P_k\left(A_s'\left(\Delta_0', \frac{\delta}{3}\right)\right) \geqslant 1 - 2([\Delta_0'^{-1}]+1)\psi\left(\Delta_0', \frac{\delta}{3}\right) > 1 - \varepsilon.$$

Further there exist $\Delta'_k > 0, k = 1, 2, \ldots$ so that $P_k(B(\Delta'_k, \delta)) > 1 - \varepsilon$. The proposition above now follows by putting $\Delta_0 = \min(\Delta'_0, \Delta'_1, \ldots, \Delta'_{k_0})$.

Let $\varepsilon > 0$ and $\delta_\nu \searrow 0, \nu \to \infty$. There is a sequence $\Delta_\nu \searrow 0, \nu \to \infty$, for which $P_k(B(\Delta_\nu, \delta_\nu)) > 1 - \varepsilon \cdot 2^{-\nu}, k, \nu = 1, 2, \ldots$. Now consider $K = \bigcap_{\nu=1}^{\infty} B(\Delta_\nu, \delta_\nu)$. The functions in K are equi-continuous with a common modulus of continuity given by the linear interpolation between the points $(\Delta_2, \delta_1), (\Delta_3, \delta_2), \ldots$. Further $P(K) > 1 - \varepsilon$ and the theorem follows easily.

Chapter 4. Central limit problems

10. *Preliminaries*

Analogues of the central limit theorem for sampling from finite populations have been extensively studied. Among the works on the problem we mention Wald and Wolfowitz [19], Madow [15], Noether [17], Hoeffding [11], Motoo [16], Erdös and Rényi [6], Hájek [8], and [9].

We state the problem we shall consider first in this chapter. As usual, $X_{k1}, \ldots, X_{kN_k}$ is a r.p. of the elements in π_k, $S_0^{(k)} = 0$, $S_n^{(k)} = X_{k1} + \ldots + X_{kn}$, $1 \leqslant n \leqslant N_k$, $k = 1, 2, \ldots$. For simplicity we write S_{n_k} instead of $S_{n_k}^{(k)}$. $S_{n_k}^*$ stands for the standardized sample sum

$$S_{n_k}^* = (S_{n_k} - ES_{n_k})/\sigma(S_{n_k}). \tag{10.1}$$

The problem is to determine the class of possible limit distributions for $S_{n_k}^*$ and the conditions for convergence. In Hájek [9] there is given a complete treatment of the problem under the infinitesimality condition (N) defined in Remark 1 to Th. 12.1. We shall consider the problem without assumption (N), but under the additional assumption, that

$$0 < \overline{\lim_k}\, n_k/N_k \leqslant \underline{\lim_k}\, n_k/N_k < 1 \tag{10.2}$$

Our analysis will be based on the fundamental Lemma 2.1 in [8], a consequence of which we now state. Let $\pi_k = \{a_{k\nu}\}_{\nu=1}^{N_k}$ and let $Y_{k1}, \ldots, Y_{kN_k}$ be independent r.v.'s with the following two-point distributions

$$P(Y_{k\nu} = a_{k\nu} - \mu_k) = 1 - P(Y_{k\nu} = 0) = n_k/N_k, \ \nu = 1, 2, \ldots, N_k, \ k = 1, 2, \ldots \tag{10.3}$$

We define $Z_{n_k} = Y_{k1} + \ldots + Y_{kN_k}$, $k = 1, 2, \ldots$ and $Z_{n_k}^* = Z_{n_k}/\sigma(Z_{n_k})$. The distributions of $S_{n_k}^*$ and $Z_{n_k}^*$ are denoted respectively $F_{n_k}^*(x)$ and $G_{n_k}^*(x)$. The lemma below is an immediate consequence of Lemma 2.1 in [8].

Lemma 10.1. *If condition* (10.2) *is fulfilled, then* $\{F_{n_k}^*\}_1^\infty$ *and* $\{G_{n_k}^*\}_1^\infty$ *possess limit distributions simultaneously, and if they have limit distributions these coincide.*

Following Hájek we will derive results about the convergence of $\{F_{n_k}^*\}_1^\infty$ by considering the sequence $\{G_{n_k}^*\}_1^\infty$. $G_{n_k}^*$ is a convolution of two-point distributions and we start by studying such distributions.

11. *On convolutions of certain two-point distributions*

We shall consider pairs (c, d) of sequences of real numbers, $c = \{c_1, c_2, \ldots\}$, $d = \{d_1, d_2, \ldots\}$. We assume throughout that such sequences satisfy $c_1 \leqslant c_2 \leqslant \ldots \leqslant 0$ and $\lim_\nu c_\nu = 0$, $d_1 \geqslant d_2 \geqslant \ldots \geqslant 0$ and $\lim_\nu d_\nu = 0$. If a sequence c or d is given only for a finite number of indices it is completed to an infinite sequence by addition of zeros. By $|(c, d)|$ we mean the sequence we get by arranging $-c_1$, $-c_2, \ldots, d_1, d_2, \ldots$ in decreasing order, and with $(0, 0)$ we mean the pair in which both sequences consist of only zeros. It will be convenient to have a notation for componentwise convergence.

Definition. $(c(k), d(k)) \Rightarrow (c, d)$ when $k \to \infty$ if $\lim_k c_\nu(k) = c_\nu$ and $\lim_k d_\nu(k) = d_\nu$, for $\nu = 1, 2, \ldots$.

The following result is easily proved.

Lemma 11.1. *The set of pairs (c, d) for which $\sum_1^\infty (c_\nu^2 + d_\nu^2) \leqslant M < \infty$ is compact under $\Rightarrow$ convergence.*

Next we define a class of two-point distributions. For $0 < \lambda < 1$ $\Lambda(x; \lambda)$ is the pr. distribution with pr. $(1-\lambda)$ in the point $-\sqrt{\lambda/(1-\lambda)}$ and pr. λ in the point $\sqrt{(1-\lambda)/\lambda}$. The corresponding c.f. $\varphi(t; \lambda)$ is

$$\varphi(t; \lambda) = (1-\lambda) \exp\{-it\sqrt{\lambda/(1-\lambda)}\} + \lambda \exp\{it\sqrt{(1-\lambda)/\lambda}\} \tag{11.1}$$

Lemma 11.2. *(c, d) is a pair for which $\sum_1^\infty (c_\nu^2 + d_\nu^2) < \infty$ and $0 < \lambda < 1$. Then*

$$F(x; \lambda, c, d) = \prod_{\nu=1}^{\infty}{}^{*} \Lambda\left(\frac{x}{c_\nu}; \lambda\right) * \prod_{\nu=1}^{\infty}{}^{*} \Lambda\left(\frac{x}{d_\nu}; \lambda\right) \tag{11.2}$$

converges. ($*$ and $\prod^*$ stand for convolution and $\Lambda(x/0; \lambda)$ is the d.f. with pr. 1 in $x = 0$). *The representation of $F(x; \lambda, c, d)$ as a convolution of factors $\Lambda(x/c_\nu; \lambda)$ and $\Lambda(x/d_\nu; \lambda)$ is unique in the following sense when $(c, d) \neq (0, 0)$:*

(i) *If $\lambda \neq \frac{1}{2}$ there are exactly two sets of parameters which yield the same $F(x; \lambda, c, d)$, namely, $F(x; \lambda, c, d) = F(x; 1-\lambda, -d, -c)$.*

(ii) *If $\lambda = \frac{1}{2}$ it holds that $F(x, \frac{1}{2}, c, d) = F(x, \lambda', c', d')$ if and only if $\lambda' = \frac{1}{2}$ and $|(c', d')| = |(c, d)|$.*

Proof. In order to show that (11.2) converges we show that the corresponding product of c.f.'s

$$\psi(t; \lambda, c, d) = \prod_{\nu=1}^{\infty} \varphi(c_\nu t; \lambda) \cdot \prod_{\nu=1}^{\infty} \varphi(d_\nu t; \lambda) \tag{11.3}$$

converges uniformly on every compact interval on the real axis. It will be convenient to study ψ for complex arguments. Easy eastimates yield that

$$|1 - \varphi(z; \lambda)| \leqslant C|z|^2 \tag{11.4}$$

holds on every compact set in the complex plane. The constant C depends on λ and the compact set. According to the assumption $\sum (c_\nu^2 + d_\nu^2) < \infty$ it follows from (11.4) that the product (11.3) converges uniformly on every compact set in the complex plane. Thus $\psi(z; \lambda, c, d)$ is an entire function. Especially we have uniform convergence of (11.3) on every compact interval on the real axis. Thus $\varphi(t; \lambda, c, d)$ is a c.f. and $F(x; \lambda, c, d)$ is a d.f.

To prove the uniqueness part we consider the zeros of ψ. From the uniform convergence of (11.3) it follows that the zeros of ψ are exactly those of the factors $\varphi(c_\nu z; \lambda)$ and $\varphi(d_\nu z; \lambda)$. The zeros of $\varphi(\varrho z; \lambda)$ are

$$z = \varrho^{-1} \sqrt{\lambda(1-\lambda)}\,[(2n+1)\pi - i \ln \{(1-\lambda)/\lambda\}], \quad n = 0, \pm 1, \pm 2, \dots \tag{11.5}$$

Suppose we have that $F(x; \lambda, c, d) = F(x; \lambda', c', d')$, where $\sum (c_\nu^2 + d_\nu^2) < \infty$ and $\sum (c_\nu'^2 + d_\nu'^2) < \infty$. The unique correspondence between c.f.'s and d.f.'s yields that (11.6) then holds for all real z. But both sides in (11.6) are entire functions and thus (11.6) holds for all z.

$$\prod_{\nu=1}^{\infty} \varphi(c_\nu z; \lambda) \prod_{\nu=1}^{\infty} \varphi(d_\nu z; \lambda) = \prod_{\nu=1}^{\infty} \varphi(c_\nu' z; \lambda') \prod_{\nu=1}^{\infty} \varphi(d_\nu' z; \lambda'). \tag{11.6}$$

First we assume $0 < \lambda < \frac{1}{2}$ and $0 < \lambda' \leqslant \frac{1}{2}$. By equating the zeros of the left- and right-hand sides of (11.6), which are closest to the origin in the quadrant $\operatorname{Im}(z) \geqslant 0$, $\operatorname{Re}(z) > 0$, we get

$$c_1^{-1} \sqrt{\lambda(1-\lambda)}\,(-\pi - i \ln \{(1-\lambda)/\lambda\}) = c_1'^{-1} \sqrt{\lambda'(1-\lambda')}\,(-\pi - i \ln \{(1-\lambda')/\lambda'\}).$$

Thus
$$\frac{c_1'}{c_1} = \frac{\sqrt{\lambda'(1-\lambda')}}{\sqrt{\lambda(1-\lambda)}} = \frac{\sqrt{\lambda'(1-\lambda')}}{\sqrt{\lambda(1-\lambda)}} \cdot \frac{\ln \{(1-\lambda')/\lambda'\}}{\ln \{(1-\lambda)/\lambda\}},$$

which yields $\lambda = \lambda'$ and $c_1 = c_1'$. By considering the zeros closest to the origin in the quadrant $\operatorname{Im}(z) \leqslant 0$, $\operatorname{Re}(z) > 0$ we obtain that $d_1 = d_1'$. Now we can cancel the factors $\varphi(c_1 z; \lambda)$ and $\varphi(d_1 z; \lambda)$ in (11.6), repeat the argument, and get that $c_2 = c_2'$ and $d_2 = d_2'$ and so on. Next we assume $0 < \lambda < \frac{1}{2}$ and $\frac{1}{2} < \lambda' < 1$. Again, by considering the zeros closest to the origin in the quadrant $\operatorname{Re}(z) > 0$, $\operatorname{Im}(z) > 0$ for the right- and left-hand sides of (11.6) we get that

$$c_1^{-1} \sqrt{\lambda(1-\lambda)}\,(-\pi - i \ln \{(1-\lambda)/\lambda\}) = d_1'^{-1} \sqrt{\lambda'(1-\lambda')}\,(\pi - i \ln \{(1-\lambda')/\lambda'\})$$

which yields $\lambda' = 1 - \lambda$ and $d_1' = -c_1$. By proceeding as above we obtain that a necessary condition for (11.6) to hold is that $\lambda' = 1 - \lambda$ and $(c', d') = (-d, -c)$. It is easily checked that this condition is also sufficient. In quite an analogous manner the proposition can be demonstrated for $\frac{1}{2} < \lambda < 1$ and $\lambda = \frac{1}{2}$. Thus the lemma is proved. The following result is immediate.

Lemma 11.3. *$F(x; \lambda, c, d)$ has mean 0 and variance $\sum_1^\infty (c_\nu^2 + d_\nu^2)$.*

Next we prove a convergence result.

Lemma 11.4. *$F_k(x) = F(x; \lambda_k, c(k), d(k))$, $k = 1, 2, \ldots$ is a sequence of distributions defined according to* (11.2) *for which it holds that $0 < \underline{\lim}_k \lambda_k \leqslant \overline{\lim}_k \lambda_k < 1$ and $\sum_1^\infty (c_\nu^2(k) + d_\nu^2(k)) = 1$.*

(i) *The possible limit distributions for F_k when $k \to \infty$ are those which can be written*

$$N(0, 1 - \sum_{\nu=1}^{\infty} (c_\nu^2 + d_\nu^2)) * \prod_{\nu=1}^{\infty}{}^{*} \Lambda(x/c_\nu; \lambda) * \prod_{\nu=1}^{\infty}{}^{*} \Lambda(x/d_\nu; \lambda) \tag{11.7}$$

where (c, d) satisfies $\sum_1^\infty (c_\nu^2 + d_\nu^2) \leqslant 1$ and $0 < \lambda < 1$. $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

(ii) *Necessary and sufficient for $F_k(x)$ to convergence i.d. to the distribution* (11.7) *when $k \to \infty$ is*

1. *if $(c, d) = (0, 0)$ that $(c(k), d(k)) \Rightarrow (0, 0)$ when $k \to \infty$.*
2. *if $(c, d) \neq (0, 0)$ and $\lambda \neq \frac{1}{2}$ that*
   (a) *$\{\lambda_k\}_1^\infty$ has at most the two limits points λ and $1 - \lambda$.*
   (b) *for every subsequence $\{k_\nu\}_{\nu=1}^\infty$ for which $\lim_\nu \lambda_{k_\nu} = \lambda$ it holds that $(c(k_\nu), d(k_\nu)) \Rightarrow (c, d)$ and for every subsequence for which $\lim_\nu \lambda_{k_\nu} = 1 - \lambda$ it holds that $(c(k_\nu), d(k_\nu)) \Rightarrow (-d, -c)$.*
3. *if $(c, d) \neq (0, 0)$ and $\lambda = \frac{1}{2}$ that*
   (a) *$\lambda_k \to \frac{1}{2}$ when $k \to \infty$.*
   (b) *$|(c(k), d(k))| \Rightarrow |(c, d)|$ when $k \to \infty$.*

We will use the following estimate in the proof.

Lemma 11.5. *$\psi(t; \lambda, c, d)$ is defined in* (11.3). *If*

$$|\lambda - \tfrac{1}{2}| \leqslant \varrho < \tfrac{1}{2} \quad \text{and} \quad |t| \leqslant \min(-c_{N+1}^{-1}, d_{N+1}^{-1}),$$

then
$$\left| \psi(t; \lambda, c, d) - \exp\left\{-\frac{t^2}{2} \sum_{N+1}^{\infty} (c_\nu^2 + d_\nu^2)\right\} \cdot \prod_{\nu=1}^{N} \varphi(c_\nu t; \lambda)\, \varphi(d_\nu t; \lambda) \right|$$

$$\leqslant \exp\left\{C |t|^3 \max(-c_{N+1}, d_{N+1}) \cdot \sum_{N+1}^{\infty} (c_\nu^2 + d_\nu^2)\right\} - 1 \tag{11.8}$$

where the constant C only depends on ϱ.

Proof. $C_1, C_2, \ldots$ denotes constants. It holds that

$$\varphi(t; \lambda) = 1 - \tfrac{1}{2} t^2 (1 + t R_1(t; \lambda)), \tag{11.9}$$

where $|R_1(t; \lambda)| \leqslant C_1$ if $|t| \leqslant 1$ and $|\lambda - \frac{1}{2}| \leqslant \varrho < \frac{1}{2}$. Further we have

$$1 - z = e^{-z}(1 + z^2 R_2(z)), \tag{11.10}$$

where $|R_2(z)| \leqslant C_2$ if $|z| \leqslant 2^{-1}(1+C_1)$. By combining (11.9) and (11.10) we get for real α's

$$\varphi(\alpha t; \lambda) = e^{-(\alpha^2 t^2)/2}(1 + |\alpha t|^3 R_3(t, \alpha, \lambda)), \tag{11.11}$$

where $|R_3(t, \alpha, \lambda)| \leqslant C_3$ if $|t| \leqslant |\alpha^{-1}|$ and $|\lambda - \frac{1}{2}| \leqslant \varrho < \frac{1}{2}$. Thus

$$\prod_{N+1}^{\infty} \varphi(c_\nu t; \lambda)\, \varphi(d_\nu t; \lambda) = \exp\left\{-\frac{t^2}{2}\sum_{N+1}^{\infty}(c_\nu^2 + d_\nu^2)\right\} \cdot \prod_{N+1}^{\infty}(1 + |c_\nu t|^3 R_3)(1 + |d_\nu t|^3 R_3')$$

and if $|t| \leqslant \min(-c_{N+1}^{-1}, d_{N+1}^{-1})$ and $|\lambda - \frac{1}{2}| \leqslant \varrho < \frac{1}{2}$ it holds that

$$\prod_{N+1}^{\infty}(1 + |c_\nu t|^3 R_3)(1 + |d_\nu t|^3 R_3') \leqslant \exp\left\{C_3 |t|^3 \sum_{N+1}^{\infty}(|c_\nu|^3 + d_\nu^3)\right\}$$

$$\leqslant \exp\left\{C_3 |t|^3 \cdot \max(-c_{N+1}, d_{N+1}) \sum_{N+1}^{\infty}(c_\nu^2 + d_\nu^2)\right\}$$

and the estimate (11.8) follows.

Proof of Lemma 11.4. First we assume

$$\lambda_k \to \lambda \text{ and } (c(k), d(k)) \Rightarrow (c, d) \text{ when } k \to \infty. \tag{11.12}$$

According to Lemma 11.1 it holds that $\sum_1^\infty (c_\nu^2 + d_\nu^2) \leqslant 1$ and we have

$$\left|\psi(t; \lambda, c, d) \exp\left\{-\frac{t^2}{2}\left(1 - \sum_{\nu=1}^{\infty}(c_\nu^2 + d_\nu^2)\right)\right\} - \psi(t; \lambda_k, c(k), d(k))\right|$$

$$\leqslant \left|\exp\left\{-\frac{t^2}{2}\left(1 - \sum_{\nu=1}^{\infty}(c_\nu^2 + d_\nu^2)\right)\right\}\left[\psi(t; \lambda, c, d) - \prod_{\nu=1}^{N}\varphi(c_\nu t; \lambda)\, \varphi(d_\nu t; \lambda)\right.\right.$$

$$\left.\left.\times \exp\left\{-\frac{t^2}{2}\sum_{N+1}^{\infty}(c_\nu^2 + d_\nu^2)\right\}\right]\right| + \left|\psi(t; \lambda_k, c(k), d(k)) - \exp\left\{-\frac{t^2}{2}\sum_{N+1}^{\infty}(c_\nu^2(k) + d_\nu^2(k))\right\}\right.$$

$$\left.\times \prod_{\nu=1}^{N}\varphi(c_\nu(k)t; \lambda_k)\, \varphi(d_\nu(k)t; \lambda_k)\right| + \left|\exp\left\{-\frac{t^2}{2}\left(1 - \sum_{\nu=1}^{N}(c_\nu^2 + d_\nu^2)\right)\right\}\right.$$

$$\times \prod_{\nu=1}^{N}\varphi(c_\nu t; \lambda)\, \varphi(d_\nu t; \lambda) - \exp\left\{-\frac{t^2}{2}\left(1 - \sum_{\nu=1}^{N}(c_\nu^2(k) + d_\nu^2(k))\right)\right\}$$

$$\left.\times \prod_{\nu=1}^{N}\varphi(c_\nu(k)t; \lambda_k)\, \varphi(d_\nu(k)t; \lambda_k))\right| = R_4(t, N) + R_5(t, N, k) + R_6(t, N, k).$$

Let T and ε be arbitrary positive numbers. From (11.12) we conclude that $\sup_k \max(-c_N(k), d_N(k)) \to 0$ when $N \to \infty$ and that $|\lambda_k - \frac{1}{2}| \leqslant \varrho < \frac{1}{2}$. The estimate (11.8) now yields that there is a N_0, depending on T and ε, such that $|R_4(t, N_0)| \leqslant \varepsilon$ and $|R_5(t, N_0, k)| \leqslant \varepsilon$ if $|t| \leqslant T$. It is an immediate consequence of (11.12) that

$R_6(t, N_0, k) \to 0$ uniformly on every compact t-interval when $k \to \infty$. Thus, in all, we have proved that if (11.12) holds then it holds uniformly on every compact t-interval that

$$\lim_k \psi(t; \lambda_k, c(k), d(k)) = \exp\left\{-\frac{t^2}{2}\left(1 - \sum_{\nu=1}^{\infty}(c_\nu^2 + d_\nu^2)\right)\right\} \cdot \psi(t; \lambda, c, d). \tag{11.13}$$

According to the continuity theorem for c.f.'s this is equivalent to the following:

$$\text{(11.12) implies that } F_k \to N\left(0, 1 - \sum_{\nu=1}^{\infty}(c_\nu^2 + d_\nu^2)\right) * F(x; \lambda, c, d) \text{ i.d.} \tag{11.14}$$

We observe that the assumption $\lambda_k \to \lambda$ was used only to assure that $R_6(t, N_0, k) \to 0$. It is easily seen that if $(c, d) = (0. 0)$ then the assumption $\lambda_k \to \lambda$ is superfluous for this, and the sufficiency part of (ii) 1 in the lemma follows.

We prove the sufficiency parts of (ii) 2 and 3 indirectly and we assume that (a) and (b) hold but that $\{F_k\}_1^\infty$ does not converge i.d. From Lemma 11.3 it readily follows that $\{F_k\}_1^\infty$ is weakly compact under convergence i.d. By combining this with Lemma 11.1 we see that we can select two subsequences $\{k_\nu\}_{\nu=1}^\infty$ and $\{k_\mu\}_{\mu=1}^\infty$ such that

$$\lambda_{k_\nu} \to \lambda, \ (c(k_\nu), \ d(k_\nu)) \Rightarrow (c, d) \text{ and } F_{k_\nu}(x) \to F(x) \text{ i.d. when } \nu \to \infty \tag{11.15}$$

$$\lambda_{k_\mu} \to \lambda', \ (c(k_\mu), d(k_\mu)) \Rightarrow (c', d') \text{ and } F_{k_\mu}(x) \to F'(x) \text{ i.d. when } \mu \to \infty \tag{11.16}$$

where

$$F(x) \not\equiv F'(x). \tag{11.17}$$

From (11.14), (11.15) and (11.16) it follows that the c.f.'s of F and F' are

$$\text{c.f. of } F = \exp\left\{-\frac{t^2}{2}\left(1 - \sum_1^{\infty}(c_\nu^2 + d_\nu^2)\right)\right\} \prod_{\nu=1}^{\infty} \varphi(c_\nu t; \lambda)\, \varphi(d_\nu t; \lambda) \tag{11.18}$$

$$\text{c.f. of } F' = \exp\left\{-\frac{t^2}{2}\left(1 - \sum_1^{\infty}(c_\nu'^2 + d_\nu'^2)\right)\right\} \prod_{\nu=1}^{\infty} \varphi(c_\nu' t; \lambda')\, \varphi(d_\nu' t; \lambda'). \tag{11.19}$$

From (b) it follows that $\sum_1^\infty (c_\nu^2 + d_\nu^2) = \sum_1^\infty (c_\nu'^2 + d_\nu'^2)$. Thus (11.17), (11.18) and (11.19) yield

$$\prod_{\nu=1}^{\infty} \varphi(c_\nu t; \lambda)\, \varphi(d_\nu t; \lambda) \not\equiv \prod_{\nu=1}^{\infty} \varphi(c_\nu' t; \lambda')\, \varphi(d_\nu' t; \lambda')$$

and this contradicts the result in Lemma 11.2. Hence the sufficiency parts of (ii) are proved. Next we show that the class of limit distributions is the one claimed. From a sequence $\{\lambda_k, (c(k), d(k))\}_1^\infty$ we can always select a subsequence for which (11.15) holds. That every limit function is of the type (11.7) thus follows from (11.14). Conversely, that every function of the type (11.7) is a limit function follows from the easily proved fact that every (c, d) for which $\sum_1^\infty (c_\nu^2 + d_\nu^2) \leqslant 1$ is obtainable as $(c(k), d(k)) \Rightarrow (c, d)$, $k \to \infty$, where $\sum_1^\infty (c_\nu^2(k) + d_\nu^2(k)) = 1$.

To prove the necessity parts we assume that $\{F_k(x)\}_1^\infty$ converges to $F(x)$ given by (11.18). Select a subsequence $\{k_\mu\}_{\mu=1}^\infty$ for which (11.16) holds. Then $F'(x)=F(x)$ and from (11.14) it follows that (11.18)$\equiv$(11.19). First, we assume that $(c,d)=(0,0)$. Then also $(c',d')=(0,0)$ because otherwise (11.19) has zeros, while (11.18) is zerofree. Next assume that $(c,d)\neq(0,0)$. The exponential functions are zerofree. By carrying through an analysis of the zeros of (11.18) and (11.19) exactly as in the proof of Lemma 11.2, the necessity can be proved. Thus the lemma is completely proved.

12. *Limit distributions for sample sums*

Let $\{\pi_k\}_1^\infty$, $\pi_k=\{a_{k\nu}\}_{\nu=1}^{Nk}$, be a population sequence and let $\alpha_{k1}\leqslant\alpha_{k2}\leqslant\ldots<\mu_k$ and $\beta_{k1}\geqslant\beta_{k2}\geqslant\ldots>\mu_k$ be the elements in π_k which differ from μ_k. We define $c_{k\nu}=(\alpha_{k\nu}-\mu_k)/D_k$, $d_{k\nu}=(\beta_{k\nu}-\mu_k)/D_k$ (for def. of D_k see (1.1), D_k is assumed >0), $c(\pi_k)=\{c_{k1},c_{k2},\ldots\}$, $d(\pi_k)=\{d_{k1},d_{k2},\ldots\}$ and $\lambda_k=n_kN_k^{-1}$, while $S_{n_k}^*$ is defined in (10.1).

Theorem 12.1. *If* (10.2) *holds and* $D_k>0$ *for all* k, *then:*

(i) *The possible limit distributions for* $S_{n_k}^*$ *when* $k\to\infty$ *are those given by* (11.7).

(ii) *Necessary and sufficient for* $S_{n_k}^*$ *to converge i.d. to* (11.7) *is*

1. *if* $(c,d)=(0,0)$, *that* $(c(\pi_k),d(\pi_k))\Rightarrow(0,0)$ *when* $k\to\infty$.
2. *if* $(c,d)\neq(0,0)$ *and* $\lambda\neq\frac{1}{2}$ *that*
   (a) $\{\lambda_k\}_1^\infty$ *has at most the two limit points* λ *and* $1-\lambda$.
   (b) *for every subsequence* $\{k_\nu\}_1^\infty$ *for which* $\lim_\nu\lambda_{k_\nu}=\lambda$ *it holds that* $(c(\pi_{k_\nu}),d(\pi_{k_\nu}))\Rightarrow(c,d)$ *and for every subsequence for which* $\lim_\nu\lambda_{k_\nu}=1-\lambda$ *it holds that* $(c(\pi_{k_\nu}),d(\pi_{k_\nu}))\Rightarrow(-d,-c)$.
3. *if* $(c,d)\neq(0,0)$ *and* $\lambda=\frac{1}{2}$ *that*
   (a) $\lambda_k\to\frac{1}{2}$ *when* $k\to\infty$.
   (b) $|(c(\pi_k),d(\pi_k))|\Rightarrow|(c,d)|$.

Remarks. 1. The condition $(c(\pi_k),d(\pi_k))\Rightarrow(0,0)$ or, rather, its equivalent

$$\lim_k D_k^{-1}\max_\nu|a_{k\nu}-\mu_k|=0 \tag{N}$$

which is called Noether's condition (see [17] and [11]) is well-known to be necessary and sufficient for S_{n_k} to converge to a $N(0,1)$-distribution i.d. when (10.2) holds, see [8].

2. We have excluded the cases $\underline{\lim}_k\lambda_k=0$ and $\overline{\lim}_k\lambda_k=1$. That the whole situation changes considerably in these cases can be seen from Theorem 5.1 in [8].

Proof of Theorem 12.1. Let $Y_{k\nu}$ and Z_{n_k} be the random variables defined in § 10. It is easily checked that $\sigma(Z_{n_k})=D_k\sqrt{\lambda_k(1-\lambda_k)}$ and that $(Y_{k\nu}-EY_{k\nu})/\sigma(Z_{n_k})$

has a $\Lambda(x/(a_{k\nu}-\mu_k);\lambda_k)$-distribution. Thus, $G^*_{nk}=F(x;\lambda_k,c(\pi_k),d(\pi_k))$. Now part (ii) follows immediately from Lemmata 10.1 and 11.4. From Lemma 11.4 it also follows that the class of limit distributions cannot exceed the class defined by (11.7). It remains to prove that every pair (c,d) for which $\sum_1^\infty(c_\nu^2+d_\nu^2)\leqslant 1$, is obtainable as a limit $(c(\pi_k),d(\pi_k))\Rightarrow(c,d)$, $k\to\infty$ for some population sequence $\{\pi_k\}_1^\infty$. Let (c,d) be given. We define π_k as the population which contains the elements $c_1,c_2,\ldots,c_k,d_1,d_2,\ldots d_k$, $2^{-1}k(k-1)$ elements $-k^{-1}(1-\sum_1^k(c_\nu^2+d_\nu^2))^{\frac{1}{2}}$, and, finally, $2^{-1}k(k-1)$ elements $k^{-1}((1-\sum_1^k(c_\nu^2+d_\nu^2))^{\frac{1}{2}}-2(k-1)^{-1}\sum_1^k(c_\nu+d_\nu))$. It is easily checked that 1) $\mu_k=0$, 2) $D_k\to 1$ when $k\to\infty$, 3) $c_\nu(\pi_k)=c_\nu+o(1)$ and $d_\nu(\pi_k)=d_\nu+o(1)$ when $k\to\infty$. Thus, $(c(\pi_k),d(\pi_k))\Rightarrow(c,d)$ when $k\to\infty$, and the proof of the theorem is complete.

13. *Convergence of random walks*

π is a finite population and $X_1,X_2,\ldots,X_N$ a r.p. of its elements. Let, as usual, $S_0=0$, $S_n=X_1+\ldots+X_n$, $n=1,2,\ldots,N$. It will be convenient to normalize the random walk $S_0,S_1,\ldots,S_n$ in the following way. By the *normalized random walk* corresponding to a sample of size n we mean the stochastic process on $[0,1]$ for which the sample path at the point $\omega\in\Omega$ is obtained by connecting the points

$$(0,0),\ (n^{-1},(\sigma_\pi\sqrt{n})^{-1}(S_1(\omega)-\mu_\pi)),$$
$$(2n^{-1},(\sigma_\pi\sqrt{n})^{-1}(S_2(\omega)-2\mu_\pi)),\ \ldots,\ (1,(\sigma_\pi\sqrt{n})^{-1}(S_n(\omega)-n\mu_\pi))$$

by straight lines. This process obviously has continuous sample paths and we denote the corresponding pr. measure on $C[0,1]$ by $P(\pi,n)$.

The following condition (L) was introduced in [6] and its relevance is further exhibited by the result in Theorem 3.1 in [8]. Let $\pi_k=\{a_{k1},\ldots,a_{kN_k}\}$ and $1\leqslant n_k\leqslant N_k$, $k=1,2,\ldots$. The pair $\{\pi_k\}_1^\infty$, $\{n_k\}_1^\infty$ is said to satisfy condition (L), for Lindeberg, if for every $\varepsilon>0$ it holds that

$$\lim_{k\to\infty} D_k^{-2}\sum_{|a_{k\nu}-\mu_k|\geqslant\varepsilon D_k\sqrt{\frac{n_k}{N_k}\left(1-\frac{n_k}{N_k}\right)}}(a_{k\nu}-\mu_k)^2=0. \tag{L}$$

For def. of D_k^2 see (1.1). We shall also consider the condition (N) defined in Remark 1 to Theorem 12.1.

Theorem 13.1. *If* $\{\pi_k\}_1^\infty$ *and* $\{n_k\}_1^\infty$ *satisfy*

1. $n_k/N_k\to\lambda$ *when* $k\to\infty$, $0\leqslant\lambda\leqslant 1$.
2. *In case* $\lambda>0$, $\{\pi_k\}_1^\infty$ *fulfills* (N) *and in case* $\lambda=0$, $\{\pi_k\}_1^\infty$, $\{n_k\}_1^\infty$ *fulfills* (L), *then*

$$P(\pi_k,n_k)\Rightarrow W(1,\lambda^{-1})\ \textit{when}\ k\to\infty. \tag{13.1}$$

Proof. It is no loss of generality to assume that $\mu_k=0$ for all k, and we do so thoughout the proof. We first consider the case $n_k=N_k$, $k=1,2,\ldots$. We shall prove the theorem by applying Theorem 9.1 and thus we shall verify that

$$P(\pi_k, N_k)^M \Rightarrow W(1,1)^M \text{ when } k \to \infty, \text{ for every marginal } M \tag{13.2}$$

$$\text{and that the family } \{P(\pi_k, N_k)\}_1^\infty \text{ is tight.} \tag{13.3}$$

We start with (13.2). Let M: $0 < t_1 < t_2 < \dots t_m < 1$ be a given marginal. Λ^M is the covariance matrix corresponding to the normal distribution $W(1,1)^M$. Λ^M is positive definite. Let $n_k^{(i)} = [t_i N_k]$, $i = 0, 1, 2, \dots, m+1$, $t_0 = 0$ and $t_{m+1} = 1$. V_k^M is the random vector

$$V_k^M = (D_k^{-1} S_{n_k^{(1)}}, \dots, D_k^{-1} S_{n_k^{(m)}}).$$

Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ be a real vector. Our aim is to show that

$$V_k^M \to N(0, \Lambda^M) \text{ i.d. when } k \to \infty. \tag{13.4}$$

According to a well-known result by Cramér ([1] p. 105), (13.4) is equivalent to

$$(V_k^M, \alpha) \to N(0, \alpha \Lambda^M \alpha') \text{ i.d. for every } \alpha \neq 0, \text{ when } k \to \infty \tag{13.5}$$

((,) stands for scalar product.) We have

$$(V_k^M, \alpha) = \sum_{\nu=1}^{N_k} \varrho_{k\nu} X_{k\nu}, \tag{13.6}$$

where $\varrho_{k\nu} = D_k^{-1} \sum_{p=s(k,\nu)}^m \alpha_p$ and $s(k,\nu) = i$ when $n_k^{(i-1)} < \nu \leqslant n_k^{(i)}$, $i = 1, 2, \dots, m+1$. Empty summation gives 0. Let $\Omega_k = \{\varrho_{k1}, \dots, \varrho_{kN_k}\}$. Computations yield that $E(V_k^M, \alpha) = 0$ and that

$$\sigma^2((V_k^M, \alpha)) = D_k^2 \sigma_{\Omega_k}^2 = \frac{1}{N_k - 1} \sum_{i=1}^m (n_k^{(i)} - n_k^{(i-1)}) \left(\sum_{p=i}^m \alpha_p \right)^2 - \frac{1}{N_k(N_k - 1)} \left(\sum_{i=1}^m \alpha_i n_k^{(i)} \right)^2. \tag{13.7}$$

From (13.7) we deduce that

$$\lim_{k \to \infty} \sigma^2((V_k^M, \alpha)) = \sum_{i=1}^m (t_i - t_{i-1}) \left(\sum_{p=i}^m \alpha_p \right)^2 - \left(\sum_{i=1}^m t_i \alpha_i \right)^2. \tag{13.8}$$

It is not difficult to verify that the right-hand side in (13.8) can be written $\alpha \Lambda^M \alpha'$. (13.5) will now follow from Theorems 4.1 and 4.2 and Lemma 4.1 in [9] if we prove that the sequence $\{\Omega_k\}_1^\infty$ satisfies any of the conditions in Lemma 4.1 in [9]. We choose to verify (iii). A sequence $\{\pi_k\}_1^\infty$ satisfies this condition if

$$(\lim_{k \to \infty} A_k = \infty) \Rightarrow \lim_{k \to \infty} \sigma_{\pi_k}^{-2} \int_{|x| > A_k \sigma_{\pi_k}} x^2 dF_{\pi_k}^c = 0. \tag{13.9}$$

Condition (13.9) is clearly invariant if all elements in π_k are multiplied with the same nonzero constant. We consider $\{\Omega_k'\}_1^\infty$, where $\Omega_k' = D_k \cdot \Omega_k$, $k = 1, 2, \dots$. We observe that all elements in Ω_k' lie on the interval

$$\left[\min_{1\leqslant s\leqslant m+1}\sum_{p=s}^{m}\alpha_p,\ \max_{1\leqslant s\leqslant m+1}\sum_{p=s}^{m}\alpha_p\right]$$

and that $\overset{2}{\sigma}\Omega_k\to\alpha\Lambda^M\alpha'>0$, because Λ^M is positive definite. Now (13.9) is easily verified for $\{\Omega_k'\}_1^\infty$ and thus for $\{\Omega_k\}_1^\infty$. Hence, (13.5) follows and thus (13.4) is proved when M is a marginal which contains none of the points 0 or 1. The extension of (13.4) to an arbitrary marginal is immediate. Finally, we have that $P(\pi_k, N_k)^M$ is the distribution of $[V_k^M+(Y_k^{(1)},\dots,Y_k^{(m)})](1-N_k^{-1})$, where $Y_k^{(i)}=(t_iN_k-[t_iN_k])X_{k,[t_iN_k]+1}\cdot D_k^{-1}$, $i=1,2,\dots m$. Condition (N) implies that $Y_k^{(i)}\to 0$ in pr. when $k\to\infty$, and hence (13.2) follows from (13.4).

Next we verify (13.3), and this can be done by applying Theorem 9.2. Let δ and Δ be positive numbers,

$$T\in[0,1-\Delta],\quad \lambda_k^{(1)}=[N_kT]N_k^{-1}\quad\text{and}\quad \lambda_k^{(2)}=([N_k(T+\Delta)]+1)N_k^{-1}.$$

For simplicity we write P_k instead of $P(\pi_k, N_k)$. We have

$$P_k(\max_{T\leqslant t\leqslant T+\Delta}|x(T)-x(t)|\geqslant\delta)\leqslant P_k(\max_{T\leqslant t\leqslant\lambda_k^{(2)}}|x(T)-x(\lambda_k^{(1)})+x(\lambda_k^{(1)})-x(t)|\geqslant\delta). \quad (13.10)$$

The measure P_k is concentrated on polygons with corners in time points which are integer multiples of N_k^{-1}. Thus we can continue the inequality (13.10)

$$\leqslant P_k\left(|x(\lambda_k^{(1)})-x(T)|\geqslant\frac{\delta}{2}\right)+P_k\left(\max_{\lambda_k^{(1)}\leqslant t\leqslant\lambda_k^{(2)}}|x(\lambda_k^{(1)})-x(t)|\geqslant\frac{\delta}{2}\right)$$

$$\leqslant 2P_k\left(\max_{\lambda_k^{(1)}\leqslant t\leqslant\lambda_k^{(2)}}|x(\lambda_k^{(1)})-x(t)|\geqslant\frac{\delta}{2}\right). \quad (13.11)$$

From the exchangeability of the X_ν's and the fact that polygons attain their maximum in corners, we conclude that

$$(13.11)\leqslant 2P\left(\max_{\nu\leqslant N_k(\lambda_k^{(2)}-\lambda_k^{(1)})}|S_\nu^{(k)}/(\sqrt{N_k}\sigma_k)|\geqslant\frac{\delta}{2}\right). \quad (13.12)$$

If $\lambda_k^{(2)}-\lambda_k^{(1)}\leqslant\frac{1}{2}$ we get from Lemma 6.2 that

$$(13.12)\leqslant 4P\left(|S_{N_k(\lambda_k^{(2)}-\lambda_k^{(1)})}/(\sqrt{N_k}\sigma_k)|\geqslant\frac{\delta}{4}-\sqrt{2(\lambda_k^{(2)}-\lambda_k^{(1)})}\right). \quad (13.13)$$

As $\lambda_k^{(2)}-\lambda_k^{(1)}\to\Delta$ when $k\to\infty$ we have if k is sufficiently large and if $\Delta<\delta^2/128$

$$(13.13)\leqslant 4P\left(\left|\frac{S_{N_k(\lambda_k^{(2)}-\lambda_k^{(1)})}}{\sigma_k\sqrt{N_k}}\right|\geqslant\frac{\delta}{8}\right). \quad (13.14)$$

As a special case of (13.2) (or of Theorem 12.1) it holds that the quantity in (13.14) converges, when $k\to\infty$, to

$$\frac{4}{\sqrt{2\pi\Delta(1-\Delta)}}\int_{|x|\geqslant\delta/8}\exp\left\{\frac{-x^2}{2\Delta(1-\Delta)}\right\}dx\leqslant 4\cdot\frac{8}{\delta}\sqrt{\frac{2}{\pi}\Delta(1-\Delta)}\exp\left\{-\frac{\delta^2}{128\Delta(1-\Delta)}\right\}$$

Summing up, we have proved: If $\Delta<\delta^2/128$, then

$$\overline{\lim_{k\to\infty}}\sup_{0\leqslant T\leqslant 1-\Delta}P_k\left(\max_{T\leqslant t\leqslant T+\Delta}|x(T)-x(t)|\leqslant\delta\right)$$
$$\leqslant 4\cdot\frac{8}{\delta}\sqrt{\frac{2}{\pi}\Delta(1-\Delta)}\exp\left\{-\frac{\delta^2}{128\Delta(1-\Delta)}\right\}. \tag{13.15}$$

By combining (13.15) with the fact $S_0^{(k)}=0$ we obtain (13.3) as a consequence of Theorem 9.2. Thus, Theorem 13.1 is proved for the case $n_k=N_k$.

It is not difficult to see that the case $0<\lambda\leqslant 1$ is contained in the case $n_k=N_k$. The difference is merely a change of the time scale. We do not carry out the details to verify this.

Finally, the case $\lambda=0$ can be treated quite analogously. We give an indication of the proof and we follow the steps in the above proof. M is the marginal, Λ^M the covariance matrix corresponding to $W(1)^M$, $n_k^{(i)}=[t_i n_k]$, $i=0,1,\ldots,m$. $V_k^M=((\sigma_k\sqrt{n_k})^{-1}S_{n_k^{(1)}},\ldots(\sigma_k\sqrt{n_k})^{-1}S_{n_k^{(m)}})$. In the representation (13.6) of (V_k^M,α) we get that

$$\varrho_{k\nu}=\begin{cases}\dfrac{1}{\sigma_k\sqrt{n_k}}\displaystyle\sum_{p=s(k,\nu)}^{m}\alpha_p, & \nu=1,2,\ldots,n_k^{(m)}\\ 0 & ,\ \nu=n_k^{(m)}+1,\ldots,N_k,\end{cases}$$

where $s(k,\nu)$ is defined as before. Again it is a matter of computation to show that $E(V_k^M,\alpha)=0$ and that $\sigma^2((V_k^M,\alpha))\to\alpha\Lambda^M\alpha'$, when $k\to\infty$. To prove the asymptotic normality of (V_k^M,α) we again apply Theorem 4.1 in [9], but this time we verify directly that $\{\pi_k,\Omega_k\}_1^\infty$ satisfies condition (4.3) in [9]. We consider instead the sequence $\{\pi_k,\Omega_k'\}_1^\infty$, where $\Omega_k'=\{\varrho_{k1}',\ldots,\varrho_{kN_k}'\}=\{\sigma_k\sqrt{n_k}\varrho_{k1},\ldots,\sigma_k\sqrt{n_k}\varrho_{kN_k}\}$. Let

$$\delta_{kij}=\frac{a_{ik}(\varrho_{kj}'-\mu_{\Omega_k'})}{\sqrt{N_k^{-1}D_{\pi_k}^2\cdot D_{\Omega_k'}^2}}.$$

It holds that
$$|\varrho_{kj}'-\mu_{\Omega_k'}|\leqslant 2\sum_{\nu=1}^{m}|\alpha_\nu| \tag{13.16}$$

and
$$D_{\Omega_k'}^2/n_k\to\alpha\Lambda^M\alpha'. \tag{13.17}$$

From (13.16) and (13.17) it follows easily that $\{\pi_k,\Omega_k'\}_1^\infty$ and thus $\{\pi_k,\Omega_k\}_1^\infty$ satisfies condition (4.3) in [9] as soon as $\{\pi_k\}_1^\infty$ satisfies condition (L) and $\overline{\lim}\, n_k/N_k<1$. The rest of the proof runs almost exactly as before. The proof of Theorem 13.1 is thus concluded.

The next theorem, which is a counterpart of the so-called invariance principle, see [3], is an immediate consequence of Theorem 13.1 and Lemma 9.1.

Theorem 13.2. *If $\{\pi_k\}_1^\infty$ and $\{n_k\}_1^\infty$ satisfy the conditions of Theorem 13.1 and if f is a functional on $C[0,1]$, which is continuous a.e. $(W(1,\lambda^{-1}))$, then*

$$P(\pi_k, n_k)\,(f(x)\leqslant\alpha)\rightarrow W(1,\lambda^{-1})\,(f(x)\leqslant\alpha)\ \ i.d.\ \ when\ \ k\rightarrow\infty.$$

Next we give an application of the invariance principle. We shall consider the r.v.'s

$$N_n = \text{the number of positive sums among } S_1, S_2, \ldots, S_n. \tag{13.18}$$

Theorem 13.3. *If $\{\pi_k\}_1^\infty$ and $\{n_k\}_1^\infty$ satisfy the conditions in Theorem 13.1 and if $\mu_k=0$, $k=1,2,\ldots$, then it holds for $0\leqslant\alpha\leqslant 1$, that*

$$\lim_{k\rightarrow\infty} P(N_{n_k}/n_k\leqslant\alpha)=\alpha\lambda+\frac{1}{2}(1-\lambda)$$
$$+\frac{1}{\pi}\left[(1-\alpha\lambda)\ \text{Arcsine}\ \sqrt{\frac{\alpha(1-\lambda)}{1-\alpha\lambda}}-(1-(1-\alpha)\,\lambda)\ \text{Arcsine}\ \sqrt{\frac{(1-\alpha)\,(1-\lambda)}{1-(1-\alpha)\,\lambda}}\right]. \tag{13.19}$$

Proof. We introduce the following functionals on $C[0,1]$, $\psi_s(x)=\mu(t\,|\,0\leqslant t\leqslant s,\ x(t)>0)$, where μ is the Lebesgue measure. We will write ψ instead of ψ_1. $\psi(x;n)=n^{-1}$ (the number of the halfopen intervals $((\nu-1)\,n^{-1}, \nu n^{-1}]$, $\nu=1,2,\ldots n$, on which $x(t)>0$). It holds that

$$P(N_{n_k}/n_k\leqslant\alpha)=P(\pi_k,n_k)\,(\psi(x;n_k)\leqslant\alpha). \tag{13.20}$$

We now proceed by showing that

$$P(\pi_k,n_k)\,(\psi(x;n_k)\leqslant\alpha)\rightarrow W(1,\lambda^{-1})\,(\psi(x)\leqslant\alpha)\ \ \text{i.d.,}\ \ k\rightarrow\infty. \tag{13.21}$$

Then we compute the distribution to the right in (13.21). First we prove that for every $\varepsilon>0$ it holds that

$$P(\pi_k,n_k)\,(|\psi(x;n_k)-\psi(x)|\geqslant\varepsilon)\rightarrow 0\ \ \text{when}\ \ k\rightarrow\infty. \tag{13.22}$$

The set of zeros of $x(t)\in C[0,1]$ is a closed set, and we denote by $I_1(x), I_2(x), \ldots$ the zero-free open intervals, arranged according to decreasing length. Let $E_{\varepsilon,m}=(x\,|\,\sum_{\nu=1}^m \mu(I_\nu(x))>1-\varepsilon)$. If $x\in E_{\varepsilon,m}$ it holds that

$$|\psi(x;n)-\psi(x)|\leqslant 2mn^{-1}+\varepsilon. \tag{13.23}$$

It is well-known (see e.g. Lévy [13] § 15) that the set of zeros of $x(t)$ has measure 0 with $W(1,\lambda^{-1})$ − pr. 1. Thus $W(1,\lambda^{-1})\,(E_{\varepsilon,m})\nearrow 1$ when $m\rightarrow\infty$. Hence we can choose $m(\varepsilon)$ so that $W(1,\lambda^{-1})\,(E_{\varepsilon,m(\varepsilon)})>1-\varepsilon/2$. As $P(\pi_k,n_k)\Rightarrow W(1,\lambda^{-1})$ and $E_{\varepsilon,m(\varepsilon)}$ is open it holds that

$$P(\pi_k,n_k)\,(E_{\varepsilon,m(\varepsilon)})>1-\varepsilon\ \ \text{if}\ \ k\geqslant k_\varepsilon. \tag{13.24}$$

Now (13.23) and (13.24) imply (13.22). By combining (13.22) with the well-known fact that $\psi(x)$ is continuous a.e. $(W(1,\lambda^{-1}))$, (13.21) follows from Theorem 13.2.

Next we compute $W(1,\lambda^{-1})(\psi(x)\leqslant\alpha)$. The result for $\lambda=1$

$$W(1,1)(\psi(x)\leqslant\alpha)=\alpha \text{ for } 0\leqslant\alpha\leqslant 1 \tag{13.25}$$

is well-known and can be obtained as follows. Let π_{2k} be the population which contains k 1's and $k(-1)$'s and let N_{2k}^* = the number of edges on the positive side in the random walk polygon corresponding to a r.p. of the elements in π_{2k}. According to Theorem 3 in § 2 of Chapter III in [7], $N_{2k}^*/2k$ has a uniform distribution over the values $0, 1/k, 2/k, \dots, 1$. Now $N_{2k}^*/2k$ and $N_{2k}/2k$ (N_{2k} is def. in (13.18)) are asymptotically equivalent. Thus (13.25) follows by letting $k\to\infty$.

To treat the case $\lambda<1$ we introduce the functional T on $C[0,1]$, $T(x)=\sup(t\,|\,x(t)=0)$. The distribution of T under $W(\sigma^2,\lambda^{-1})$-measure can be found in Lévy [13], p. 39. It is

$$\frac{d}{dt}W(\sigma^2,\lambda^{-1})(T(x)\leqslant t)=f(t,\lambda)=\frac{\sqrt{1-\lambda}}{\pi(1-\lambda t)\sqrt{t(1-t)}}\quad 0<t<1. \tag{13.26}$$

It is clear that the distribution should be independent of σ^2. Let $0<T_0<1$.

The process "$W(\sigma^2,\lambda^{-1})$ under the condition $T=T_0$" has the following properties. For $t\in[0,T_0]$ it is a tied-down Wiener process with tying point T_0. For $t\in[T_0,1]$ it is independent of what happened on $[0,T_0]$ and it is equally likely either strictly positive or strictly negative. This is realized by considering time reversed on $[0,\lambda^{-1}]$. Reversal of time leaves $W(\sigma^2,\lambda^{-1})$ invariant and makes T independent of the future. By the strong Markov property we get for $0\leqslant\alpha\leqslant 1$

$$\begin{aligned}W(1,\lambda^{-1})(\psi(x)\leqslant\alpha)&=\frac{1}{2}\int_0^1 W(1,T)(\psi_T(x)\leqslant\alpha)f(T,\lambda)\,dT\\&\quad+\frac{1}{2}\int_0^1 W(1,T)(\psi_T(x)\leqslant\alpha-(1-T))f(T,\lambda)\,dT\\&=\frac{1}{2}\int_0^1\min\left(1,\frac{\alpha}{T}\right)f(T,\lambda)\,dT+\frac{1}{2}\int_0^1\max\left(0,\frac{1}{T}(\alpha-(1-T))\right)f(T,\lambda)\,dT\\&=\frac{1}{2}\int_0^\alpha f(T,\lambda)\,dT+\frac{\alpha}{2}\int_\alpha^1\frac{f(T,\lambda)}{T}\,dT+\frac{1}{2}\int_{1-\alpha}^1 f(T,\lambda)\,dT-\frac{1-\alpha}{2}\int_{1-\alpha}^1\frac{f(T,\lambda)}{T}\,dT.\end{aligned} \tag{13.27}$$

The second equality in (13.27) follows from (13.25). It holds

$$f(T,\lambda)=\frac{d}{dT}\left(\frac{2}{\pi}\operatorname{Arcsine}\sqrt{\frac{T(1-\lambda)}{1-\lambda T}}\right) \tag{13.28}$$

$$\frac{f(T,\lambda)}{T}=\frac{d}{dT}\left(-\frac{2\sqrt{1-\lambda}}{\pi}\left[\sqrt{\frac{1-T}{T}}+\frac{\lambda}{\sqrt{1-\lambda}}\operatorname{Arctg}\sqrt{\frac{1-T}{T(1-\lambda)}}\right]\right). \tag{13.29}$$

By using the primitive functions in (13.28) and (13.29) to evaluate the integrals in (13.27) we get the pr. $W(1, \lambda^{-1})\,(\psi(x) \leqslant \alpha)$ on the form

$$\tfrac{1}{2} + \pi^{-1}\,[\text{Arcsine}\,\sqrt{\alpha(1-\lambda)\,(1-\alpha\lambda)^{-1}} - \text{Arcsine}\,\sqrt{(1-\lambda)\,(1-\alpha)\,(1-\lambda(1-\alpha))^{-1}}$$
$$+\,\alpha\lambda\,\text{Arctg}\,\sqrt{(1-\alpha)\,\alpha^{-1}\,(1-\lambda)^{-1}} - (1-\alpha)\,\lambda\,\text{Arctg}\,\sqrt{\alpha(1-\lambda)^{-1}\,(1-\alpha)^{-1}}],$$

which is easily transformed into (13.19). This concludes the proof.

Chapter 5. Asymptotic behaviour of empirical distributions and empirical fractiles

14. *On the empirical distribution*

The *empirical distribution function* $F^*(\omega, t, n)$ corresponding to the sample $X_1, X_2, \ldots, X_n$ of size n from π is the random distribution function (in t) defined by

$$F^*(\omega, t, n) = n^{-1} \sum_{\nu=1}^{n} H(t - X_\nu(\omega)), \;\; \omega \in \Omega, \;\; -\infty < t < \infty, \tag{14.1}$$

where $H(t)$ is 0 for $t<0$ and 1 for $t \geqslant 0$. We will often use the less cumbersome notations $F^*(t, n)$ or $F^*(t)$.

The following two ideas have been central in the study of the behaviour of $F^*(t, n)$ when the X's are independent with common distribution F.

1. (Kolmogorov.) When F is continuous, many problems about $F^*(t, n)$ can be reduced to the case where F is uniform on $[0, 1]$. This reduction is obtained by the transformation $Y_\nu = F(X_\nu)$, $\nu = 1, 2, \ldots n$.

2. (Doob [5] and Donsker [4].) When F is uniform on $[0, 1]$, it holds that $\sqrt{n}\,(F^*(t, n) - t)$, regarded as a stochastic process with time parameter $t \in [0, 1]$, converges to the $W(1, 1)$-process, when $n \to \infty$.

Our treatment of the behaviour of $F^*(t, n)$, when $X_1, \ldots, X_n$ is a sample from a finite population, leans heavily on the above ideas.

First we define some stochastic processes related to $F^*(t, n)$, defined in (14.1). The process $Z(t, n)$ is

$$Z(\omega, t, n) = F^*(\omega, t, n) - F_\pi(t), \;\; \omega \in \Omega, \;\; -\infty < t < \infty. \tag{14.2}$$

The sample paths of a Z-process consist of horizontal line segments and are thus discontinuous. We define the process $Q(t, n)$ as a "continuization" of $Z(t, n)$. For $\omega \in \Omega$ let $(-\infty, \alpha(\omega))$ and $[\beta(\omega), \infty)$ be the infinite intervals of constancy for $Z(\omega, t, n)$. $Q(\omega, t, n)$ is for $t \in [\alpha(\omega), \beta(\omega)]$ the linear interpolation between the left endpoints of constancy intervals for $Z(\omega, t, n)$. $Q(\omega, t, n) = Z(\omega, \alpha(\omega), n)$ for $t < \alpha(\omega)$ and for $t > \beta(\omega)$ $Q(\omega, t, n) = 0$. Finally we introduce a convenient normalization of the Q-process:

$$R(\omega, t, n) = (n^{-1} - N^{-1})^{-\frac{1}{2}}\, Q(\omega, t, n), \;\; \omega \in \Omega, \;\; -\infty < t < \infty. \tag{14.3}$$

The processes introduced possess a symmetry property which is analogous to the duality principle (p. 385). We need a notation. Let $X(t)$ and $Y(t)$ be stochastic processes, $-\infty < t < \infty$. If all their marginal distributions coincide, we write $X(t) =: Y(t)$.

Lemma 14.1. *Let $T(t, n)$ be any of the processes $Z(t, n)$, $Q(t, n)$ or $Z(t, n) - Q(t, n)$. Then it holds that*

$$(n^{-1} - N^{-1})^{-\frac{1}{2}} T(t, n) =: -((N-n)^{-1} - N^{-1})^{-\frac{1}{2}} T(t, N-n), \tag{14.4}$$

where N is the size of π. Especially we have

$$R(t, n) =: -R(t, N-n). \tag{14.5}$$

Proof. We prove the proposition for $T = Z$. Let $\omega \leftrightarrow \omega'$ be the one-one mapping of Ω onto itself, which is reversal of order in the permutations, i.e. $(i_1, i_2, \ldots, i_N) = \omega \leftrightarrow \omega' = (i_N, i_{N-1}, \ldots, i_1)$. Then

$$(n^{-1} - N^{-1})^{-\frac{1}{2}} Z(\omega, t, n) = (n^{-1} - N^{-1})^{-\frac{1}{2}} \left[\frac{1}{n} \sum_{\nu=1}^{n} H(t - X_\nu(\omega)) - \frac{1}{N} \sum_{\nu=1}^{N} H(t - X_\nu(\omega))\right]$$

$$= -(n^{-1} - N^{-1})^{-\frac{1}{2}} \left[\frac{1}{n} \sum_{\nu=n+1}^{N} H(t - X_\nu(\omega)) - \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{\nu=1}^{N} H(t - X_\nu(\omega))\right]$$

$$= -((N-n)^{-1} - N^{-1})^{-\frac{1}{2}} \left[\frac{1}{N-n} \sum_{\nu=1}^{N-n} H(t - X_\nu(\omega')) - \frac{1}{N} \sum_{\nu=1}^{N} H(t - X_\nu(\omega'))\right]. \tag{14.6}$$

The proposition (14.4) now follows from (14.6) and the fact that every point $\omega \in \Omega$ has the same probability.

In the sequel we shall consider sequences of empirical distributions corresponding to samples from π_k, $k = 1, 2, \ldots$. The processes introduced depend on the population π_k and the size n_k of the sample from π_k. We consider that notations like $Z(t, n_k)$, $R(t, n_k)$ etc. are sufficient to indicate this dependence. We introduce two conditions on population sequences. $\{\pi_k\}_1^\infty$ is said to satisfy respectively conditions (A) and (B) if

(A): N_k^{-1} (Maximal number of equal elements in $\pi_k) \to 0$, when $k \to \infty$.

(B): The elements $a_{k\nu}$, $\nu = 1, 2, \ldots, N_k$ of π_k satisfy $0 \leqslant \alpha_{k\nu} \leqslant 1$ and $\lim_k F_k(t) = t$ for $0 \leqslant t \leqslant 1$.

Verbally (A) means that $F_k(t)$ is asymptotically continuous and (B) means that $F_k(t)$ is asymptotically a uniform distribution on [0. 1]. It is easy to see that (B) implies (A) and also that (B) is equivalent to the condition.

(B′): The elements $a_{k\nu}$ of π_k, $\nu = 1, 2, \ldots, N_k$ satisfy $0 \leqslant a_{k\nu} \leqslant 1$ and

$$\lim_k \sup_{0 \leqslant t \leqslant 1} |F_k(t) - t| = 0.$$

The next lemma states that if (A) holds, then the two processes $(n^{-1} - N^{-1})^{-\frac{1}{2}} Z(t, n)$ and $R(t, n)$ are asymptotically "equivalent".

Lemma 14.2. *If* $\{\pi_k\}_1^\infty$ *satisfies* (*A*) *and if* $\lim_k \min(n_k, N_k - n_k) = \infty$, *then it holds for every* $\varepsilon > 0$ *that*

$$\lim_k P((n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} \sup_t |Z(t, n_k) - Q(t, n_k)| \geqslant \varepsilon) = 0. \tag{14.7}$$

Proof. From Lemma 14.1 it follows that it is sufficient to prove the lemma under the assumption $n_k \leqslant N_k/2$. Then (14.7) is equivalent to

$$\lim_k P(\sqrt{n_k} \sup_t |Z(t, n_k) - Q(t, n_k)| \geqslant \varepsilon) = 0. \tag{14.8}$$

$\operatorname{Sup}_t |Z(t, n_k) - Q(t, n_k)|$ equals the maximal jump in the Z-process. Let $\alpha_{k1}, \alpha_{k2}, \ldots, \alpha_{kM_k}$ be the distinct elements in π_k, $p_{k\nu} = P(X_{k1} = \alpha_{k\nu})$ and $T_{k\nu} = n_k(Z(\alpha_{k\nu}, n_k) - Z(\alpha_{k\nu} - 0, n_k))$, $\nu = 1, 2, \ldots, M_k$. $T_{k\nu} + n_k p_{k\nu}$ has the hypergeometric distribution $H(N_k, n_k, p_{k\nu})$. From Theorem 3.1 it follows that $ET_{k\nu}^4$ is not greater than the 4th central moment in the $Bi(n_k, p_{k\nu})$-distribution, which is (see e.g. [2], p. 195)

$$3n_k^2 p_{k\nu}^2 q_{k\nu}^2 + n_k p_{k\nu} q_{k\nu}(1 - 6p_{k\nu} q_{k\nu}), \text{ where } q_{k\nu} = 1 - p_{k\nu}.$$

From Tchebychev's inequality with 4th moments we thus get

$$P(\sqrt{n_k}\,|T_{k\nu}/n_k| \geqslant \varepsilon) \leqslant \varepsilon^{-4}(3p_{k\nu}^2 + p_{k\nu} \cdot n_k^{-1}).$$

Thus

$$\begin{aligned} &P(\sqrt{n_k} \sup_t |Z(t, n_k) - Q(t, n_k)| \geqslant \varepsilon) \\ &\leqslant \sum_{\nu=1}^{M_k} P(\sqrt{n_k}\,|T_{k\nu}/n_k| \geqslant \varepsilon) \leqslant \varepsilon^{-4}\left(3 \sum_{\nu=1}^{M_k} p_{k\nu}^2 + n_k^{-1} \sum_{\nu=1}^{M_k} p_{k\nu}\right) \\ &\leqslant \varepsilon^{-4}(3 \max_\nu p_{k\nu} + n_k^{-1}) \to 0, \text{ when } k \to \infty, \end{aligned}$$

because condition (A) states that $\lim_k \max_\nu p_{k\nu} = 0$. Thus the lemma is proved.

If π has all its elements on the interval $[0, 1]$ the significant parts of the processes Z, Q and R are those for which $t \in [0, 1]$. The process $R(t, n)$ has continuous sample paths and we can, according to § 8, identify its restriction to $t \in [0, 1]$, with a measure on $C[0, 1]$. We denote this measure by $E(\pi, n)$.

Theorem 14.1. *If* $\{\pi_k\}_1^\infty$ *satisfies condition* (*B*) *and if* $\lim_k \min(n_k, N_k - n_k) = \infty$, *then*

$$E(\pi_k, n_k) \Rightarrow W(1, 1) \text{ when } k \to \infty.$$

Proof. We prove the theorem by applying Theorem 9.1, and first we prove marginal convergence

$$E(\pi_k, n_k)^M \Rightarrow W(1, 1)^M \text{ for every marginal } M, \text{ when } k \to \infty. \tag{14.9}$$

Let M: $0 < t_1 < t_2 < \ldots < t_m < 1$ be a given marginal. We define the random vector $Y_{k\nu} = (Y_{k\nu}^{(1)}, Y_{k\nu}^{(2)}, \ldots, Y_{k\nu}^{(m)})$ by $Y_{k\nu}^{(i)} = 1$ if $X_{k\nu} \leqslant t_i$ and $Y_{k\nu}^{(i)} = 0$ if $X_{k\nu} > t_i$, $i = 1, 2, \ldots, m$. Then the distribution $E(\pi_k, n_k)^M$ equals the distribution of

$$(n_k^{-1}-N_k^{-1})^{-\frac{1}{2}}\,n_k^{-1}\sum_{\nu=1}^{n_k}(Y_{k\nu}-EY_{k\nu})$$
$$+(n_k^{-1}-N_k^{-1})^{-\frac{1}{2}}\,((Q-Z)(t_1,n_k),\,\ldots,(Q-Z)(t_m,n_k)). \tag{14.10}$$

From Lemma 14.2 it follows that the last term in (14.10) tends to 0 in pr. when $k\to\infty$. Thus it suffices to study the random vector

$$n_k^{-1}(n_k^{-1}-N_k^{-1})^{-\frac{1}{2}}\sum_{\nu=1}^{n_k}(Y_{k\nu}-EY_{k\nu}). \tag{14.11}$$

$Y_{k1}, Y_{k2}, \ldots$ is a sample from the m-dimensional population Ω_k, in which the ith component contains $N_kF_k(t_i)$ 1's and $N_k(1-F_k(t_i))$ 0's, $i=1,2,\ldots,m$. The covariance matrix $[\sigma_{ij}^{(k)}]$ of the vector (14.11) is (cf. Theorem 1.1)

$$\sigma_{ij}^{(k)}=N_k(N_k-1)^{-1}F(t_i)\,(1-F(t_j)),\;\;1\leqslant i\leqslant j\leqslant m.$$

From condition (B) it follows that

$$\lim_k\,[\sigma_{ij}^{(k)}]=\Lambda^M, \tag{14.12}$$

where Λ^M is the covariance matrix corresponding to the normal distribution $W(1,1)^M$. To prove (14.9) we apply the result in remark 3.2 in [8]. From the fact that Λ^M is non-singular and from (14.12) we conclude that condition (3.16) in [8] is fulfilled. We also have to verify that every component $\{\Omega_k^{(i)}\}_{k=1}^{\infty}$, $i=1,2,\ldots,m$ of $\{\Omega_k\}_1^{\infty}$ satisfies condition (L) with the sequence $\{n_k\}_1^{\infty}$ as soon as $\lim_k \min(n_k, N_k-n_k)=\infty$. According to Th. 4.2 and Lemma 4.1 in [9] (13.9) is sufficient for that. That $\{\Omega_k^{(i)}\}_1^{\infty}$ satisfies (13.9) is clear from the fact that $F_{\Omega_k^{(i)}}(t)$ is a 0–1 distribution which tends to a distribution with positive variance when $k\to\infty$. Thus (14.9) is proved for every marginal M not containing $t=0$ or 1. From the easily proved facts that $R(0,n_k)$ and $R(1,n_k)$ tend to 0 in pr. when $k\to\infty$, it follows that (14.9) holds for every marginal M. Thus convergence of marginal distributions is proved and it remains to verify tightness of the family $\{E(\pi_k,n_k)\}_1^{\infty}$. Here we insert a lemma.

Lemma 14.3. *Let $Z(t,n)$ be the process defined in (14.2). For every $x>0$ it holds that*

$$P(\sup_{T_1\leqslant t\leqslant T_2}|Z(t,n)-Z(T_1,n)|\geqslant x)\leqslant 2P(|Z(T_2,n)-Z(T_1,n)|$$
$$\geqslant x(1-p(T_1,T_2)/(1-p(T_1,T_2)))-\sqrt{2p(T_1,T_2)/(1-p(T_1,T_2))\,n}).$$

where $p(T_1,T_2)=F_\pi(T_2)-F_\pi(T_1)$.

Proof. We fix n and we write $Z(t)$ instead of $Z(t,n)$. The first point t (if any) on $(T_1,T_2]$, where $Z(t)-Z(T_1)$ jumps over the level x or under $-x$, must be one of the jump-points $j_1,j_2,\ldots,j_s$ of $F_\pi(t)$ on $(T_1,T_2]$. Let for $\nu=1,2,\ldots,s$ and $|x_0|\geqslant x$, $A_\nu(x_0)=(|Z(t)-Z(T_1)|<x$ for $T_1<t<j_\nu$ and $Z(j_\nu)-Z(T_1)=x_0)$ and let $A(y)=(|Z(T_2)-Z(T_1)|\geqslant y)$. Then

$$A(y) \supset \bigcup_{|x_0| \geq x} \bigcup_{\nu=1}^{s} A(y) A_\nu (x_0). \tag{14.13}$$

As two events $A_\nu (x_0)$ are disjoint, if both arguments ν and x_0 do not agree, we get from (14.13)

$$P(A(y)) \geq \sum_{|x_0| \geq x} \sum_{\nu=1}^{s} P(A(y) A_\nu (x_0)) = \sum_{|x_0| \geq x} \sum_{\nu=1}^{s} P(A_\nu (x_0)) \cdot P(A(y) \mid A_\nu (x_0)). \tag{14.14}$$

Fix ν and let Y_1, and Y_2 be the number of observations X on the resp. intervals $(T_1, j_\nu]$ and $(j_\nu, T_2]$, i.e.

$$Y_1 = n(F^* (j_\nu) - F^* (T_1)), \; Y_2 = n(F^* (T_2) - F^* (j_\nu)). \tag{14.15}$$

Let $p_1 = F_\pi (j_\nu) - F_\pi (T_1)$ and $p_2 = F_\pi (T_2) - F_\pi (j_\nu)$.

The joint distribution of Y_1 and Y_2 is the hypergeometric $H(N, n, p_1, p_2)$. We use, without formal proof, the following conditioning property. If A is an event defined in terms of the observations X which satisfy $T_1 < X \leq j_\nu$ and $Y_1 = n_0$ ($=$const.) on A, then the conditional distribution of Y_2 given A equals that of Y_2 given $(Y_1 = n_0)$. The event $A_\nu (x_0)$ is of the type described, with $n_0 = n(x_0 + p_1)$. Thus, the conditional distribution of Y_2, given $A_\nu (x_0)$ is the hypergeometric $H(N - n_0, \; n - n_0, p_2/(1 - p_1))$. We use the suffix $_0$ to indicate quantities under the condition that $A_\nu (x_0)$ occurs. From (14.15) it follows that

$$E_0 (Z(T_2) - Z(T_1)) = E_0 (n^{-1} (Y_1 + Y_2) - (p_1 + p_2)) = x_0 \left(1 - \frac{p_2}{1 - p_1}\right) \tag{14.16}$$

$$\sigma_0^2 (Z(T_2) - Z(T_1)) = \sigma_0^2 \left(\frac{Y_2}{n}\right) = \frac{n - n_0}{n^2} \frac{p_2}{1 - p_1} \left(1 - \frac{p_2}{1 - p_1}\right) \left(1 - \frac{n - n_0}{N - n_0 - 1}\right)$$

$$\leq \frac{p_2}{n(1 - p_1)} \leq \frac{p(T_1, T_2)}{n(1 - p(T_1, T_2))}. \tag{14.17}$$

From Tchebychev's inequality it now follows that if $|E_0 (Z(T_2) - Z(T_1))| > y$, then

$$P_0 (A(y)) \geq P_0 (|Z(T_2) - Z(T_1) - E_0 (Z(T_2) - Z(T_1))|$$

$$\leq |E_0 (Z(T_2) - Z(T_1))| - y) \geq 1 - \frac{\sigma_0^2 (Z(T_2) - Z(T_1))}{(|E_0| - y)^2}. \tag{14.18}$$

By choosing

$$y = x(1 - p(T_1, T_2)/(1 - p(T_1, T_2))) - \sqrt{\frac{2 p(T_1, T_2)}{n(1 - p(T_1, T_2))}}$$

and by inserting (14.16) and (14.17) into (14.18) we get that $P(A(y) \mid A_\nu (x_0)) \geq \frac{1}{2}$. The desired inequality now follows from (14.14) because

$$P(\sup_{T_1 \leqslant t \leqslant T_2} |Z(t) - Z(T_1)| \geqslant x) = \sum_{|x_0| \geqslant x} \sum_{\nu=1}^{s} P(A_\nu(x_0)).$$

Thus Lemma 14.3 is proved, and we continue the proof of Theorem 14.1.

We want to prove that the family $\{E(\pi_k, n_k)\}_1^\infty$ is tight, and we do that by applying Theorem 9.2. Let δ and Δ be positive numbers and $T \in [0, 1-\Delta]$. Then

$$\begin{aligned} E(\pi_k, n_k)&(\max_{T \leqslant t \leqslant T+\Delta} |x(t) - x(T)| \geqslant \delta) \\ &\leqslant P\left((n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} \sup_{T \leqslant t \leqslant T+\Delta} |Z(t, n_k) - Z(T, n_k)| \geqslant \frac{\delta}{2}\right) \\ &+ P\left((n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} \sup_{T \leqslant t \leqslant T+\Delta} |(Z-Q)(t, n_k) - (Z-Q)(T, n_k)| \geqslant \frac{\delta}{2}\right). \end{aligned} \tag{14.19}$$

According to Lemma 14.2 the last term in (14.19) tends to 0 uniformly in T when $k \to \infty$. To estimate the first term we apply Lemma 14.3 and we obtain

$$\begin{aligned} &P\left((n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} \sup_{T \leqslant t \leqslant T+\Delta} |Z(t, n_k) - Z(T, n_k)| \geqslant \frac{\delta}{2}\right) \\ &\leqslant 2P((n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} |Z(T+\Delta, n_k) - Z(T, n_k)| \geqslant \frac{\delta}{2}\left(1 - \frac{p_k(T, T+\Delta)}{1 - p_k(T, T+\Delta)}\right) \\ &- \sqrt{2(1 - n_k/N_k)^{-1} p_k(T, T+\Delta)/(1 - p_k(T, T+\Delta))}. \end{aligned} \tag{14.20}$$

According to Lemma 14.1 we can, and do, assume that $n_k \leqslant N_k/2$. Condition (B) implies that $p_k(T, T+\Delta) = \Delta + r_k(T, \Delta)$, where $r_k(T, \Delta) \to 0$ uniformly in T, when $k \to \infty$. Thus, if $\Delta \leqslant \Delta_0(\delta)$ and if k is sufficiently large, we can continue the inequality (14.20)

$$\leqslant 2P\left((n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} |Z(T+\Delta, n_k) - Z(T, n_k)| \geqslant \frac{\delta}{4}\right). \tag{14.21}$$

The marginal convergence proved earlier implies that for every fixed T it holds that

$$(14.21) \to \frac{2}{\sqrt{2\pi\Delta(1-\Delta)}} \int_{|x| \geqslant \delta/4} e^{-x^2/2\Delta(1-\Delta)}\,dx, \text{ when } k \to \infty. \tag{14.22}$$

We show that the convergence in (14.22) is actually uniform in T. Let

$$\Lambda(T, \Delta, n_k) = (n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} \cdot (Z(T+\Delta, n_k) - Z(T, n_k)).$$

We give an indirect proof and we assume that the convergence in (14.22) is not uniform. Then, by restriction to a subsequence (for simplicity we introduce no new notation), it holds that there is an $\varepsilon > 0$ and a sequence $\{T_k\}_1^\infty$ such that $T_k \to T_0$, when $k \to \infty$ and

$$\left| P\left(\left| \Lambda(T_k, \Delta, n_k) \right| \geqslant \frac{\delta}{4} \right) - (2\pi\Delta(1-\Delta))^{-\frac{1}{2}} \int_{|x| \geqslant \delta/4} e^{-x^2/2\Delta(1-\Delta)}\, dx \right| \geqslant \varepsilon. \quad (14.23)$$

It is no essential restriction to assume that $T_k \searrow T_0$ and we do so. Then we have

$$\Lambda(T_k, \Delta, n_k) = \Lambda(T_0, \Delta, n_k) - \Lambda(T_0, T_k - T_0, n_k) + \Lambda(T_0 + \Delta, T_k - T_0, n_k). \quad (14.24)$$

An estimate with Tchebychev's inequality yields that if $\lambda_k \searrow 0$ then

$$\Lambda(T, \lambda_k, n_k) \to 0 \text{ in pr. uniformly in } T \text{ when } k \to \infty. \quad (14.25)$$

It is now easy to combine (14.22), (14.24) and (14.25) so that they yield a contradiction to (14.23). Thus the uniform convergence in (14.22) is proved. The tightness of the family $\{E(\pi_k, n_k)\}_1^\infty$ is now a consequence of Theorem 9.2, and Theorem 14.1 is thereby completely proved.

We define the following random variables, which are analogues of the well-known Kolmogorov statistics. Let $F^*(t, n)$ be defined according to (14.1). Then

$$D_n^{(1)} = (n^{-1} - N^{-1})^{-\frac{1}{2}} \sup_t \, (F^*(t, n) - F_\pi(t))$$

$$D_n^{(2)} = (n^{-1} - N^{-1})^{-\frac{1}{2}} \sup_t \, |F^*(t, n) - F_\pi(t)|.$$

Theorem 14.2. *If* $\{\pi_k\}_1^\infty$ *satisfies condition* (A) *and if* $\lim_k \min(n_k, N_k - n_k) = \infty$, *then*

$$\lim_{k \to \infty} P(D_{n_k}^{(1)} \leqslant \alpha) = \begin{cases} 1 - e^{-2\alpha^2} & \text{for } \alpha \geqslant 0 \\ 0 & \text{for } \alpha < 0 \end{cases}$$

$$\lim_{k \to \infty} P(D_{n_k}^{(2)} \leqslant \alpha) = \begin{cases} 1 - 2\sum_{k=1}^{\infty} (-1)^k e^{-2k^2\alpha^2} & \text{for } \alpha > 0 \\ 0 & \text{for } \alpha \leqslant 0. \end{cases}$$

Proof. First we assume that $\{\pi_k\}_1^\infty$ satisfies condition (B). We have that

$$D_{n_k}^{(1)} = \sup_t \, ((n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} Z(t, n_k)) = \sup_t \, ((n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} Q(t, n_k)$$
$$+ (n_k^{-1} - N_k^{-1})^{-\frac{1}{2}} (Z(t, n_k) - Q(t, n_k)) = \sup_t \, (R(t, n_k) + r(t, n_k)).$$

Lemma 14.2 states that $\sup_t r(t, n_k) \to 0$ in pr. when $k \to \infty$. Thus the limiting distribution of $D_{n_k}^{(1)}$ equals that of $\sup_t R(t, n_k)$. The functional $\sup_{0 \leqslant t \leqslant 1} x(t)$ is a continuous functional on $C[0, 1]$ and from Theorem 14.1 and Lemma 9.1 it follows that this limiting distribution is the distribution of $\sup_{0 \leqslant t \leqslant 1} x(t)$ under $W(1, 1)$-measure. In the same manner it can be shown that the limiting distribution of $D_{n_k}^{(2)}$ is the distribution of $\sup_{0 \leqslant t \leqslant 1} |x(t)|$ under $W(1, 1)$-measure. These distributions are those claimed in the theorem, see Doob [5].

Next we reduce the case when (A) holds to the case when (B) holds. Let the *T-transform* of the population $\pi=\{a_1, a_2, \ldots, a_N\}$ be the population $T\pi=\{F_\pi(a_1), F_\pi(a_2), \ldots, F_\pi(a_N)\}$. We will use the following two properties of the transformation T.

1. If $\{\pi_k\}_1^\infty$ satisfies (A), then $\{T\pi_k\}_1^\infty$ satisfies (B).
2. The distributions of $D^{(1)}$ and $D^{(2)}$ are invariant under T-transformation of the populations.

1. follows from the facts that $0 \leqslant F_\pi(t) \leqslant 1$ and that distinct elements in π are mapped onto distinct elements in $T\pi$. To prove 2 let $\alpha_1, \alpha_2, \ldots, \alpha_M$ be the distinct elements in π. We define $\psi(t)$ as $\psi(\alpha_\nu)=F_\pi(\alpha_\nu)$, $\nu=1, 2, \ldots, M$. $\psi(t)$ is strictly monotone on $\alpha_1, \alpha_2, \ldots, \alpha_M$ and can be extended to a function on $-\infty<t<\infty$, which is continuous and strictly monotone, which tends to ∞ with t and tends to 0 when $t\to-\infty$. Let ψ^{-1} be the inverse of ψ. For a distribution function F, we define F^ψ by $F^\psi(t)=F(\psi^{-1}(t))$. Then we have $F^*_{T\pi}(\omega, t, n)=F^{*\psi}_\pi(\omega, t, n)$ and $F_{T\pi}(t)=F^\psi_\pi(t)$. Thus

$$F^*_{T\pi}(\omega, t, n)-F^*_{T\pi}(t)=F^*_\pi(\omega, t, n)-F^\psi_\pi(t)=F^*_\pi(\omega, \psi^{-1}(t), n)-F_\pi(\psi^{-1}(t)). \quad (14.26)$$

By taking supremum over t in (14.26), proposition 2 follows. The proof of the theorem is then concluded.

15. *On empirical fractiles*

Let $0<p<1$. The *p-fractile* of a d.f. $F(t)$ is defined as supremum over the t-values for which $F(t)<p$. Analogously, we define the *empirical p-fractile* corresponding to a sample of size n from π as the supremum over the t-values for which $F^*(t, n)<p$, where $F^*(t, n)$ is defined in (14.1). The following theorem gives an analogue of the result on p. 369 in [2].

Theorem 15.1. *Let $Y(p, n_k)$ be the empirical p-fractile in a sample of size n_k from π_k, $k=1, 2, \ldots$. We assume that there is a continuous d.f. $F(t)$ such that*

$$\lim_k \sqrt{n_k} \sup_t |F_{\pi_k}(t)-F(t)|=0 \quad (15.1)$$

and, furthermore, that $F'(t)$ is continuous and positive in a vicinity of the p-fractile ξ_p of $F(t)$. Then, if $\lim_k n_k=\infty$ and $\overline{\lim}_k n_k/N_k<1$ it holds for every real α that

$$\lim_{k\to\infty} P\left(\frac{F'(\xi_p)(Y(p, n_k)-\xi_p)}{\sqrt{p(1-p)(n_k^{-1}-N_k^{-1})}}\leqslant\alpha\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha} e^{-x^2/2}\,dx. \quad (15.2)$$

Proof. First we assume that $\{\pi_k\}_1^\infty$ also satisfies condition (A), i.e. that $F(t)$ in (15.1) equals t for $0\leqslant t\leqslant 1$. p is fixed and we define $U(n_k)=p-F^*(p, n_k)$ and $V(n_k)=Y(p, n_k)-p$. We shall show that for every $\varepsilon>0$ it holds that

$$\lim_k P(\sqrt{n_k}\,|U(n_k)-V(n_k)|\geqslant\varepsilon)=0. \quad (15.3)$$

Let $A_k(\lambda)=(U(n_k)=\lambda/\sqrt{n_k})$, i.e. the event that among the observations X_{k1},

$X_{k2}, \ldots, X_{kn_k}$ there are exactly $n_k(\lambda) = n_k(p - \lambda/\sqrt{n_k})$ which are $\leqslant p$. We assume that $A_k(\lambda)$ is non-empty and we introduce the conditioned process

$$G^*(t, n_k; \lambda) = F^*(t, n_k) \mid A_k(\lambda), \quad p \leqslant t \leqslant 1. \tag{15.4}$$

Simple conditioning arguments yield that $G^*(t, n_k; \lambda)$ can be described as follows: $G^*(t, n_k; \lambda) = p - \lambda/\sqrt{n_k} + n_k^{-1} \cdot$ (The number of observations $\leqslant t$ in a sample of size $n_k - n_k(\lambda)$ from $\pi_k(\lambda)$), where $\pi_k(\lambda)$ is the population of size $N_k(1 - F_k(p))$ with d.f. $G_k(t) = (F_k(t) - F_k(p))/(1 - F_k(p))$, $p \leqslant t \leqslant 1$. Thus, we get

$$\begin{aligned} EG^*(t, n_k; \lambda) &= p - \frac{\lambda}{\sqrt{n_k}} + \left(1 - \frac{n_k(\lambda)}{n_k}\right) \frac{F_k(t) - F_k(p)}{1 - F_k(p)} \\ &= p - \frac{\lambda}{\sqrt{n_k}} + \left(1 - p + \frac{\lambda}{\sqrt{n_k}}\right)\left(\frac{t-p}{1-p} + \frac{R_k^{(1)}(t)}{\sqrt{n_k}}\right) \end{aligned} \tag{15.5}$$

$$\begin{aligned} \sigma^2(G^*(t, n_k; \lambda)) &= \frac{n_k - n_k(\lambda)}{n_k^2}\left(1 - \frac{n_k - n_k(\lambda)}{N_k(1 - F_k(p)) - 1}\right) \\ &\quad \times \frac{F_k(t) - F_k(p)}{1 - F_k(p)} \cdot \frac{1 - F_k(t)}{1 - F_k(p)} \leqslant \frac{1}{n_k}\left(\frac{t-p}{1-p} + \frac{R_k^{(2)}(t)}{\sqrt{n_k}}\right), \end{aligned} \tag{15.6}$$

where $R_k^{(1)}(t)$ and $R_k^{(2)}(t)$ tend to 0 uniformly in t when $k \to \infty$ (cf. (15.1)). Let $0 < \varepsilon < 1$. From (15.5), (15.6), and Tchebychev's inequality we obtain that, when $\lambda > 0$, it holds

$$\begin{aligned} P\left(G^*\left(p + (1-\varepsilon)\frac{\lambda}{\sqrt{n_k}}, n_k; \lambda\right) < p\right) &\geqslant 1 - \frac{1}{n_k}\left(\frac{(1-\varepsilon)\lambda}{(1-p)\sqrt{n_k}} + \frac{R_k^{(2)}}{\sqrt{n_k}}\right) \\ &\times \left[\frac{\lambda}{\sqrt{n_k}} - \left(1 - p + \frac{\lambda}{\sqrt{n_k}}\right)\left(\frac{(1-\varepsilon)\lambda}{(1-p)\sqrt{n_k}} + \frac{R_k^{(1)}}{\sqrt{n_k}}\right)\right]^{-2} \to 1 \text{ when } k \to \infty. \end{aligned} \tag{15.7}$$

In a similar manner it can be proved that

$$P\left(G^*\left(p + (1+\varepsilon)\frac{\lambda}{\sqrt{n_k}}, n_k; \lambda\right) > p\right) \to 1 \text{ when } k \to \infty. \tag{15.8}$$

We observe that, when ε is fixed, then the convergence in (15.7) and (15.8) is uniform in λ on every interval $0 < \lambda \leqslant \Lambda < \infty$. We now claim that for every $\varepsilon > 0$ it holds that

$$\lim_k P(\sqrt{n_k} \mid U(n_k) - V(n_k) \mid \geqslant \varepsilon \mid A_k(\lambda)) = 0 \tag{15.9}$$

and that the convergence in (15.9) is uniform in λ on every interval $|\lambda| \leqslant \Lambda < \infty$.

To prove (15.9) we observe that (15.7) and (15.8) state that when $\lambda \geqslant 0$ it holds for every fixed $\varepsilon > 0$ that, with a probability tending to 1 uniformly on $0 \leqslant \lambda \leqslant \Lambda < \infty$, the process $G^*(t, n_k; \lambda)$ crosses the level p on the interval

$$[p+(1-\varepsilon)\lambda/\sqrt{n_k},\ p+(1+\varepsilon)\lambda/\sqrt{n_k}].$$

This proposition is equivalent to (15.9), which is thus proved when $\lambda \geqslant 0$. The case $\lambda < 0$ can be treated in the same manner by considering instead the process $G^*(t, n_k; \lambda)$ for $0 \leqslant t \leqslant p$. Thus we regard (15.9) proved and next we prove (15.3). We have

$$P(\sqrt{n_k}\,|U(n_k)-V(n_k)| \geqslant \varepsilon) = \sum_{\lambda} P(\sqrt{n_k}\,|U(n_k)-V(n_k)| \geqslant \varepsilon \,|\, A_k(\lambda))\,P(A_k(\lambda)). \quad (15.10)$$

Now (15.3) follows from (15.10), (15.9) and the formula

$$\varliminf_{k\to\infty} \sum_{|\lambda|\leqslant\Lambda} P(A_k(\lambda)) \geqslant 1-\psi(\Lambda), \quad (15.11)$$

where $\psi(\Lambda)\to 0$ when $\Lambda\to\infty$. To prove (15.11) we consider

$$\frac{U(n_k)}{\sqrt{p(1-p)(n_k^{-1}-N_k^{-1})}} = \frac{F_k(p)-F^*(p,n_k)}{\sqrt{p(1-p)(n_k^{-1}-N_k^{-1})}} + \frac{p-F_k(p)}{\sqrt{p(1-p)(n_k^{-1}-N_k^{-1})}} \quad (15.12)$$

The last term in (15.12) tends to 0 when $k\to\infty$ (cf. (15.1) and the assumption that $\overline{\lim}\, n_k/N_k < 1$) and from the marginal convergence proved in Theorem 14.1 it follows that

$$[p(1-p)(n_k^{-1}-N_k^{-1})]^{-\frac{1}{2}}\, U(n_k) \to N(0,1) \text{ i.d., } k\to\infty. \quad (15.13)$$

Now (15.13) and (15.12) easily yield (15.11) and thus (15.3) is proved. From (15.3) it follows that (15.13) still holds if $U(n_k)$ is changed into $V(n_k)$. Thus the theorem is proved for the case $F(t)=t$, $0\leqslant t\leqslant 1$.

We treat the general case with the aid of the T-transform introduced in § 14. The following two properties of the T-transform are easily proved.

3. If $\{\pi_k\}_1^\infty$ satisfies (15.1), then $\{T\pi_k\}_1^\infty$ satisfies (15.1) with $F(t)=t$ for $0\leqslant t\leqslant 1$.

4. Let $Y(p,n)$ and $Y_T(p,n)$ be the empirical p-fractiles corresponding to samples of size n from respectively π and $T\pi$. Then $Y_T(p,n)$ and $F_\pi(Y(p,n))$ have the same distribution.

It is well-known that if $(X_k-\mu)/\sigma_k \to N(0,1)$ i.d. and $\sigma_k\to 0$ when $k\to\infty$, then $(g(X_k)-g(\mu))/g'(\mu)\sigma_k \to N(0,1)$ i.d., provided that $g'(\mu)\neq 0$. By combining this with 3, 4 and the above result for $F(t)=t$ we can conclude that

$$\frac{F'(\xi_p)\,[F^{-1}(F_k(Y(p,n_k)))-F^{-1}(p)]}{\sqrt{p(1-p)(n_k^{-1}-N_k^{-1})}} \to N(0,1) \text{ i.d.} \quad (15.14)$$

F^{-1} stands for the inverse of F, and it is well-defined, at least in a vicinity of ξ_p. The theorem now follows from (15.14) and the proposition

$$\sqrt{n_k}\,[F^{-1}(F_k(Y(p,n_k)))-Y(p,n_k)] \to 0 \text{ in pr. when } k\to\infty \quad (15.15)$$

(15.15) is an easy consequence of the following two propositions, the verifications of which we omit,

(a) $Y(p, n_k) \to \xi_p$ in pr., when $k \to \infty$.

(b) $\sqrt{n_k}\, \sup_{t \in I} |F^{-1}(F_k(t)) - t| = 0$ holds for some neighbourhood I of ξ_p.

The proof is concluded.

ACKNOWLEDGEMENTS

The author is deeply grateful to Professor Lennart Carleson for all advice and encouragement offered during the course of this work.

Dr. R. Elston is thanked for kind assistance in correcting the language.

REFERENCES

1. CRAMÉR, H., *Random variables and probability distributions*. Cambridge University Press, London, 1937.
2. CRAMÉR, H., *Mathematical methods of statistics*. University Press, Princeton, 1946.
3. DONSKER, M. D., An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 6 (1951), pp. 1–12.
4. DONSKER, M. D., Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Stat.*, Vol. 23 (1952), pp. 277–281.
5. DOOB, J. L., Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.
6. ERDÖS, P. and RÉNYI, A., On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 4 (1959), pp. 49–61.
7. FELLER, W., *An Introduction to Probability Theory and Its Applications*. Vol. 1, 2nd ed., John Wiley & Sons, 1957.
8. HÁJEK, J., Limiting distributions in simple random sampling from a finite population. *Publ. Math. Inst. Hung. Acad. Sci.*, Vol. 5 (1960), pp. 361–374.
9. HÁJEK, J., Some extensions of the Wald–Wolfowitz–Noether theorem. *Ann. Math. Stat.*, Vol. 32 (1961), pp. 506–523.
10. HÁJEK, J. and RÉNYI, A., Generalization of an inequality of Kolmogorov. *Acta. Math. Acad. Sci. Hung.*, Vol. 6 (1955), pp. 281–283.
11. HOEFFDING, W., A combinatorial central limit theorem. *Ann. Math. Stat.*, Vol. 22 (1951), pp. 558–566.
12. HOEFFDING, W., Probability inequalities for sums of bounded random variables. *Journ. Amer. Stat. Ass.*, Vol. 58 (1963), pp. 13–30.
13. LÉVY, P., *Le mouvement brownien*. Gauthier-Villars, Paris 1954.
14. LOÈVE, M., *Probability Theory*. 2nd ed. D. van Nostrand, Princeton 1960.
15. MADOW, W. G., On the limiting distributions of estimates based on samples from finite universes. *Ann. Math. Stat.*, Vol. 19 (1948), pp. 535–545.
16. MOTOO, M. M., On the Hoeffding's combinatorial central limit theorem. *Ann. Inst. Stat. Math.*, Vol. 8 (1957), pp. 145–154.
17. NOETHER, G. E., On a theorem by Wald and Wolfowitz. *Ann. Math. Stat.*, Vol. 20 (1949), pp. 455–458.
18. PROKHOROV, YU. V., Convergence of random processes and limit theorems in probability theory. *Theory Prob. and Appl.* (*SIAM trans.*), Vol. 1 (1956), pp. 157–214.
19. WALD, A. and WOLFOWITZ, J., Statistical tests based on permutations of the observations. *Ann. Math. Stat.*. Vol. 15 (1944), pp. 358–372.
20. WILKS, S. S., *Mathematical Statistics*. John Wiley & Sons, 1962.

Tryckt den 10 november 1964

Uppsala 1964. Almqvist & Wiksells Boktryckeri AB