Ministry of Education and Science of the Russian Federation
National Research University - Novosibirsk State University (NRU NSU)
Department of Mechanics and Mathematics

TEXTBOOK

“Poissonian and Gaussian approximations in linear spaces”

Main educational program for teaching foreign students in English ”Probability and Statistics” (State Master Standard for preparation direction 010100 - ”Mathematics”)

BASIC COURSE

M.2-E-10 ”Limit Theorems for Sums of Multivariate Random Variables”

qualification (degree) of the graduate Master
Full-time tuition

Author: BORISOV IGOR S., Professor of the Chair of Probability and Statistics NRU NSU

Novosibirsk 2013

Contents

1 Introduction

This textbook involves important topics of the limit theory for sums of independent and weakly dependent random variables. In particular, this textbook provides an overview of modern aspects of this theory in the case of multivariate random variables. Basic elements of Gaussian and Poissonian approximations (including some recent results of the author) are presented in the case when the random variables under consideration take their values in linear spaces, in particular, in Banach spaces. We would like specially to note, firstly, the problem of Poissonian approximation of sums of independent random variables with values in Abelian groups, and secondly, the problem of constructing complete asymptotic expansions for expectations of functions of sums of independent random variables in both the Poissonian and Gaussian cases. In frame of the textbook we study rates of convergence in the Poisson limit theorem using various probability distances.

A special interest for the reader is investigation of limit behavior of nonlinear functionals of the empirical distribution like U- and V-statistics based on weakly dependent observations. As an example of observations of such a kind, we study the case on moving averages (linear processes) as the original observations. It is worth noting that the limiting random variables for these statistics are represented as some multilinear functionals of infinite dimensional Gaussian random variables.

The textbook consists of 13 chapters. The content of the chapters is as follows.

In Chapter 2, exact inequalities are obtained which connect expectations of some functions of sums of independent random variables taking values in a measurable Abelian group, and those for the accompanying infinitely divisible laws. Some applications to empirical processes are studied.

In Chapter 3, a more general version of the well-known Dobrishin's result connected with an optimal coupling of two random variables is proven. An application to the problem of Poisson approximation in Abelian groups is considered. In particular, an optimal coupling in Poisson approximation of empirical processes is studied.

Chapter 4 is dedicated to deriving upper bounds for the Strassen distance in the invariance principle in Banach spaces under the Poissonian setting when the distributions of the random variables have large atoms at zero.

In Chapter 5, we study limit behavior of χ^2-distance between the distributions of the nth partial sum of independent not necessarily identically distributed Bernoulli random variables and the accompanying Poisson law. As a consequence in the i.i.d. case we make more precise the multiplicative constant in the classical upper bound for the rate of convergence in the Poisson limit theorem.

In Chapter 6, the total variation distance is estimated between distributions of the so-called rescaled empirical process and a Poisson point process which are indexed by all Borel subsets of a bounded Borel set in R^k.

In Chapter 7, under minimal moment conditions, complete asymptotic expansions are obtained for expectations of unbounded functions of a finite family of independent random variables in the Poissonian setting.

In Chapter 8, we study the second term in the asymptotic expansion for the expectations of smooth functions of sums of independent identically distributed random variables; moreover, the order of smallness of the remainder is optimal. The method proposed in the chapter allows us to strengthen the corresponding results of the predecessors, since, to obtain the asymptotic expansions, we enlarge the class of smooth functions under a fixed moment condition.

In Chapter 9, we study optimal connection between smoothness of functions and the corresponding asymptotic expansions for the moments of these functions in the classical one-dimensional central limit theorem (CLT). In particular, under the fixed moment conditions, we prove optimality of some smoothness conditions necessary for asymptotic expansions for expectation of smooth functions in the CLT with optimal bounds of the remainder terms.

In Chapter 10, we prove the central limit theorem for normalized von Mises statistics based on an array of degenerate kernel functions. In the case under consideration the limiting random variable is represented as a multifold stochastic integral constructed by a Gaussian stochastic product-measure.

In Chapter 11, it is studied limit behavior of canonical Von Mises statistics based on samples from a sequence of weakly dependent stationary observations satisfying ψ-mixing condition. The corresponding limit distributions are defined by multiple stochastic integrals of nonrandom functions with respect to nonorthogonal Hilbert noises generated by Gaussian processes with nonorthogonal increments.

In Chapter 12, a functional limit theorem (the invariance principle) is proven for a sequence of normalized U-statistics (i.e., for the so-called U-processes) of arbitrary order with canonical kernels defined on samples of φ-mixing observations of growing size. The corresponding limit distribution is described as that of a polynomial of a sequence of dependent Wiener processes with some known covariance function.

In Chapter 13, we study approximation to the partial sum processes which is based on the stationary sequences of random variables having the structure of the so-called moving averages of independent identically distributed observations. In particular, the rates of convergence both in Donsker's and Strassen's invariance principles are obtained in the case when the limit Gaussian process is a fractional Brownian motion with an arbitrary Hurst parameter.

The results of the textbook are based on fundamental knowledge of the students in Probability Theory, Mathematical Statistics, Functional Analysis, and Theory of Functions. This textbook corresponds to high international scientific level in Probability Theory and Mathematical Statistics.

2 Moment inequalities connected with accompanying Poisson laws in Abelian groups

1. Main definitions and results

Let $X_1, X_2, \ldots$ be independent random variables taking values in a measurable Abelian group $(\mathcal{G}, \mathcal{A})$ with respective distributions $P_1, P_2, \ldots$. If the random variables are identically distributed (the i.i.d. case), then we denote by P their common distribution. We suppose that $\{0\} \in \mathcal{A}$ and the operation " $+$ " is measurable. In other words, a sum of arbitrary random variables in $(\mathcal{G}, \mathcal{A})$ is a random variable, too.

Denote by $Pois(\mu)$ the generalized Poisson distribution with the Lévy measure μ:

$$Pois(\mu) := e^{-\mu(\mathcal{G})} \sum_{k=0}^{\infty} \frac{\mu^{*k}}{k!},$$

where μ^{*k} is the k-fold convolution of a finite measure μ with itself; μ^{*0} is the unit mass concentrated at zero. Under the measurability conditions above the convolution is well defined because we can define the convolution of probability (i.e.,normed finite) measures.

Put $S_n := \Sigma_{i \le n} X_i$. Generalized Poisson distribution with the Lévy measure $\mu := \sum_{i \le n} P_i$ is called the accompanying Poisson law for S_n (for example, see Araujo and Giné, 1980). We will denote by τ_μ a random variable having this distribution.

The main goal of the chapter is to obtain sharp moment inequalities for some measurable functions of S_n via the analogous moments of the accompanying Poisson law. Results of such a kind are connected with the Kolmogorov problem of approximation of the sum distributions by infinitely divisible laws as well as with an improvement of the classical probability inequalities for the sums.

For an arbitrary measurable function f satisfying the condition $\mathbf{E}|f(\tau_\mu)| < \infty$, introduce the following notations:

$$\phi(k) := \mathbf{E}f(S_k), \quad \phi_{m,z}(k) := \mathbf{E}f(S_{m,k} + z), \tag{1}$$

where $S_{m,k} := \sum_{i \le k} X_{m,i}$, $S_{m,0} = S_0 = 0$, and $\{X_{m,i};\, i \ge 1\}$ are independent copies of the random variable X_m. We assume that all the sequences $\{X_i\}$, $\{X_{1,i}\}$, $\{X_{2,i}\}, \ldots$ are independent. Note that, under the moment condition above, the functions $\phi(k)$ exist as well as the functions $\phi_{m,z}(k)$ are well defined at least for almost all z with respect to the distribution of $S_{j,k}$ for each $j \ne m$ and integer $k \ge 0$ (for details, see Section 3).

We say that a function $g(k)$ is convex if the difference $\Delta g(k) := g(k+1) - g(k)$ is nondecreasing.

Theorem 1. *Let one of the following two conditions be fulfilled:*

a) *the random variables $\{X_i\}$ are identically distributed and $\phi(k)$ is a convex function;*

b) *for all z and m, all the functions $\phi_{m,z}(k)$ are convex.*

Then, for each n,

$$\mathbf{E}f(S_n) \le \mathbf{E}f(\tau_\mu). \tag{2}$$

For the initial random variables which are nondegenerate at zero, let $\{X_i^0\}$ be independent random variables with respective distributions

$$P_i^0 := \mathcal{L}(X_i | X_i \neq 0).$$

For this sequence we introduce the notations S_k^0, $S_{m,k}^0$, $\phi^0(k)$, and $\phi^0_{m,z}(k)$ as above.

Proposition. *Convexity of the functions $\phi^0(k)$ or $\phi^0_{m,z}(k)$ implies convexity of the functions $\phi(k)$ or $\phi_{m,z}(k)$ respectively. The converse implication is false.*

R e m a r k 1. If the functions in the conditions of the above two assertions are concave then inequality (2) is changed to the opposite. It follows from the well-known connection between convex and concave functions.

A simple sufficient condition for the functions $\phi(k)$ and $\phi_{m,z}(k)$ as well as $\phi^0(k)$ and $\phi^0_{m,z}(k)$ to be convex is as follows:

For all $x \in \mathcal{G}$ and all $z, h \in \bigcup_{i \leq n} suppX_i$ the function f satisfies the inequality

$$f(x+h) - f(x) \leq f(x+h+z) - f(x+z), \tag{3}$$

where $suppX_i$ denotes a measurable subset such that $X_i \in suppX_i$ with probability 1.

For example, in the i.i.d. case, the convexity (say, of $\phi(k)$) easily follows from (3):

$$\phi(k+1) - \phi(k) \leq \mathbf{E}(f(S_{k+2}) - f(S_k + X_{k+2})) = \phi(k+2) - \phi(k+1).$$

For the Banach-space-valued summands the following result is valid.

Theorem 2. *Let $\mathcal{G}$ be a separable Banach space. Suppose that at least one of the following two conditions is fulfilled*:

1) *the function f is continuously differentiable in Fréchet sense* (*i.e., $f'(x)[h]$ is continuous in x for each fixed h*) *, and, for each $x \in \mathcal{G}$ and every $z, h \in \bigcup_{i \leq n} suppX_i$,*

$$f'(x)[h] \leq f'(x+z)[h]; \tag{4}$$

2) $\mathbf{E}X_k = 0$ *for all k, f is twice continuously differentiable in Fréchet sense, and $f''(x)[h,h]$ is convex in x for each fixed $h \in \bigcup_{i \leq n} suppX_i$.*

Then all the functions in the conditions of Theorem 1 and in Proposition are convex.

Corollary 1. *If $X_i \geq 0$ a.s. and f is an arbitrary convex function on $[0, \infty)$, then inequality* (3) *is true. Moreover, if X_i are random vectors in R^k, $k \geq 2$,* (*as well as in the Hilbert space l_2*) *with nonnegative coordinates, then the function $f(x) := \|x\|^{2+\alpha}$, where $\|\cdot\|$ is the corresponding Euclidean norm and $\alpha \geq 0$, satisfies inequalities* (3) *and* (4)*. For the mean zero Hilbert-space-valued summands, the function $f(x) := \|x\|^\beta$, where $\beta = 2, 4$ or $\beta \geq 6$, satisfies condition* 2) *of Theorem* 2*. Therefore, in these cases, inequality* (2) *holds under the additional necessary restriction $\mathbf{E}|f(\tau_\mu)| < \infty$.*

R e m a r k 2. In the multivariate case, conditions (3) and (4) are slightly stronger than convexity. In particular, in general, the Euclidean norm does not satisfy these conditions.

R e m a r k 3. There exist functions $f(x)$ which do not satisfy the conditions of Theorem 2 but the corresponding functions in Theorem 1 and Proposition are convex. For example, in the i.i.d. one-dimensional case, we consider the function $f(x) := x^5$ and the centered summands $\{X_i\}$. It is clear that the conditions of Theorem 2 are not fulfilled. In this case we have

$$\phi(k) = \mathbf{E}(\sum_{i=1}^{k} X_i)^5 = k\mathbf{E}X_1^5 + 10k(k-1)\mathbf{E}X_1^3\mathbf{E}X_1^2.$$

Thus, if $\mathbf{E}X_1^3 \geq 0$, then the function $\phi(k)$ (as well as the function $\phi^0(k)$) is convex, otherwise it is concave. In other words, in this case we have various inequality signs in (2) depending on positivity or negativity of the third moment of the summands.

Given a finite measure μ on $(\mathcal{G}, \mathcal{A})$ satisfying the condition $\mu(\{0\}) = 0$, we denote by $\phi_\mu(k)$ the function $\phi(k)$ in (1) computed in the i.i.d. case for the summand distribution $\mu(\cdot)/\mu(\mathcal{G})$. Exactness of inequality (2) is characterized by the following result:

Theorem 3. *In the i.i.d. case, let the function $\phi_\mu(k)$ be convex. Then*

$$\sup_{n,\,P} \mathbf{E}f(S_n) = \mathbf{E}f(\tau_\mu) \tag{5}$$

whenever the expectation on the right-hand side of (5) *is well defined, where $\mathcal{L}(\tau_\mu) = Pois(\mu)$ and the supremum is taken over all n and P such that $nP(A \setminus \{0\}) = \mu(A)$ for all $A \in \mathcal{A}$.*

R e m a r k 4. Taking inequality (2) into account we can easily reformulate Theorem 3 for the non-i.i.d. case. Perhaps, for the first time the idea of employing generalized Poisson distributions for constructing upper bounds for moments of the sums was proposed by Prokhorov (1960, 1962). In particular, relations (2) and (5) were obtained by Prokhorov (1962) for the functions $f(x) := x^{2m}$ (m is an arbitrary natural) and $f(x) := \mathrm{ch}(tx)$, $t \in \mathrm{R}$, and for one-dimensional symmetric $\{X_i\}$. Moreover, in the case of mean zero one-dimensional summands, these relations for the functions $f(x) := \exp(hx)$, $h \geq 0$, can be easily deduced from Prokhorov (1960) (see also Pinelis and Utev, 1989).

The most general result in this direction was obtained by Utev (1985) which, in fact, rediscovered and essentially employed some results of Cox and Kemperman (1983) regarding lower bounds for moments of sums of independent centered random variables. Under condition 2) of Theorem 2 he proved extremal equality (5) for nonnegative functions $f(x)$ having an exponential majorant. Moreover, he required some additional unnecessary restrictions on the sample Banach space. In our opinion , the corresponding proof proposed in this chapter, is simpler than that of Utev and need no additional restrictions on $f(x)$ and the sample space.

Relations like (2) and (5) can be also applied for obtaining sharp moment and tail probability inequalities for sums of independent random variables (for details, see Kemperman, 1972; Pinelis and Utev, 1985, 1989; Utev, 1984, 1985; Ibragimov and Sharakhmetov, 1997, 2001).

The above results deal with some type of convexity. However, we can obtain moment inequalities close to those mentioned above without any convexity conditions.

Theorem 4. *In the i.i.d. case, for every nonnegative measurable function f, the following inequality holds:*

$$\mathbf{E}f(S_n) \leq \frac{1}{1-p}\mathbf{E}f(\tau_\mu), \tag{6}$$

where $p := \mathbf{Pr}(X_1 \neq 0)$.

In the non-i.i.d. case, the factor $(1-p)^{-1}$ *in* (6) *should be replaced by* $\exp(\sum_{i \le n} p_i)$, *where* $p_i := \mathbf{Pr}(X_i \neq 0)$.

It is clear that inequality (6) provides a sufficiently good upper bound under the so-called Poissonian setting when the summand distributions have large atoms at zero (i.e., the probabilities p_i are small enough). Some particular cases of inequality (6) are contained in Araujo and Giné (1980), and in Giné, Mason, and Zaitsev (2001).

2. Applications to empirical processes

In this section we formulate some consequences of the above theorems as well as some new analogous results for empirical processes. For the sake of simplicity we study the empirical processes with one-dimensional time parameter although the results below can be reformulated for empirical processes indexed by subsets of an arbitrary measurable space (moreover, for abstract empirical processes indexed by a family of measurable functions). These results are a basis for the so-called Poissonization method for empirical processes. Sometimes it is more convenient to replace an empirical process under study by the corresponding accompanying Poisson point process having a simpler structure for analysis (for example, independent "increments"). Some versions of this sufficiently popular and very effective method can be found in many papers. In particular, some probability inequalities connecting the distributions of empirical processes (in various settings) and those of the corresponding Poisson processes are contained in Borisov (1983, 1990, 1991), Einmahl (1987), Deheuvels and Mason (1992), Giné, Mason, and Zaitsev (2001), and others.

Introduce the so-called tail (or local) empirical process on the interval $[0, n]$:

$$\nu_n(t) := nF_n(t/n),$$

where $F_n(\cdot)$ is the empirical distribution function (right-continuous version) based on a sample of size n from the (0,1)-uniform distribution. We consider ν_n as a random variable in the space $LS([0, n])$ which is defined as the linear span of the set of all piecewise constant right-continuous functions on $[0, n]$ with finitely many jumps, endowed with the cylinder σ-field. It is easy to verify that the standard Poisson process $\pi(t)$ on $[0, n]$ (with right-continuous paths) has the accompanying Poisson distribution for ν_n in this space.

Theorem 5. *Let* $\Phi(\cdot)$ *be a convex nonnegative functional on* $LS([0, n])$ *which is nondecreasing on the subset of all nonnegative functions with respect to the standard partial order in function spaces. Suppose that, for each function* $x(\cdot) \in LS([0, n])$, *the following relation holds:* $\lim_{m\to\infty} \Phi(x^{(m)}) = \Phi(x)$, *where* $x^{(m)}(t) = x([mt]/m)$, *with* $[\cdot]$ *the integer part of a number. Moreover, if* $\mathbf{E}\Phi(\pi) < \infty$ *then*

$$\mathbf{E}\Phi(\nu_n) \le \mathbf{E}\Phi(\pi). \tag{7}$$

R e m a r k 5. It is well known that if a convex functional defined on a topological linear space (say, on a Banach space) is bounded in a neighborhood of some point, then it is continuous (for example, see Kutateladze, 1995). Thus, if the functional in Theorem 5 is defined, say, on $L_m([0,n],\lambda)$, where λ is a finite measure, and satisfies the local boundedness condition, then the continuity condition connected with the step functions $x^{(m)}(t)$ can be omitted.

In the sequel, in the case of Banach-space-valued random variables, we will consider only continuous convex functionals. For example, the functional $\Phi(x) := \|x\|_m^q \equiv (\int_0^n |x(t)|^m \lambda(dx))^{q/m}$ with arbitrary parameters $m \geq 1$ and $q \geq 1$, where λ is an arbitrary finite measure on $[0,n]$, satisfies the conditions of Theorem 5.

Note that the accompanying Poisson process for the centered empirical process $\nu_n^0(t) := \nu_n(t) - t$, say, in $L_m([0,n],\lambda)$ differs from the corresponding centered Poisson process. This process can be defined as $\pi^0(t) := \pi(t) - \pi(n)t/n$ and, by analogy with the definition of a Brownian bridge, can be called a Poissonian bridge on $[0,n]$. For such processes the second assertion of Theorem 2 can be reformulated as follows:

Corollary 2. *Let $\Phi(x)$ be a functional on $L_m([0,n],\lambda)$ having convex second Fréchet derivative. Then*

$$\mathbf{E}\Phi(\nu_n^0) \leq \mathbf{E}\Phi(\pi^0) \tag{8}$$

whenever the expectation on the right-hand side of (8) *exists.*

As an example of such a functional we can consider $\Phi(x) := \|x\|_m^{mq}$ for any $m \geq 2$ and $q \geq 3$.

If we consider the processes ν_n and π as random elements in $LS([o,\delta n])$, where $\delta < 1$, then the following direct consequence of Theorems 4 and 5 above, and Lemma 1 and Corollary 6 below holds:

Corollary 3. *For every measurable functional Φ on $LS([0,\delta n])$ under the minimal restriction $\mathbf{E}|\Phi(\pi)| < \infty$, the following inequality holds:*

$$\mathbf{E}|\Phi(\nu_n)| \leq \frac{1}{1-\delta}\mathbf{E}|\Phi(\pi)|. \tag{9}$$

Moreover, if $\delta = N/n$, N does not depend on n, and the functional Φ satisfies the conditions of Theorem 5*, then*

$$\sup_n \mathbf{E}\Phi(\nu_n) = \lim_{n\to\infty} \mathbf{E}\Phi(\nu_n) = \mathbf{E}\Phi(\pi). \tag{10}$$

Finally, we formulate some useful moment inequalities which deal with one-dimensional projections of the processes $\nu_n(\cdot)$ and $\pi(\cdot)$. A direct consequence of Corollary 1 is as follows:

Corollary 4. *For every natural n and m, and every $t \geq 0$, the following inequality holds:*

$$\mathbf{E}(\nu_n(t) + x)^m \leq \mathbf{E}(\pi(t) + x)^m,$$

where x is arbitrary for even m, and $x \geq 0$ for odd m.

In the following assertion which complements this inequality, the above-mentioned convexity conditions need not be fulfilled.

Theorem 6. *For every natural n and m, and every $t \geq 0$, the following inequality holds:*

$$|\mathbf{E}(\nu_n(t)+x)^{2m-1}| \leq \mathbf{E}(\pi(t)+x)^{2m-1},$$

where $x \in [-t, 0]$ is arbitrary.

Corollary 5. *Let $f(x)$ be an entire function on $[0, \infty)$, i.e., an analytic function which admits Taylor expansion at all points with a converging power series on the whole positive half-line. Assume that, for a point $x_0 \geq 0$, the kth derivative of this function at x_0 is nonnegative for each $k \geq 2$. Then, for every $t \geq x_0$,*

$$\mathbf{E}f(\nu_n(t)) \leq \mathbf{E}f(\pi(t)).$$

3. Proof of the results

First we formulate two important lemmas which play a key role in proving the above results.

Lemma 1. *In the i.i.d. case, under the above notations, the following relations hold*:

$$Pois(n\mathcal{L}(X_1)) = \mathcal{L}(S_{\pi(n)}), \tag{11}$$

where the standard Poisson process $\pi(\cdot)$ is independent of $\{X_i\}$;

$$\mathcal{L}(S_n) = \mathcal{L}(S^0_{\nu(n,p)}), \quad Pois(n\mathcal{L}(X_1)) = \mathcal{L}(S^0_{\pi(np)}), \tag{12}$$

where $p := \mathbf{Pr}(X_1 \neq 0)$, $\mathcal{L}(\nu(n,p)) = B_{n,p}$ is the binomial distribution with parameters n and p; the pair $(\nu(n,p), \pi(np))$ does not depend on the sequence $\{X^0_i\}$.

The relation (11) is well known. It immediately follows from the above-mentioned definition of a generalized Poisson distribution: the probability law $Pois(\mu)$ may be interpreted as the distribution of $\sum_{i\leq\pi_\mu} Y_i$, where $\{Y_i\}$ are i.i.d. random variables with the common distribution $\mu(\cdot)/\mu(\mathcal{G})$ and π_μ is a Poissonian random variable with parameter $\mu(\mathcal{G})$, which is independent of the sequence $\{Y_i\}$.

The equalities in (12) which are more convenient in studying accuracy of Poisson approximation of the sums, are contained in various forms in many papers (see, for example, Khintchine, 1933, 1936; Le Cam, 1960, 1965; Borovkov, 1988; Borisov, 1993, 1996; and others). Actually, these relations also represent versions of the total probability formula and are easily proven.

Taking into account the representations in Lemma 1 we can reduce the problem to the simplest one-dimensional case when we estimate the analogous moments of the binomial distribution introduced in Remark 1. However, in this case, we can obtain sufficiently exact inequalities for moments of arbitrary functions using the following lemma.

Lemma 2. *For each $p \in (0, 1)$,*

$$\sup_{n,j} \frac{B_{n,p}(j)}{\mathcal{L}(\pi(np))(j)} \leq \frac{1}{1-p}. \tag{13}$$

Proof. For every nonnegative integer $j \le n$, we have

$$\begin{aligned}\frac{\mathbf{P}(\nu(n,p)=j)}{\mathbf{P}(\pi(np)=j)} &= \frac{n(n-1)\cdots(n-j+1)}{n^j(1-p)^j}(1-p)^n e^{np} = \\ &\exp\left\{n(p+\log(1-p)) - j\log(1-p) + \sum_{i=0}^{j-1}\log\left(1-\frac{i}{n}\right)\right\} \le\end{aligned}$$

$$\exp\Bigg\{-\log(1-p) + n(p+\log(1-p)) - (j-1)\log(1-p)$$

$$+n\int_0^{(j-1)/n}\log(1-x)dx\Bigg\} \le \exp\left\{-\log(1-p) - nH_p\left(\frac{j-1}{n}\right)\right\},$$

where $H_p(x) = -p + x + (1-x)\log((1-x)/(1-p))$. The following properties of H_p are obvious:

$$H_p(1) = 1-p, \quad H_p(p) = 0, \quad \frac{d}{dx}H_p(p) = 0, \quad \frac{d^2}{dx^2}H_p(x) = 1/(1-x),$$

which implies $H_p(x) \ge 0$ if $x \le 1$, i.e., inequality (13) is proven.

R e m a r k 6. Inequality (13) is a part of a more general result in Borisov and Ruzankin (2002). It is worth noting that this upper bound is an estimate for the so-called Radon-Nikodym derivative of a binomial distribution with respect to the accompanying Poisson law. This problem was studied by a number of authors (Le Cam, 1960; Chen, 1975; Barbour, Chen, and Choi, 1995; and others). In particular, under some additional restriction on n and p, a slightly stronger estimate is contained in Le Cam (1960). However, in general, estimate (13) cannot be essentially improved. Under some restrictions on n and p, a lower bound for the left-hand side of (13) has the form $(1-cp)^{-1}$, where c is an absolute positive constant.

Corollary 6. *Let g be an arbitrary function with the restriction $\mathbf{E}|g(\pi(\lambda))| < \infty$ for some λ. Then, for every n and p satisfying the condition $np \le \lambda$, the following inequality holds*:

$$\mathbf{E}|g(\nu(n,p))| \le \frac{e^{\lambda-np}}{1-p}\mathbf{E}|g(\pi(\lambda))|. \tag{14}$$

Moreover,

$$\lim_{n\to\infty,\, np\to\lambda-0}\mathbf{E}g(\nu(n,p)) = \mathbf{E}g(\pi(\lambda)). \tag{15}$$

Proof. Inequality (14) follows from Lemma 2 and the simple estimate

$$\sup_j \frac{\mathbf{P}(\pi(np)=j)}{\mathbf{P}(\pi(\lambda)=j)} \le e^{\lambda-np}.$$

Relation (15) follows from the classical Poisson limit theorem and inequality (14) which provides fulfillment of the uniform integrability condition. The corollary is proven.

R e m a r k 7. Inequality (6) in Theorem 4 immediately follows from Corollary 6 and representations (12). In the case $n = 1$ in Lemma 2 there exists a slightly stronger upper bound for the Radon–Nikodym derivative. It is easy to see that, in this case, the right-hand side of (13) can be replaced by e^p. In the non-i.i.d. case evaluation of the moment $\mathbf{E}f(S_n)$ can be reduced to that for a new function of n independent Bernoulli random variables $\nu_1(1,p),\dots,\nu_n(1,p)$ (for details, see the proof of Theorem 1 below). In this case, the approximating moment is calculated by independent Poisson random variables $\pi_1(p),\dots,\pi_n(p)$ with the same parameter p. Thus, the corresponding upper bound for the Radon–Nikodym derivative (as well as the corresponding factor on the right-hand side of (6)) equals $\exp(\sum_{i\le n} p_i)$. However, in the special case when $S_n = \sum_{i\le n}\nu_i(1,p)$, there exist better upper bounds for this derivative. For example, in this case we can replace the factor $\exp(\sum_{i\le n} p_i)$ by $(1-\tilde{p})^{-2}$, where $\tilde{p} = \max\{p_i;\ i \le n\}$ (see Barbour, Chen, and Choi, 1995; Borisov and Ruzankin, 2002).

It is worth noting that, under the minimal moment condition above, we cannot replace the one-sided double limit in (15) by the classical double limit as well as the condition $np \le \lambda$ in (14) cannot be omitted. For example, the function $g(k) = (1 \vee (k-2))!\lambda^{-k}$ satisfies the above-mentioned moment condition, however it is easy to prove the relation

$$\lim \sup_{n\to\infty,\, np\to\lambda} \mathbf{E}g(\nu(n,p)) = \infty.$$

Proof of Theorem 1. In the i.i.d. case inequality (2) is a simple consequence of relation (11) and the classical Jensen inequality:

$$\mathbf{E}f(\tau_\mu) = \mathbf{E}\phi(\pi(n)) \ge \phi(n) = \mathbf{E}f(S_n).$$

In order to prove inequality (2) in the non-i.i.d. case we introduce the sequence of independent identically distributed random variables $\{\pi_i;\ i \ge 1\}$ having Poisson distribution with parameter 1, which are independent of all the sequences of random variables introduced in (1) (including the initial random variables). Then we can define the random variable τ_μ in the following way:

$$\tau_\mu := \sum_{m=1}^{n} S_{m,\pi_m}, \tag{16}$$

where $S_{m,k}$ are defined in (1). The further reasoning is quite analogous to the above. Put $z_1 := \sum_{m=2}^{n} S_{m,\pi_m}$. Using the above arguments, we have

$$\mathbf{E}f(\tau_\mu) = \mathbf{E}\mathbf{E}_{z_1}\phi_{1,z_1}(\pi_1) \ge \mathbf{E}\mathbf{E}_{z_1}\phi_{1,z_1}(1) = \mathbf{E}f(X_1 + z_1),$$

where the symbol $\mathbf{E}_{z_1}$ denotes the conditional expectation given z_1. Now we put $z_2 := X_1 + \sum_{m=3}^{n} S_{m,\pi_m}$. Then, repeating the same calculation, we obtain the estimate

$$\mathbf{E}f(X_1 + z_1) = \mathbf{E}\mathbf{E}_{z_2}\phi_{2,z_2}(\pi_2) \ge \mathbf{E}\mathbf{E}_{z_2}\phi_{2,z_2}(1) = \mathbf{E}f(X_1 + X_2 + \sum_{m=3}^{n} S_{m,\pi_m}).$$

Continuing the calculations in this way, we finally obtain inequality (2). Theorem 1 is proven.

Proof of Proposition. The first assertion is easily verified. Indeed, by Corollary 1 and Lemma 1 (see (12)) we have

$$\phi(k+1)-\phi(k)=\mathbf{E}\phi^0(\nu(k+1,p))-\mathbf{E}\phi^0(\nu(k,p))$$
$$\le \phi(k+2)-\phi(k+1).$$

The analogous inequality holds for the functions $\phi_{m,z}(k)$.

In order to prove the second assertion of this Proposition we consider the subclass of random variables satisfying the conditions $\mathbf{E}X_1^4<\infty$, $\mathbf{P}(X_1=0)\neq 0$, and $\mathbf{E}X_1\neq 0$. Put $f(x):=x^4$. Then

$$\phi(k)=A_k^4(\mathbf{E}X_1)^4+6A_k^3\mathbf{E}X_1^2(\mathbf{E}X_1)^2+3A_k^2(\mathbf{E}X_1^2)^2+4A_k^2\mathbf{E}X_1\mathbf{E}X_1^3+k\mathbf{E}X_1^4, \tag{17}$$

where $A_k^m:=k(k-1)\cdots(k-m+1)$. The second derivative has the form

$$\phi''(k)=Ak^2+Bk+C,$$

where $A:=12(\mathbf{E}X_1)^4$, $B:=36\mathbf{E}X_1^2(\mathbf{E}X_1^2-(\mathbf{E}X_1)^2)$, and

$$C:=22(\mathbf{E}X_1)^4+6(\mathbf{E}X_1^2)^2+8\mathbf{E}X_1\mathbf{E}X_1^3-36\mathbf{E}X_1^2(\mathbf{E}X_1)^2. \tag{18}$$

Because of positivity of A and B, the function $\phi(k)$ in (17) is convex for $C\ge 0$. If $C<0$ and at least $\phi''(2)<0$, then the function $\phi(k)$ replaces concavity with convexity. The same representations with the above comments hold for the function $\phi^0(k)$ (with the replacement of X_1 by X_1^0 and C by C^0 in (17) and (18)).

Consider the case in which the first moment of X_1^0 is positive and the third moment equals zero. It is clear that we can choose the distribution of X_1^0 so that the constant C^0 will be negative with its absolute value large enough. In this case the function $\phi^0(k)$ will be of the mixed type. For example, we can define this distribution as follows: Given a positive constant K, we put $X_1^0=K$ with probability $8/9$, and $X_1^0=-2K$ with probability $1/8$. In this case, $\mathbf{E}X_1^0=6K/9$, $\mathbf{E}(X_1^0)^2=4K/9$, and $C^0<-K^4$.

Since $\mathbf{E}X_1^k=p\mathbf{E}(X_1^0)^k$ for each integer k, given the above-mentioned distribution of X_1^0, we can consider p as a free parameter. Substituting this representation into (18) we conclude that, for sufficiently small p (say, $p\le 0.1$), the constant C will be positive. Proposition is proven.

Proof of Theorem 2. The first assertion is trivial because, under condition 1), from Taylor's formula we have

$$f(x+h)-f(x)=\int_0^1 f'(x+th)[h]dt\le\int_0^1 f'(x+z+th)[h]dt$$
$$=f(x+z+h)-f(x+z)$$

for every $x\in\mathcal{G}$ and $z,h\in\bigcup_{i\le n} suppX_i$, that is, inequality (3) is fulfilled.

To prove the second assertion we only need to prove this in the i.i.d. case because, using the arguments in proving Theorem 1 above, we can reduce the problem to the i.i.d. case. It remains to observe that, under condition 2) and given z, the function $f(x+z)$ has convex second derivative with respect to x. So, we prove the assertion in the i.i.d. case. Taking into account continuity in x of the function $f''(x)[h,h]$ for any fixed h and using Taylor's formula, we have

$$f(S_{k+1}) - f(S_k) = f'(S_k)[X_{k+1}] + \int_0^1 (1-t) f''(S_k + tX_{k+1})[X_{k+1}, X_{k+1}]dt. \tag{19}$$

First we average both sides of (19) with respect to the distribution of X_{k+1} and use the fact that, for any centered (in Bochner sense) random variable X and an arbitrary linear continuous functional $l(\cdot)$, the equality $\mathbf{E}l(X) = 0$ holds. Averaging both sides of this identity with respect to the other distributions we then obtain the equality (with more convenient representation of the remainder in (19))

$$\phi(k+1) - \phi(k) = \frac{1}{2}\mathbf{E}f''(S_k + \zeta X_{k+1})[X_{k+1}, X_{k+1}], \tag{20}$$

where ζ is a random variable with the density $2(1-t)$ on the unit interval, which is defined on the main probability space and independent of the sequence $\{X_k\}$ (we may assume here that this space is reach enough). It is worth noting that, because of integrability of the left-hand side of (19), the expectation on the right-hand side of (20) is well defined due to Fubini's theorem. In the i.i.d. case, by the classical Jensen inequality (for the conditional expectation $\mathbf{E}_{\zeta, X_{k+2}}$) we finally obtain the inequality we need:

$$\phi(k+1) - \phi(k) = \frac{1}{2}\mathbf{E}\mathbf{E}_{\zeta, X_{k+2}} f''(S_k + \zeta X_{k+2})[X_{k+2}, X_{k+2}]$$

$$\le \frac{1}{2}\mathbf{E}f''(S_{k+1} + \zeta X_{k+2})[X_{k+2}, X_{k+2}] = \phi(k+2) - \phi(k+1).$$

The proof of convexity of $\phi^0(k)$ and $\phi^0_{m,z}(k)$ is the same because, for the centered initial summands, $\mathbf{E}X^0_k = 0$. The theorem is proven.

Proof of Theorem 3. Put $n > \mu(\mathcal{G})$ and consider the independent random variables $X_k \equiv X_k(n)$ with the following common distribution:

$$P(A \setminus \{0\}) = \mu(A)/n, \quad P(\{0\}) = 1 - \mu(\mathcal{G})/n.$$

Then the corresponding random variables X^0_i have the common distribution $P^0(A) = \mu(A)/\mu(\mathcal{G})$. Therefore, for each n, by Proposition we have the corresponding inequality for the moments under study. It is easy to see that, in this case, the function $\phi^0(k) \equiv \phi_\mu(k)$ does not depend on n, and we can apply Lemma 1 and relation (15). Thus, we have

$$\lim_{n\to\infty} \mathbf{E}f(S_n) = \lim_{n\to\infty} \mathbf{E}\phi_\mu(\nu(n,p))$$

$$= \mathbf{E}\phi_\mu(\pi(\lambda)) = \mathbf{E}f(\tau_\mu),$$

where $p = \mu(\mathcal{G})/n$ and $\lambda = \mu(\mathcal{G})$. The theorem is proven.

Proof of Theorem 4. The claim follows immediately from Lemmas 1 and 2 and Remark 7.

Proof of Theorem 5. Because of the monotonicity and continuity conditions it is sufficient to prove the assertion for any finite-dimensional projection $\nu_n^{(m)}$ and $\pi^{(m)}$ of the processes under consideration. To this end it we consider an arbitrary nonnegative function $\psi(x_1, \ldots, x_k)$ which is convex and increasing in every coordinate x_i. We will study the moment $\mathbf{E}\psi(\nu_n(t_1), \ldots, \nu_n(t_k))$, where $t_i \in [0, n)$ are arbitrary points and $t_i < t_{i+1}$ for every $i < k$. We will also assume that the corresponding Poisson moment exists.

The following so-called Markov property of $\nu_n(\cdot)$ is well known: Given the quantity $\nu_n(t)$ (the number of the sample points to the left of t/n), two new samples constituted by the points to the left and to the right of t/n respectively, are independent and distributed as samples (of the corresponding sizes) from the uniform distributions on $[0, t/n]$ and $[t/n, 1]$ respectively. In other words, given $\nu_n(t_1)$, the increment $\nu_n(t_2) - \nu_n(t_1)$ coincides in distribution with $\nu_N^*((t_2 - t_1)N/n)$, where $N := n - \nu_n(t_1)$, and the process $\nu_n^*(\cdot)$ is an independent copy of $\nu_n(\cdot)$. Thus, taking into account Corollary 1 and convexity and monotonicity of the function $\psi_1(x) := \mathbf{E}\psi(\nu_n(t_1), \ldots, \nu_n(t_{k-1}), \nu_n(t_{k-1}) + x)$, we have

$$\mathbf{E}\psi(\nu_n(t_1), \ldots, \nu_n(t_k)) = \mathbf{E}\mathbf{E}_N\psi_1(\nu_N^*((t_k - t_{k-1})N/n))$$

$$\leq \mathbf{E}\mathbf{E}_N\psi_1(\pi((t_k - t_{k-1})N/n)) \leq \mathbf{E}\psi_1(\pi(t_k) - \pi(t_{k-1})),$$

where $\pi(\cdot)$ is a Poisson process independent of $\nu_n(\cdot)$.

Therefore, we reduced the problem to evaluating the moment of a function of the analogous $(k-1)$-dimensional projection $\nu_n(t_1), \ldots, \nu_n(t_{k-1})$. It remains to observe that the function $\psi_2(x_1, \ldots, x_{k-1}) := \mathbf{E}\psi(x_1, \ldots, x_{k-1}, x_{k-1} + \pi(t_k) - \pi(t_{k-1}))$ is convex and monotone, too. In other words, to prove the assertion we may use induction on k. The theorem is proven.

Proof of Theorem 6. It is clear that, under the above notations, we deal with the random variable $\nu(n, p)$ having the binomial distribution $B_{n,p}$. First we consider the case $n = 1$.

Lemma 4. *For every natural m, the function $g_m(t) := \mathbf{E}(\pi(t) - t)^m$ is a polynomial on $[0, \infty)$ with nonnegative coefficients, and the following inequalities hold*:

$$\mathbf{E}(\nu(1, p) + x)^{2m-1} \leq \mathbf{E}(\pi(p) + x)^{2m-1} \tag{21}$$

if $x \geq -1$; and

$$|\mathbf{E}(\nu(1, p) + x)^{2m-1}| \leq \mathbf{E}(\pi(p) + x)^{2m-1} \tag{22}$$

if $x \geq -p$.

Proof. The properties of the functions $g_m(t)$ immediately follow from the relation

$$g_m(t) = \sum_{k=0}^{\infty}(k-t)^{m-1}(k-t)\frac{t^k}{k!}e^{-t} = \sum_{k=1}^{\infty}(k-t)^{m-1}\frac{t^k}{(k-1)!}e^{-t} - tg_{m-1}$$

$$= t\mathbf{E}(\pi(t) - t + 1)^{m-1} - tg_{m-1} = t\sum_{k=0}^{m-2} \frac{(m-1)!}{k!(m-k-1)!} g_k(t),$$

where $m \geq 2$, $g_0(t) \equiv 1$, and $g_1(t) \equiv 0$.

In order to prove (21) we first study the case $x = -1$. We have

$$\mathbf{E}(\nu(1,p) - 1)^{2m-1} = p - 1,$$

$$\mathbf{E}(\pi(p) - 1)^{2m-1} = -e^{-p} + \sum_{k=2}^{\infty} \frac{(k-1)^{2m-1}}{k!} p^k e^{-p}$$

$$> -e^{-p} + \frac{1}{2}\sum_{k=2}^{\infty} \frac{(k-1)^{2m-3}}{(k-2)!} p^k e^{-p} = -e^{-p} + \frac{p^2}{2}\mathbf{E}(1 + \pi(p))^{2m-3}$$

$$> p - 1 - \frac{p^2}{2} + \frac{p^2}{2}\mathbf{E}(1 + \pi(p))^{2m-3} > p - 1, \tag{23}$$

where $m \geq 2$ (in the case $m = 1$ the assertion is trivial). Inequality (21) follows from (23) and the analogous inequality for the corresponding derivatives with respect to x (see Corollary 4).

To prove (22) we need to deduce only the inequality

$$\mathbf{E}(p - \nu(1,p))^{2m-1} \leq \mathbf{E}(\pi(p) - p)^{2m-1}. \tag{24}$$

First we assume that $p \leq 1/2$. Then we have

$$\mathbf{E}(p - \nu(1,p))^{2m-1} = p(1-p)(p^{2m-2} - (1-p)^{2m-2}) \leq 0,$$

and (24) holds because of nonnegativity of the functions $g_m(t)$.

In the case $p > 1/2$ we put $\tilde{\nu}(1,\tilde{p}) := 1 - \nu(1,p)$, where $\tilde{p} := 1 - p$. By (24) we then obtain

$$\mathbf{E}(p - \nu(1,p))^{2m-1} = \mathbf{E}(\tilde{\nu}(1,\tilde{p}) - \tilde{p})^{2m-1} \leq g_{2m-1}(\tilde{p}) \leq g_{2m-1}(p)$$

due to monotonicity of the functions $g_m(t)$. The lemma is proven.

Since $\nu(n,p)$ coincides in distribution with a sum of independent copies of the random variables $\nu(n-1,p)$ and $\nu(1,p)$, the further proof of the theorem can be continued by induction on n (using (22) and the binomial formula). The theorem is proven.

3 Couplings in Poissonian approximation in Abelian groups

1. Introduction and the main result

Let $\mathcal{S}$ be a separable metric space. Denote by $V(\mathcal{L}(X), \mathcal{L}(Y))$ the total variation distance (with respect to the Borel σ-field) between the distributions of random variables (r.v-s) X and Y taking values in $\mathcal{S}$. We discuss the following fundamental result of Dobrushin, 1970: There exist r.v-s X_0 and Y_0 defined on a common probability space, with the same distributions as X and Y respectively such that

$$V(\mathcal{L}(X), \mathcal{L}(Y)) = \mathbf{Pr}(X_0 \neq Y_0) \tag{1}$$

and, moreover,

$$V(\mathcal{L}(X), \mathcal{L}(Y)) = \inf_{X,Y} \mathbf{Pr}(X \neq Y), \tag{2}$$

where the infimum is taken over all possible constructions of r.v-s X and Y on a common probability space.

Relations (1) and (2) are a duality theorem in the Monge–Kantorovich problem for the indicator metric (for details, see Rachev, 1984). These relations mean that if the total variation distance between the distributions of some r.v-s is sufficiently small, then there exist versions of these r.v-s coinciding with probability close to 1, and this probability is the largest possible.

R e m a r k 1. If $\mathcal{S}$ is a Polish space with a metric r, then relations (1) and (2) follow from the more general result of Strassen, 1965: For every $t \geq 0$, we have

$$\rho(t, P, Q) = \beta(t, P, Q), \tag{3}$$

where

$$\rho(t, P, Q) = \inf_{\xi,\eta} \{\mathbf{P}(r(\xi, \eta) > t)\}$$

with $P = \mathcal{L}(\xi)$, $Q = \mathcal{L}(\eta)$, and

$$\beta(t, P, Q) = \sup\{P(Z) - Q(Z_t) : Z \text{ is closed in } \mathcal{S}\}$$

with $Z_t = \{y \in U : r(Z, y) \leq t\}$. In particular, from (3) it follows that $\rho(0, P, Q) = V(P, Q)$.

Under less restrictive conditions on S (for the so-called inner regular separable metric spaces which may be incomplete) relation (3) was proved by Dudley (1968).

Another proof of the Dudley-Strassen result was proposed by Schay (1974). But at least transition from the discrete case to the general one in that paper (see Theorem 2 in Schay, 1974) is incorrect. Actually, the Schay method of discrete approximation implies continuity of the functional $\rho(t, P, Q)$ on P and Q with respect to the weak convergence topology (for every fixed $t \geq 0$). However, this statement is false (the functional $\rho(t, P, Q)$ is continuous with respect to a stronger metric like the total variation distance).

R e m a r k 2. The original Dobrushin proof of (1) and (2) which was carried out for an arbitrary metric space $\mathcal{S}$ (not necessarily separable), contains an incorrectness because, in general, the event $\{X \neq Y\}$ does not belong to the product σ-field on which every joint distribution is well defined. In other words, it is not clear how to define the probability on the right-hand side of (2) for an arbitrary construction of r.v-s X and Y on a common probability space without separability. In order to correct the proof we need a more careful construction of the optimal joint distribution of (X_0, Y_0) on the product space and the probabilities on the right-hand side of (2) as well.

Now let $(\mathcal{S}, \mathcal{A})$ be an arbitrary measurable space. Denote by $\Sigma = \Sigma(\mathcal{A} \times \mathcal{A}, D)$ the minimal σ-field generated by the standard product σ-field $\mathcal{A} \times \mathcal{A}$ and diagonal $D = \{(x, y) : x = y\}$ of the product space. The following assertion is a more general and somewhat corrected version of the Dobrushin result:

Theorem 1. *On the probability space* $(\mathcal{S} \times \mathcal{S}, \Sigma)$*, there exist copies* X_0 *and* Y_0 *of r.v-s* X *and* Y *respectively such that relations* (1) *and* (2) *hold, where the infimum in* (2) *is taken over all pairs* (X, Y) *based on the above probability space and having the initial marginal distributions.*

Proof. Let P_1 and P_2 be the distributions of X and Y respectively. The Dobrushin proof is based on the classical Hahn–Jordan decomposition $\mathcal{S} = \mathcal{S}^+ \bigcup \mathcal{S}^-$ for the signed measure $G := P_1 - P_2$. In other words, we have $G(A \cap \mathcal{S}^+) \geq 0$ and $G(A \cap \mathcal{S}^-) \leq 0$ for every $A \in \mathcal{A}$.

Put $\nu(A) := P_2(A \cap \mathcal{S}^+) + P_1(A \cap \mathcal{S}^-)$ and $\alpha := \nu(\mathcal{S})$. It is clear that $\alpha \leq 1$ because, from the definition of S^+ and S^-, we have

$$\nu(A) \leq \min\{P_1(A),\, P_2(A)\}.$$

We omit the cases $\alpha = 0$ and $\alpha = 1$ because in the first case, we have $P_1 \perp P_2$ and in the second case, $P_1 = P_2$, and the problem is trivially solved.

Let $0 < \alpha < 1$. Introduce the following three probability measures:

$$\mu(A) := \nu(A)/\alpha,$$

$$\lambda_1(A) := \frac{P_1(A) - \nu(A)}{1 - \alpha},$$

$$\lambda_2(A) := \frac{P_2(A) - \nu(A)}{1 - \alpha}.$$

The construction of the optimal joint distribution $\mathbf{P}(\cdot)$ on the probability space $(\mathcal{S} \times \mathcal{S},\, \Sigma)$ is as follows: For all $A, B \in \mathcal{A}$, we put

$$\mathbf{P}(A \times B) = \alpha\mu(A \cap B) + (1 - \alpha)\lambda_1(A)\lambda_2(B),$$

$$\mathbf{P}((A \times B) \cap D) = \alpha\mu(A \cap B).$$

Note that, for a separable metric space, it is sufficient to define the optimal distribution only on rectangles $A \times B$ (as in Dobrushin, 1970). In other words, we have constructed an elementary measure on the minimal semiring generated by all canonical rectangles and the diagonal of the product space. Thus, by the classical extension theorem of measure theory, we can extend this elementary probability to the minimal σ-field containing this semiring of subsets.

It is easy to see that $\mathbf{P}(A \times \mathcal{S}) = P_1(A)$, $\mathbf{P}(\mathcal{S} \times B) = P_2(B)$, and

$$\mathbf{P}(D) = \nu(\mathcal{S}) = P_2(\mathcal{S}^+) + P_1(\mathcal{S}^-) = 1 - G(\mathcal{S}^+)$$
$$= 1 - V(\mathcal{L}(X), \mathcal{L}(Y)).$$

Relations (1) and (2) follow from the above and the elementary inequality

$$\mathbf{Pr}(X \in A) \leq \mathbf{Pr}(Y \in A) + \mathbf{Pr}(X \neq Y)$$

which is true for all constructions of X and Y on a common probability space with measurable diagonal. In other words, in this case,

$$V(\mathcal{L}(X), \mathcal{L}(Y)) \leq \mathbf{Pr}(X \neq Y).$$

The theorem is proved.

Remark 3. In the formulation of Theorem 1, the infimum in (2) can be taken over all constructions of r.v-s X and Y on a common probability space if we replace $\mathbf{Pr}$ in (2) with the symbol of the outer probability $\mathbf{Pr}^*$ which is defined by the above-introduced extended product σ-field Σ.

Remark 4. Actually, the construction of the optimal joint distribution of (X_0, Y_0) in the Dobrushin theorem is also contained in Le Cam (1965) (however, without proving equalities (1) and (2)). Le Cam's proof is based on the notion of the minimal measure that is equivalent to the above-mentioned Hahn-Jordan decomposition. It is easy to verify that the Le Cam coupling provides equality (1) as well (at least for separable metric spaces).

2. Poisson approximation in Abelian groups

Let $\{X_k;\, k \geq 1\}$ be independent identically distributed (i.i.d.) r.v-s taking values in an arbitrary measurable Abelian group $(\mathcal{G}, \mathcal{A})$ with measurable operation "+". In this case, we have no measurability difficulties mentioned above. Let $\{X_i^0\}$ be i.i.d. r.v-s with distribution

$$\mathcal{L}(X_1^0) = \mathcal{L}(X_1 | X_1 \neq 0),$$

where the distribution on the right-hand side of this equality is conditional under the condition $\{X_1 \neq 0\}$. Put $p := \mathbf{Pr}(X_1 \neq 0)$, $S_n := \sum_{i \leq n} X_i$, and $S_n^0 := \sum_{i \leq n} X_i^0$. Introduce the so-called accompanying (for S_n) Poisson r.v. Π_n. It is well known that this r.v. coincides in distribution with the sum $S_{\pi(n)}$, where the r.v. $\pi(n)$ has the Poisson distribution with parameter n and is independent of the initial sequence of r.v-s. We need more convenient representations of the r.v-s introduced above. Actually, these representations in various forms are contained in many papers (see, for example, Khintchine, 1933, 1936; Le Cam, 1960, 1965; K.Borovkov, 1988; Borisov, 1993, 1996) and they can be briefly defined by the relations

$$\mathcal{L}(S_n) = \mathcal{L}(S^0_\nu), \quad \mathcal{L}(\Pi_n) = \mathcal{L}(S^0_\pi), \tag{4}$$

where $\mathcal{L}(\nu) = B_{n,p}$ is the binomial distribution with parameters n and p; $\mathcal{L}(\pi) = P_{np}$ is the classical Poisson distribution with parameter np; the pair $(\nu,\, \pi)$ does not depend on the sequence $\{X^0_i\}$.

Whence the following inequality immediately follows (say, by Theorem 1):

$$V(\mathcal{L}(S_n), \mathcal{L}(\Pi_n)) \leq V(B_{n,p}, P_{np}). \tag{5}$$

Moreover, if there exists (with probability 1) a measurable one-to-one mapping $n \to S^0_n$ (typical property of empirical processes) then it is easy to obtain the equality

$$V(\mathcal{L}(S_n), \mathcal{L}(\Pi_n)) = V(B_{n,p}, P_{np}). \tag{6}$$

In other words, the problem of approximating the distributions of the sums in Abelian groups by the accompanying Poisson laws is easily reduced to the simplest one-dimensional case. In our opinion, employment of relations (5) and (6) allows us to obtain the corresponding upper bounds by the shortest way. Using Theorem 1 we easily obtain the corresponding coupling with an exact upper bound.

Sometimes we study approximation of the distributions by Poisson laws different from the accompanying. In this case, we can use the well-known representation of such a Poisson law as an operator exponential. Denote by $Pois(\mu)$ the generalized Poisson distribution in $(\mathcal{G}, \mathcal{A})$ with the Lévy measure μ:

$$Pois(\mu) := e^{-\mu(\mathcal{G})} \sum_{k=0}^{\infty} \frac{\mu^{*k}}{k!}, \tag{7}$$

where μ^{*k} is the k-fold convolution of a finite measure μ with itself; μ^{*0} is the unit mass concentrated at zero. In particular, using relations (4), it is easy to verify the double equality

$$\mathcal{L}(\Pi_n) = Pois(n\mathcal{L}(X_1)) = Pois(np\mathcal{L}(X^0_1)). \tag{8}$$

¿From (7) it is easy to obtain the estimate (for details, see Borisov, 2000)

$$V(Pois(\mu), Pois(\lambda)) \leq 2V(\mu, \lambda), \tag{9}$$

where the factor “2” in the right-hand side of (9) can be omitted if either $\lambda(\mathcal{G}) = \mu(\mathcal{G})$ or $\lambda \geq \mu$ ($\lambda \leq \mu$). This inequality improves the corresponding result in Reiss (1993, p. 87).

As a consequence of Theorem 1, relations (4), (8), (9), and the classical Prohorov-Le Cam estimate for the right-hand side in (5) as well (see Prohorov, 1953; Le Cam, 1960; Barbour and Hall, 1984), we can formulate the following result:

Theorem 2. *Let $\Pi(\mu)$ be an r.v. in $(\mathcal{G}, \mathcal{A})$ with distribution Pois(μ). Then r.v-s S_n and $\Pi(\mu)$ can be defined on a common probability space such that*

$$\mathbf{Pr}(S_n \neq \Pi(\mu)) \leq \min\{p, np^2\} + 2V(np\mathcal{L}(X_1^0), \mu). \tag{10}$$

Corollary. *Let $\{X_{nk};\, k \leq n\}$, $n = 1, 2, ...,$ be row-wise i.i.d. r.v-s in $(\mathcal{G}, \mathcal{A})$. Given a finite measure μ, let the condition*

$$\lim_{n\to\infty} V(np_n\mathcal{L}(X_{n1}^0), \mu) = 0 \tag{11}$$

be fulfilled, where $p_n = \mathbf{Pr}(X_{n1} \neq 0)$. Then, for each n, r.v-s $S_n^{(n)} := \sum_{k\leq n} X_{nk}$ and $\Pi(\mu)$ can be defined on a common probability space such that

$$\lim_{n\to\infty} \mathbf{Pr}(S_n^{(n)} \neq \Pi(\mu)) = 0. \tag{12}$$

Note that, under the conditions of the Corollary, we have $np_n \to \mu(\mathcal{G})$. Thus, using the upper bound in (10), we can estimate the rate of convergence in (12).

Remark 5. The above-introduced measurability condition on the operation "+" can be weakened. We only need a correct definition of the distributions of S_n^0 for all n. In other words, if $0 \in \mathcal{A}$, we can axiomatically define the nth convolutions we need, and convolutions of the initial distribution of the summands as well (for details, see Borisov, 1993). In this case, the accompanying Poisson law is well defined. However, for every construction of r.v-s S_n and $\Pi(\mu)$ on a common probability space, we cannot guarantee measurability of the event $\{S_n \neq \Pi(\mu)\}$ as well as correct definition of the corresponding probability. Theorem 2 provides such a construction.

3. Poisson approximation of empirical processes

Let $\{Y_{nk};\, k \leq n\}$, $n = 1, 2, ...,$ be row-wise i.i.d. r.v-s taking values in an arbitrary measurable space $(\mathcal{S}, \mathcal{A})$. For each n, introduce the normalized empirical process based on the sample $\{Y_{nk};\, k \leq n\}$ and indexed by a family $\mathcal{F}$ of real-valued Borel measurable functions on $(\mathcal{S}, \mathcal{A})$:

$$S_n(f) := \sum_{k\leq n} f(Y_{nk}), \quad f \in \mathcal{F}.$$

We consider S_n as an r.v. in the measurable space $(R^{\mathcal{F}}, \mathcal{C})$ of all real-valued functions indexed by $\mathcal{F}$, with the cylinder σ-field $\mathcal{C}$ (Kolmogorov's space).

Denote by $\nu_n(A) := S_n(I_A)$ the normalized empirical measure, where $I_A(\cdot)$ is the indicator function. The following theorem is a simple consequence of the Corollary above:

Theorem 3. *Given some $B \in \mathcal{A}$ and some finite measure λ with support in B, suppose that $V(np_n\mathcal{L}(Y_{n1}^B), \lambda) \to 0$ as $n \to \infty$, where $\mathcal{L}(Y_{n1}^B) = \mathcal{L}(Y_{n1} \,|\, Y_{n1} \in B)$ and*

$p_n = \mathbf{Pr}(Y_{n1} \in B)$. *Then the sample* $\{Y_{nk};\, k \le n\}$ *and a Poisson point process* π_λ *on* $\mathcal{A}$ *with mean measure* λ *can be defined on a common probability space such that*

$$\lim_{n\to\infty} \mathbf{Pr}\{\sup_{A\subseteq B} |\nu_n(A) - \pi_\lambda(A)| \ne 0\} = 0.$$

Because of the identity

$$S_n(f) = \int_{\mathcal{S}} f(x)\nu_n(dx)$$

we can reformulate Theorem 3 in the following equivalent form:

Theorem 3′. *Under the conditions of Theorem* 3, *we can define the random process* $S_n(f)$ *and the generalized Poisson process*

$$\Pi(f) := \int_{\mathcal{S}} f(x)\pi_\lambda(dx),$$

on a common probability space such that

$$\lim_{n\to\infty} \mathbf{Pr}\{\bigcup_{f\in\mathcal{F}} \{S_n(f) \ne \Pi(f)\}\} = 0.$$

Proof of Theorem 3. Introduce the notations

$$X_{nk} := \{I_{A\cap B}(Y_{nk});\, A \in \mathcal{A}\},$$

$$X^0_{nk} := \{I_A(Y^B_{nk});\, A \in \mathcal{A}\},$$

$$p_n := \mathbf{Pr}(Y_{n1} \in B) = \mathbf{Pr}(X_{n1} \ne 0).$$

We consider the random processes $\{X_{nk}\}$ as elements of the corresponding Kolmogorov space. In order to apply the Corollary of Theorem 2, we need to verify condition (11). Denote by $\{Z_i;\, i \ge 1\}$ a sequence of i.i.d. r.v-s in $\mathcal{S}$ with the distribution $\lambda(\cdot)/\lambda(\mathcal{S})$. It is easy to see that the Lévy measure of the Poisson point process π_λ introduced in the statement of Theorem 3 coincides with $\mu := \lambda(\mathcal{S})\mathcal{L}\{I_A(Z_1);\, A \in \mathcal{A}\}$. It only remains to evaluate the total variation distance in the Kolmogorov space between μ and $np_n\mathcal{L}\{I_A(Y^B_{nk});\, A \in \mathcal{A}\}$. In order to do this, we need to estimate closeness of the measures only over all finite-dimensional cylindrical subsets. Note that, for arbitrary measurable subsets $A_1, \dots, A_k$ and an arbitrary $\mathcal{S}$-valued r.v. ξ, we have the equality of the events

$$\{I_{A_1}(\xi) = e_1, \dots, I_{A_k}(\xi) = e_k\} = \{\xi \in \bigcap_{i\le k} C_i\},$$

where $e_i \in \{0, 1\}$, $C_i = A_i$ whenever $e_i = 1$, and $C_i = A^c_i$ (A^c is the complement of A) whenever $e_i = 0$.

Therefore,

$$V(np_n\mathcal{L}(X^0_{n1}), \mu) \leq V(np_n\mathcal{L}(Y^B_{n1}), \lambda).$$

Finally, using Theorem 1, we complete the proof.

Remark 6. Actually, the convergence in the total variation distance of the distributions of restrictions to a subset B of the point processes $\nu_n(\cdot)$ to that of the analogous restriction of $\pi_\lambda(\cdot)$ was studied in Reiss (1993). Thus the statement of Theorem 3 can be deduced from Theorem 1 and Theorem 3.2.3 in Reiss (1993). Under some separability conditions on the parametric set such a proof of the coupling was proposed by Einmahl (1997).

Remark 7. Many authors studied some problems of Poisson approximation close to that mentioned above in Theorem 3 (say, the so-called strong Poisson approximation of empirical measures) in various particular cases and applied this approximation in limit theorems of various kinds. For example, the cases were studied in which $A \subseteq B$, where B is a bounded Borel subset in R^k, or $A = \{x \in R^k : x \geq a\}$, $a \geq a_0 > 0$, and $X_{n1} = m(n)X_1$, where $m(n) \to \infty$ for the first case, and $m(n) \to 0$ for the second case (the so-called local and tail empirical processes; see Deheuvels and Pfeifer, 1988; Deheuvels and Mason, 1995; Major, 1990; Horvath, 1990; Borisov, 1993, 1996, 2000; Einmahl, 1997; and others). Some of them used the so-called Serfling coupling because of studying the approximation by accompanying Poisson processes (see Remark 8 below).

Note that analogs of inequality (5) for the above-mentioned particular cases can be found in the corresponding papers by Deheuvels and Pfeifer (1988), Major (1990), Borisov (1993, 1996), and some others. Actually, for the empirical and accompanying Poisson point processes, equality (6) holds.

Note also that, under some additional separability restrictions on the parametric set, the corresponding references to the above-mentioned Dobrushin's result are contained in papers by Borisov (1993, 1996, 2000), Borisov and Mironov (2000), and Einmahl (1997).

Remark 8. In the mid 1970s, to prove the Poisson limit theorem, two close coupling methods were proposed independently by Serfling (1975) and A.Borovkov (1976) (see also English translation of the last textbook by A.Borovkov, 1998, p.100). For example, Borovkov's method is based on comparison of the standard quantile transforms for Bernoulli and the corresponding Poisson distribution functions (for details, see also Borisov, 1993, 1996). The Lebesgue measure of noncoincidence (as well as in Serfling's coupling, although his construction is slightly different from that of Borovkov) of these quantile transforms on the unit interval is equal to $p(1 - e^{-p}) < p^2$. It is worth noting that the above-mentioned optimal coupling in Le Cam (1965) provides a construction of these r.v-s on a common probability space with a smaller probability of their noncoincidence (for details, see Le Cam, 1965, p. 186). Whence the Le Cam upper bound np^2 immediately follows for the probability of noncoincidence of r.v-s with the (n, p)-Binomial and accompanying Poisson distributions based on the probability space $[0, 1]^n$ with the n-variate Lebesgue measure. Moreover, it is easy to obtain a lower bound of the same order np^2 for the probability of noncoincidence under Borovkov's and Serfling's constructions. However, Dobrushin's theorem allows us to obtain optimal coupling in the Poisson approximation with the probability of noncoincidence of order $\min\{p,\, np^2\}$. This is Prohorov's

(1953) and Le Cam's (1960) combined estimate for the total variation distance between the distributions under study. In other words, the above-mentioned three couplings for each pair of the summands in the sums of independent r.v-s under consideration provide worse Poisson approximation for the (n, p)-Binomial distribution than Dobrushin's (as well as Le Cam's) coupling for the whole sums.

It is very important to note that Borovkov's and Serfling's couplings are applicable only in approximation of the sum distributions by the accompanying Poisson laws (see the proof of Theorem 2). However, the limit Poisson processes in Theorems 3 and $3'$ differ from the corresponding accompanying Poisson processes. In this case, Borovkov's and Serfling's couplings do not work.

R e m a r k 9. Borovkov's coupling mentioned in Remark 8 can be interpreted as a particular case of the well-known ε-coupling proposed by Prohorov (1956) studying the convergence rates in the Donsker invariance principle. This method was also based on comparison of quantile transforms for the distribution functions of independent increments (on subintervals of a suitable length) of a partial sum process and a Wiener process.

To study proximity between the distributions of two partial sum processes based on Bernoulli and accompanying Poisson r.v-s respectively (Poisson "invariance principle" or strong Poisson approximation), we can use Dobrushin's and Le Cam's constructions as well as Borovkov's and Serfling's couplings because a lower bound for the total variation distance between the distributions of these partial sum processes in the corresponding functional sample space has order np^2 (see Borisov, 1996).

4 Poissonian approximation of the partial sum processes in Banach spaces

1. Statement of the Main Results

Let $X, X_1, X_2, \ldots$ be independent random variables taking values in a measurable Banach space $(\mathbf{B}, \mathcal{A}, \|\cdot\|)$ and having the respective distributions $P, P_1, P_2, \ldots$; here $\mathcal{A}$ is a σ-field in $\mathbf{B}$. Hereafter, we suppose that the linear operations and the norm $\|\cdot\|$ are measurable with respect to $\mathcal{A}$. If $\mathbf{B}$ is separable, then $\mathcal{A}$ is usually considered to be the Borel σ-field. The measurability conditions are fulfilled in this case. For a nonseparable $\mathbf{B}$, such choice of $\mathcal{A}$ is not natural, since in this case the family of all measurable random variables may become too narrow (see [30]). Thus, for nonseparable Banach spaces, the class $\mathcal{A}$ is, as a rule, smaller than the Borel σ-field (for example, the cylindrical or ball σ-fields).

Denote by $Pois(\mu)$ the generalized Poisson distribution with the Lévy measure μ:

$$Pois(\mu) := e^{-\mu(\mathbf{B})} \sum_{k=0}^{\infty} \frac{\mu^{*k}}{k!},$$

where μ^{*k} is the k-fold convolution of a finite measure μ with itself; μ^{*0} is the unit mass concentrated at zero. It is easy to see that $Pois(\mu)$ coincides with the distribution of the sum $\sum_{i \le \pi(\mu(\mathbf{B}))} Y_i$, where $\{Y_i\}$ are independent identically distributed random variables having the common distribution $\mu(\cdot)/\mu(\mathbf{B})$, and $\pi(\mu(\mathbf{B}))$ is a Poisson random variable having mean $\mu(\mathbf{B})$ and independent of $\{Y_i\}$.

The main goal of this chapter is to obtain path proximity estimates between the random process (vector)

$$S_n := \left\{ \sum_{i=1}^{k} X_i;\ k = 1, \ldots, n \right\} \tag{1}$$

and the generalized Poisson process

$$\Pi_n := \left\{ \sum_{i=1}^{k} \hat{\pi}_i(P_i);\ k = 1, \ldots, n \right\}, \tag{2}$$

where $\{\hat{\pi}_k(P_k)\}$ are independent random variables with the respective distributions $\{Pois(P_k)\}$. We study the proximity in terms of the distance

$$d(z, S_n, \Pi_n) := \inf_{\{X_k, \hat{\pi}_k(\cdot);\ k \le n\}} \mathbf{P}\left(\max_{k \le n} \left\| \sum_{i=1}^{k} X_i - \sum_{i=1}^{k} \hat{\pi}_i(P_i) \right\| > z \right), \tag{3}$$

where the infimum is taken over all families of $\{X_k\}$ and $\{\hat{\pi}_k(\cdot)\}$ based on a common probability space. Alongside (3), we also need the total variation distance between the distributions $\mathcal{L}(X_i)$ of arbitrary $\mathbf{B}$-valued random variables X_i, $i = 1, 2$:

$$V(X_1, X_2) := \sup_{A \in \mathcal{A}} |\mathbf{P}(X_1 \in A) - \mathbf{P}(\mathbf{X_2} \in \mathbf{A})|.$$

We now recall the most significant relevant results. One of the first results was obtained by Yu. V. Prokhorov [98] and was later generalized by L. Le Cam [82]. In the case when $\mathbf{B} = \mathbf{R}$ and $X_i \equiv \nu_i$ are Bernoulli random variables with the respective success probabilities p_i, the following inequality was proved in [98] and [82]:

$$V\Big(\sum_{i \le n} \nu_i, \pi\Big(\sum_{i \le n} p_i\Big)\Big) \le c p_0, \tag{4}$$

where c is an absolute constant, $p_0 = \max_{i \le n} p_i$, and $\pi(s)$ is a Poisson random variable with mean s. Somewhat later, L. Le Cam proved (see [83], [84]) the following inequality for every random variable X taking values in an Abelian group $\mathbf{B}$ (under the above measurability conditions):

$$V(X, \hat{\pi}(P)) \le p^2, \tag{5}$$

where $p := \mathbf{P}(X \ne 0)$. Moreover, if $\mathbf{B}$ is a Polish space, then it is easy to obtain from (5) the following estimate

$$V\Big(\sum_{i \le n} X_i, \sum_{i \le n} \hat{\pi}_i(P_i)\Big) \le \sum_{i \le n} p_i^2, \tag{6}$$

where $p_i := \mathbf{P}(X_i \ne 0)$.

Therefore, in the Bernoulli case, (4) and (6) yield a more universal upper bound

$$V\Big(\sum_{i \le n} \nu_i, \pi\Big(\sum_{i \le n} p_i\Big)\Big) \le \min\Big(c p_0, \sum_{i \le n} p_i^2\Big). \tag{7}$$

The best two-sided estimate was obtained by A. Barbour and P. Hall (see [5]):

$$\frac{1}{32} \varepsilon_n \sum_{i \le n} p_i^2 \le V\Big(\sum_{i \le n} \nu_i, \pi\Big(\sum_{i \le n} p_i\Big)\Big) \le \varepsilon_n \sum_{i \le n} p_i^2, \tag{8}$$

where $\varepsilon_n := \min\big\{1, \big(\sum_{i \le n} p_i\big)^{-1}\big\}$. Note that, under the condition $\sum_{i \le n} p_i \le 1$, the right-hand sides in (6) and (8) coincide, and for the identically distributed ν_i the right-hand sides in (7) and (8) coincide up to an absolute multiplicative constant. Note also that, in the Bernoulli case under the condition $\sum_{i \le n} p_i \to \infty$, the exact asymptotics of the total variation distance under consideration is known (see [7]) and differs from the right-hand side of (8) only by the factor $(2\pi e)^{-1/2}$.

For non-Bernoulli sequences $\{X_i\}$, estimate (6) cannot be improved without additional restrictions on the distributions P_i. For example, if each random variable X_i has a two-points distribution concentrated at zero and the point $(a_i)^{1/2}$ where $\{a_i\}$ are different prime numbers then, because of one-to-one correspondence between rational numbers $\{e_i\}$ and the sums $\sum_{i\le n}\sqrt{a_i}e_i$, we have (see [83])

$$V\Big(\sum_{i\le n}X_i, \sum_{i\le n}\hat\pi_i(P_i)\Big) \ge 1-\mathbf{P}\Big(\sum_{i\le n}\hat\pi_i(P_i)\in\Big\{\sum_{i\le n}\sqrt{a_i}e_i;\ e_i=0,1,\ i\le n\Big\}\Big)$$

$$=1-\prod_{i\le n}e^{-p_i}(1+p_i)\ge 1-\prod_{i\le n}\big(1-p_i^2(1-p_i)/2\big)\ge 1-\exp\Big\{-\frac{1}{2}\sum_{i\le n}p_i^2(1-p_i)\Big\}. \qquad (9)$$

In other words, if $\sum_{i\le n}p_i^2\to 0$, then the right-hand side of (9) is equivalent to $\frac{1}{2}\sum_{i\le n}p_i^2$ (where, possibly, $\sum_{i\le n}p_i\to\infty$).

We would like to emphasize the significance of the fundamental result by R. L. Dobrushin (see [45]):

$$V(X,Y)=\inf_{X,Y}\mathbf{P}(X\ne Y), \qquad (10)$$

where the infimum in the right-hand side of (10) is taken over all pairs (X,Y) based on a common probability space, and $\mathcal{A}$ is the Borel σ-field in $\mathbf{B}$. If $\mathcal{A}$ is smaller than the Borel σ-field then the sign "=" in (10) must be replaced by the sign "$\le$."

Relation (10) means that if the total variation distance between some random variables is sufficiently small, then there exist versions of these random variables coinciding with probability close to 1.

R e m a r k. The original proof of (10) in [45] which was carried out for arbitrary metric spaces $\mathbf{B}$ (not necessarily separable) is incorrect since the event $\{X\ne Y\}$ may be nonmeasurable. Postulating the measurability of linear operations (or the group operation when $\mathbf{B}$ is an Abelian group) allows us to avoid this inconvenience.

Before evaluating $d(z,\cdot)$, let us agree to consider the partial sum processes (1) and (2) as random elements in the measurable Banach space $\mathbf{B}^n:=\mathbf{B}\times\cdots\times\mathbf{B}$ endowed with the norm $\|Y\|_\infty:=\max_{k\le n}\|Y^{(k)}\|$ and the σ-field $\mathcal{A}^n:=\mathcal{A}\times\cdots\times\mathcal{A}$. If $\mathbf{B}$ is separable then $\mathbf{B}^n$ is separable too. In this case, it follows from (10) that the total variation distances between the distributions of S and Π_n, and those between $(X_1,\dots,X_n)$ and $(\hat\pi_1(P_1),\dots,\hat\pi_n(P_n))$ coincide. Moreover, by (5) and (10) we obtain

$$d(0,S_n,\Pi_n)\le \inf_{S_n,\Pi_n:\{(X_k,\hat\pi_k(P_k))\}\text{ are independent}} P\Big(\bigcup_{k\le n}\{X_k\ne\hat\pi_k(P_k)\}\Big)$$

$$\le\sum\inf_{X_k,\pi_k(P_k)}P(X_k\ne\hat\pi_k(P_k))\le\sum_{k\le n}p_k^2. \qquad (11)$$

Note that, even in the Bernoulli case, the upper bound in (11) cannot be essentially improved (in contrast to (6)) since the distance $d(0,S_n,\Pi_n)$ admits the lower bound by the analogy with (9):

$$d(0,S_n,\Pi_n)\ge 1-\mathbf{P}(\pi_i(p_i)\in\{0,1\};\ i=1,\dots,n)\ge 1-\exp\Big\{-\frac{1}{2}\sum_{i\le n}p_i^2(1-p_i)\Big\}.$$

If $\mathbf{B}$ is not separable then, generally speaking, (10) does not hold. Thus, to estimate $d(z,\cdot)$, we need a new approach. Denote by $\{X_i^0\}$ the independent random variables with the following distributions:

$$\mathbf{P}\big(X_i^0 \in A\big) = \mathbf{P}(X_i \in A \mid X_i \neq 0). \tag{12}$$

Note that the random variables are well defined because, by the above assumptions, the event $\{X_i \neq 0\} = \{\|X_i\| \neq 0\}$ is measurable. It is proved in [17] that for every k the following equalities hold:

$$\mathcal{L}(X_k) = \mathcal{L}\Bigg(\sum_{i=\overline{\nu}_n(k-1)+1}^{\overline{\nu}_n(k)} X_{k,i}^0\Bigg), \quad Pois(P_k) = \mathcal{L}\Bigg(\sum_{i=\bar{\pi}_n(k-1)+1}^{\bar{\pi}_n(k)} X_{k,i}^0\Bigg), \tag{13}$$

where $\big\{X_{k,i}^0; i = 1, 2, \dots\big\}$ are independent copies of X_k^0, the random processes $\overline{\nu}_n(k) := \sum_{i\le k} \nu_i$ and $\bar{\pi}_n(k) := \sum_{i\le k} \pi_i(p_i)$, $k \le n$, do not depend on $\{X_{k,i}^0\}$, and $p_i := \mathbf{P}(X_i \neq 0)$, $\sum_{i=1}^{0} = 0$. If X_i^0 are identically distributed (whereas X_i can be nonidentically distributed with arbitrary p_i), then (13) and independence of the increments of the processes $\bar{\nu}_n(\cdot)$ and $\bar{\pi}_n(\cdot)$ readily imply more informative representations for the distributions of S_n and Π_n in the Banach space $\mathbf{B^n}$:

$$\mathcal{L}(S_n) = \mathcal{L}\left\{\sum_{i=1}^{\bar{\nu}_n(k)} X_i^0;\ k = 1, \dots, n\right\}, \quad \mathcal{L}(\Pi_n) = \mathcal{L}\left\{\sum_{i=1}^{\bar{\pi}_n(k)} X_i^0;\ k = 1, \dots, n\right\}, \tag{14}$$

We formulate several simple and useful consequences of (13), (14), and (10) (see [17]).

Corollary 1. *If* $\mathbf{B}$ *is a separable Banach space then the inequality*

$$V(S_n, \Pi_n) \le V(\bar{\nu}_n(\cdot), \bar{\pi}_n(\cdot)) \tag{15}$$

holds for arbitrary distributions $\{P_i\}$.

Corollary 2. *If* $\mathbf{B}$ *is a separable Banach space and* $\mathbf{P}\big(\sum_{i\le m} X_{k,i}^0 = 0\big) = 0$ *for all natural numbers* m *and* k *then the following relation holds*

$$V(S_n, \Pi_n) = V(\bar{\nu}_n(\cdot), \bar{\pi}_n(\cdot)). \tag{16}$$

Corollary 3. *If* $\|X_i\| \le 1$, $i = 1, 2 \dots$, *almost surely and random variables* $\{X_i^0\}$ *are identically distributed then the inequality*

$$d(z, S_n, \Pi_n) \le d(z, \bar{\nu}_n(\cdot), \bar{\pi}_n(\cdot)) \tag{17}$$

holds for all $z \ge 0$.

Finally, the last consequence is connected with a special structure of the random variables $\{X_i\}$. We say that X_i *are elements of the indicator type with conforming supports* if for all natural m and k the following identity holds with probability 1:

$$\left\|\sum_{i=m+1}^{m+k} X_i^0\right\| = k. \tag{18}$$

Corollary 4. *If $\{X_i\}$ are random elements of the indicator type with conforming supports, and $\left\{X_i^0\right\}$ are identically distributed then the identity*

$$d(z, S_n, \Pi_n) = d(z, \bar{\nu}_n(\cdot), \bar{\pi}_n(\cdot)). \tag{19}$$

holds for all $z \geq 0$

Hence, for the identically distributed $\left\{X_i^0\right\}$, evaluation of the distances $V(\cdot)$ and $d(\cdot)$ in arbitrary Banach spaces reduces to an analogous problem for sequences of independent, generally speaking, nonidentically distributed Bernoulli random variables with special success probabilities.

The main result of the chapter is as follows:

Theorem 1. *Let $\{\nu_i\}$ be independent identically distributed Bernoulli random variables with the success probability p. Then*

$$d(z, \bar{\nu}_n(\cdot), \bar{\pi}_n(\cdot)) \leq (np^2)^{[z]+1} \exp\{-C_1 z \ln\ln(z+2) + C_2\} \quad \text{if } np \geq 1,$$

$$np^{[z]+2} \exp\{-C_3 z + C_4\}, \quad \text{if } np \leq 1, \tag{20}$$

where C_1–C_4 are absolute positive constants.

Therefore, for bounded identically distributed $\mathbf{B}$-valued random variables $\{X_i\}$, (15)–(20) allow us to obtain an appropriate estimate.

As an example consider the problem of the uniform Poisson approximation of the sequence of multivariate empirical distribution functions. Let $F_n(\mathbf{t})$, $\mathbf{t} \in \mathbf{R^m}$, be the empirical distribution function based on the sample $Y_1, \ldots, Y_n$ from an arbitrary distribution in $\mathbf{R}^m$. Put $X_i \equiv X_i(\mathbf{t}) := \mathbf{I}(\mathbf{Y_i} < \mathbf{t})$, where $\mathbf{I}(\cdot)$ is the indicator function and the sign "$<$" means the standard partial order in $\mathbf{R^m}$. We consider X_i as elements of the Banach space $\mathbf{B}$ consisting of all bounded left-continuous functions (in the sense of the partial order mentioned) defined on the set $\{\mathbf{t} \in \mathbf{R^m} : \mathbf{t} \leq \mathbf{a}\}$ and endowed with the uniform norm. Denote by $\mathcal{A}$ the cylindrical σ-field. The norm introduced is $\mathcal{A}$-measurable although the space $\mathbf{B}$ is not separable. Without loss of generality, we may assume that $\mathbf{P}(X_i \neq 0) \equiv \mathbf{P}(Y_i \leq \mathbf{a}) \neq \mathbf{0}$. Then the random variables X_i^0 in (12) are well defined and satisfy (18). Therefore, from (19) and (20) we obtain the estimate

$$d(z, \widetilde{S}_n, \widetilde{\Pi}_n) \leq \delta(z, n, p), \tag{21}$$

where

$$\widetilde{S}_n := \{kF_k(\mathbf{t});\ k = 1, \ldots, n\}, \quad \widetilde{\Pi}_n := \left\{\sum_{i=1}^{\pi(k)} \mathbf{I}(Y_i < \mathbf{t});\ k = 1, \ldots, n\right\},$$

$\pi(\cdot)$ is the standard Poisson process (with mean s) independent of $\{Y_i\}$, $\delta(z, n, p)$ is the right-hand side of (20), $p := \mathbf{P}(Y_1 \leq \mathbf{a})$.

The particular case considered above was studied by a number of authors (see [1, 17, 62, 86]). In [62], the upper bound for the left-hand side of (21) had the form $\sqrt{n}p(z^{-1} + z^{-2})$. In [17], it was improved by means of (17): $np^2 \exp\{-C_1\sqrt{z} + C_2\}$. Under the additional restriction $np \geq 1$ in [86] (in the special case $p = p(n)$) and [1], it was proved that the left-hand side of (21) admits the upper bound $C(z)(np^2)^{[z]+1}$, but the structure of $C(z)$ was not investigated. Note that in the particular case mentioned, in [1] and [86], an equality like (19) was used which also reduced the problem to the one-dimensional case.

Therefore, Theorem 1 and Corollaries 1–4 strengthen and essentially generalize the above results.

Note that, for $z = 0$, the right-hand side of (20) has order $O(np^2)$, i.e. it is unimprovable up to an absolute multiplicative constant. Moreover, under the condition $np \leq 1$, the dependence of the right-hand side of (20) on p is optimal, while the dependence on z is close to the optimal for any fixed n and p.

Theorem 2. *For any p, n and z, the following relation holds*

$$d(z, \bar{\nu}_n(\cdot), \bar{\pi}_n(\cdot)) > (2\pi)^{-1/2} p^{[z]+2} \exp\{-(z + 5/2)\ln(z + 2)\}. \tag{22}$$

R e m a r k. In [62], a simpler coupling was used which is based on the equality mentioned above

$$\mathcal{L}(\diamond_\backslash) = \mathcal{L}\Big\{ \sum_{\rangle=\infty}^{\pi(\|)} \mathcal{X}_\rangle;\ \| = \infty, \dots, \backslash \Big\}, \tag{23}$$

where $\pi(t)$ is the standard Poisson process independent of X_i. As an example, we compare the nth coordinates of the random vectors S_n and Π_n in (23). By the total probability formula we have

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i \neq \sum_{i=1}^{\pi(n)} X_i \right) = 1 - \mathbf{E}(1 - p)^{|\pi(n) - n|}. \tag{24}$$

On the other hand, it is clear that $\mathbf{E}|\pi(n) - n| = O(\sqrt{n})$. Thus, under the conditions $n \to \infty$ and $p\sqrt{n} \to 0$, the left-hand side of (24) tends to zero at the rate of $p\sqrt{n}$. Hence, the coupling based on representation (23) cannot provide the optimal upper bound for $d(z, S_n, \Pi_n)$.

2. Proof of the Main Results

Proof of Theorem 1. We begin with constructing the families of random variables $\{v_i;\ i \leq n\}$ and $\{\pi_i(p);\ i \leq n\}$ on a common probability space by the method proposed independently in [14] and [15]. Given a random variable ζ, we introduce the notation $F_\zeta^{-1}(\omega) := \inf\{t \in \overline{\mathbf{R}} : F_\zeta(t) \geq \omega\}$, where $\overline{\mathbf{R}}$ is the extended real line, $F_\zeta(t)$ is the distribution function of ζ, and $\omega \in [0, 1]$. It is well known that if ω has the uniform distribution on $[0, 1]$, then $\mathcal{L}(F_\zeta^{-1}(\omega)) = \mathcal{L}(\zeta)$. We set

$$\nu_i^* := F_{\nu_i}^{-1}(\omega_i), \quad \pi_i^* := F_{\pi_i(p)}^{-1}(\omega_i),$$

where $\{\omega_i;\ i \le n\}$ are independent random variables with the uniform distribution on $[0,1]$ It is easy to see that

$$\max_{k\le n}\left|\sum_{i=1}^{k}\nu_i^* - \sum_{i=1}^{k}\pi_i^*\right| \le \sum_{i=1}^{n}\zeta_i, \tag{25}$$

where $\zeta_i = (\pi_i^*-1)\mathbf{I}\{\pi_i^* \ge 2\}+\mathbf{I}\{\omega_i \in [1-p, e^{-p}]\}$. It is easy to verify that the distribution of ζ_i is determined as follows:

$$\mathbf{P}(\zeta_i = 0) = 1 - p(1-e^{-p}), \quad \mathbf{P}(\zeta_i = 1) = e^{-p} - 1 + p + \frac{p^2}{2}e^{-p},$$

$$\mathbf{P}(\zeta_i = k) = \frac{p^{k+1}}{(k+1)!}e^{-p}, \quad k \ge 2. \tag{26}$$

First, assume that $np \ge 1$. Denote $z_0 = [z] + 1$. Then, for every $t \ge 0$, by (25) we have

$$d(z, \bar{\nu}_n(\cdot), \bar{\pi}_n(\cdot)) \le \mathbf{P}\left(\sum_{i=1}^{n}\zeta_i \ge z_0\right) \le (\mathbf{E}\exp\{t\zeta_1\})^n e^{-tz_0}. \tag{27}$$

The moment generating function in the right-hand side of (27) admits the following estimate:

$$\mathbf{E}\exp\{t\zeta_1\} \le 1 + e^t p^2 + \sum_{k=2}^{\infty} e^{kt}\frac{p^{k+1}}{(k+1)!} = 1 + e^{-t}\left(\exp\{e^t p\} - 1 - e^t p + \frac{e^{2t}p^2}{2}\right)$$

$$\le \exp\left\{\frac{e^t p^2}{2}(\exp\{e^t p\} + 1)\right\}. \tag{28}$$

Put $t = \ln\bigl(\frac{1}{2np^2}\ln(z_0+2)\bigr)$, where we may assume that $\ln(z_0+2) \ge 2np^2$ without losing generality (see (29)). Then (27) and (28) imply the estimate

$$d(z, \bar{\nu}_n(\cdot), \bar{\pi}_n(\cdot)) \le \exp\{e^t np^2 \exp\{e^t p\} - tz_0\}$$

$$= \exp\left\{\frac{1}{2}\ln(z_0+2)(z_0+2)^{1/2np} - z_0\ln\left(\frac{1}{2np^2}\ln(z_0+2)\right)\right\}$$

$$\le (2np^2)^{z_0}\exp\left\{-z_0\ln\ln(z_0+2) + \frac{1}{2}\sqrt{z_0+2}\,\ln(z_0+2)\right\}. \tag{29}$$

¿From here we obtain the first inequality in (20).

Before proving the second inequality in (20), we note that the first holds also in the case $np \ge \delta$ for any $\delta \in (0,1)$. To verify this fact, it suffices to put in (29) $t = \ln\bigl(\frac{\delta}{2np^2}\ln(z_0+2)\bigr)$. Moreover, because of the identity $(np^2)^{k+1} = (np)^k np^{k+2}$, $\delta \le np \le 1$, the second inequality in (20) follows from the first in case $\delta \le np \le 1$. Hence, it suffices to prove the second estimate in (2) under the condition $np \le \delta$, where δ will be chosen later.

Assume that $np \le \delta$. We prove the second inequality in (20) by induction on n. Denote $Q_n(k) := \mathbf{P}\big(\sum_{i=1}^n \zeta_i \ge k\big)$, $k = 1, 2, \ldots$. Then, by the total probability formula, we have

$$Q_n(k) = Q_1(k) + \sum_{m=0}^{k-1} \mathbf{P}(\zeta_1 = m) Q_{n-1}(k-m). \tag{30}$$

To prove the assertion, it suffices to find a positive constant $c < 1$ such that, for all integers $k \ge 2$ and $n \ge 1$, the inequality

$$Q_n(k) \le n(cp)^{k+1} \tag{31}$$

holds which together with the trivial estimate $Q_n(1) \le np^2$ implies (20). For $n = 1$, relation (31) with the constant $c = e/3$ follows from the elementary estimate $Q_1(k) \le p^{k+1}/(k+1)!$ that is true for all $k \ge 2$ (see (26)), and from the Stirling formula (see [15]). Further, let (31) be fulfilled for all $n \le N-1$ with $Np \le \delta$. Then from (30) we obtain

$$\begin{aligned} Q_N(k) \le \frac{p^{k+1}}{(k+1)!} + \sum_{m=2}^{k-2} \frac{p^{m+1}}{(m+1)!}(N-1)(cp)^{k-m+1} \\ +(N-1)(cp)^{k+1} + (N-1)pc^{-1}(cp)^{k+1} + \frac{p}{k!}Np^{k+1}. \end{aligned} \tag{32}$$

Consider separately the second summand in the right-hand side of (32). We have

$$\begin{aligned} \sum_{m=2}^{k-2} \frac{p^{m+1}}{(m+1)!}(N-1)(cp)^{k-m+1} &\le \frac{(N-1)}{6}(cp)^{k+2} \sum_{m=0}^{\infty} \frac{(c^{-1})^m}{m!} \\ &= (cp)^{k+1} \frac{(N-1)cp}{6} e^{1/c}. \end{aligned} \tag{33}$$

Substituting (33) into (32) and taking into account the estimate $p \le \delta N^{-1}$, we obtain

$$Q_N(k) \le N(cp)^{(k+1)} R(N, c, \delta, k), \tag{34}$$

where

$$\begin{aligned} R(N, c, \delta, k) := \frac{1}{Nc^{k+1}(k+1)!} + \frac{(N-1)c\delta}{N^2} e^{1/c} \\ + \frac{N-1}{N} + \frac{(N-1)\delta}{cN^2} + \frac{\delta}{c^{k+1}k!N}. \end{aligned} \tag{35}$$

¿From (34) it follows that the theorem will be proved if we find some absolute constants $e/3 \le c < 1$ and $\delta < 1$ such that the inequality $R(N, \delta, c, k) \le 1$ holds for all $N \ge 2$ and $k \ge 2$,. By the Stirling formula, from (35) we obtain a more convenient upper bound

$$R(N, c, \delta, k) \le 1 - \frac{1}{N}\left[1 - \left(\frac{e}{c(k+1)}\right)^{k+1} \frac{1}{\sqrt{2\pi(k+1)}}\right.$$

$$-\frac{(N-1)\delta}{N}(ce^{1/c}+c^{-1})-\frac{\delta}{c\sqrt{2\pi k}}\left(\frac{e}{ck}\right)^k\Bigg]. \tag{36}$$

Now it is easy to see that, for $c = e/3$ and $\delta = 10^{-1}$, the expression in the square brackets in the right-hand side of (36) is nonnegative for all $k \geq 2$ and $N \geq 2$. The theorem is proved.

Proof of Theorem 2 is based on the following simple inequality. For every coupling of the random processes $\bar{\nu}_n(\cdot)$ and $\bar{\pi}_n(\cdot)$, the following lower bound holds:

$$\mathbf{P}\Big(\max_{k\leq n}\Big|\sum_{i\leq k}\nu_i-\sum_{i\leq k}\pi_i(p)\Big|>z\Big)\geq\mathbf{P}(\pi_1(p)>z+1)$$

$$=\mathbf{P}(\pi_1(p)\geq[z]+2)>\frac{p^{[z]+2}}{([z]+2)!}e^{-p}.$$

It remains merely to use the Stirling formula and to carry out an elementary evaluation of the right-hand side of the inequality obtained. Theorem 2 is proved.

R e m a r k. In the case when $np \to \infty$, the coupling method for the processes $\bar{\nu}_n(\cdot)$ and $\bar{\pi}_n(\cdot)$ used in Theorem 1 could fail to be optimal. For the random variables $\{X_i\}$ that are not too degenerate, there exist better approximations of the sum distributions by the accompanying Poisson laws in terms of the distance $d(z,\cdot)$. The corresponding example was considered in [86].

5 Rate of the χ^2-distance approximation in the Poisson theorem

1. Preliminaries

An approximation of the sum distribution of independent Bernoulli random variables (in the non i.i.d. case we deal with the so-called *generalized binomial distributions*) to the accompanying Poisson law or to other Poisson laws close to the accompanying is one of the classical problems of Probability Theory. This study has a rich semi–centennial prehistory which started with the paper [98] by Yu. V. Prokhorov published in 1953. We do not plan here to review all papers in this direction due to a huge number of such papers. We will mention only the results close to the topic under consideration.

First of all, we note that the most papers in this direction were dedicated to estimating the total variation distance between the above–mentioned distributions (for example, see [5, 8, 82]). In these papers, the corresponding results from [98] were improved and generalized. We also note the recent paper [57], in which the rate of convergence in the classical Poisson theorem was studied in terms of the Kulback – Leibler distance which is stronger than the total variation distance. In particular, by Pinsker's inequality in [57] (see § 2) the authors obtained some refinement of the above-mentioned results for the total variation distance.

In this chapter, we study the so-called χ^2-distance between a generalized (or a classical) binomial distribution and the accompanying Poisson law. As a consequence of Theorem 4.1 of the chapter, we improve the corresponding result in [57] concerning the estimating of the total variation distance.

The chapter has the following structure. In § 2 we recall the definitions of three probability distances and their properties we need. In § 3 we study an asymptotic form of χ^2-distance for a generalized binomial distribution (Theorem 3.1). In § 4 we obtain asymptotic expansions and two-sided estimates for this distance in the case of a classical binomial distribution. In § 5 we compare the upper bounds obtained in § 4 with the corresponding results of predecessors in the case of a classical binomial distribution.

2. Main definitions and notation

Let P and Q be some distributions on the set $\mathbf{Z}_+ := \{0, 1, 2 \ldots\}$. Introduce the following distances in the space of all such distributions:

1. *The total variation distance* between P and Q (see [1, 2]):

$$\|P - Q\| := \sum_{i=0}^{\infty} |P(i) - Q(i)|.$$

Sometimes (see [5, 8]) this distance is defined as $\sup_{A \subset \mathbf{Z}_+} |P(A) - Q(A)|$ which is half the value $\|P - Q\|$, i. e., half the total variation of the signed finite measure $P - Q$.

2. *Information divergence* or *the Kulback – Leibler distance* between P and Q:

$$D(P, Q) := \sum_{i=0}^{\infty} P(i) \log \frac{P(i)}{Q(i)}.$$

3. χ^2*-distance* between P and Q:

$$\chi^2(P, Q) := \sum_{i=0}^{\infty} \frac{(P(i) - Q(i))^2}{Q(i)} = \sum_{i=0}^{\infty} Q(i) \left(\frac{P(i)}{Q(i)} - 1 \right)^2 = \sum_{i=0}^{\infty} P(i) \frac{P(i)}{Q(i)} - 1.$$

In the case when the supports of the distributions P and Q does not coincide with the set $\mathbf{Z}_+$, in the items 2 and 3, we put that $0/0 = 0$, $a/0 = \infty$ for $a > 0$, $0 \cdot \log(0) = 0$ and $\log(\infty) = \infty$. Hence, on the set of all distributions, we can define these distances with values on the extended positive half–line. If the support of $P(\cdot)$ is finite and it is included into the support of $Q(\cdot)$ then, under the above–mentioned agreement regarding the arithmetical operations, the values of $\chi^2(P, Q)$ and $D(P, Q)$ are finite. Notice that the distances $\chi^2(P, D)$ and $D(P, Q)$ (in contrast with $\| \cdot \|$) are not metrics. They do not possess the symmetry, and, in general, do not satisfy the triangle inequality.

These three distances are connected by the following inequalities:

(a) Cauchy — Bunyakovskii inequality

$$\|P - Q\|^2 \le \chi^2(P, Q);$$

(b) Pinsker's inequality (see [5, 6])

$$\|P - Q\|^2 \le 2D(P, Q);$$

(c) The inequality

$$D(P, Q) \le \chi^2(P, Q)$$

follows from the simple estimate $\log x \le x - 1$ for all $x > 0$.

Thus, χ^2-distance is the strongest of these three distances. Therefore, it seems that it is more logical to evaluate the total variation distance by the Kulback – Leibler distance (as in [5]). However, χ^2-distance has a simpler structure than that of $D(P,Q)$. Moreover, in spite of coincidence of the leading terms of the asymptotic expansions of the right–hand sides in (a) and (b) (say, in the conditions of the classical Poisson theorem), limit behavior of the smaller terms in these expansions differ in order: For χ^2-distance this term is asymptotically better than that for $D(P,Q)$ (see § 5). In fact, this phenomenon provides an improvement of the upper bounds for the total variation distance obtained in [57] by Pinsker's inequality.

In the sequel, we will use the following notation:

1) $\xi_1,\ldots,\xi_n$ are independent Bernoulli random variables with the respective success probabilities $p_1,\ldots,p_n$;

2) η is a Poisson random variable with the parameter $\lambda := \sum\limits_{i=1}^{n} p_i$;

3) P is the distribution of the sum $\sum\limits_{i=1}^{n}\xi_i$, and Q is the distribution of η (the accompanying Poisson distribution);

4) $p := \sum\limits_{i=0}^{n} p_i^2/\lambda$; in particular, in the case of identically distributed Bernoulli random variables, the value p coincides with the corresponding success probability;

5) $\tilde{p}_i := \frac{p_i}{1-p_i}$, $\tilde{\lambda} := \sum\limits_{i=1}^{n}\tilde{p}_i$, $\tilde{p} := \sum\limits_{i=0}^{n}\tilde{p}_i^2/\tilde{\lambda}$.

We agree to interpret the limit relations like $\to$, $\sim$, o, O, and lim as those as $n\to\infty$. And a limit relation of the type $\varphi_n = O(f(\lambda,p))$ for positive φ_n and $f(\cdot)$ means that the the upper limit of the corresponding ratio can be evaluated by an absolute positive constant. In the paper, such constants will be denoted by the symbols c or c_i. Dependence of constants on some parameters will be denoted by the corresponding arguments in the recording $c(\cdot)$.

3. Approximation to a generalized binomial distribution

The main goal of this Section is to prove the following assertion.

Theorem 3.1 *Let $\lambda^7 p\to 0$. Then*

$$\frac{\chi^2(P,Q)}{p^2}\to\frac{1}{2}.$$

Proof To study the asymptotics of $\chi^2(P,Q)$-distance, for $\Delta(k) := \frac{(P(k)-Q(k))^2}{Q(k)^2}$, we obtain a two-sided estimate of the form

$$A(k)+C(k)\le\Delta(k)\le A(k)+B(k),$$

where the expectations $\mathbf{E}(C(\eta))$ and $\mathbf{E}(B(\eta))$ should be much smaller than the expectation $\mathbf{E}(A(\eta))$ which should be easy calculated. Then we will obtain the asymptotic representation

$$\chi^2(P,Q)=\mathbf{E}(\Delta(\eta))\sim\mathbf{E}(A(\eta)).$$

The corresponding construction consists of several stages.

Lemma 3.1. *Let* $\lambda p/2 + \sum\limits_{i=1}^{n} \tilde{p}_i{}^3/3 \le 1$. *Then, for every* $k \in [0, n]$, *the following two–sided inequality is valid:*

$$R_1 R_2 \le \frac{P(k)}{Q(k)} \le R_1^* R_2^*,$$

where $R_1 := 1 - \lambda p/2 - \sum\limits_{i=1}^{n} \tilde{p}_i^3/3$, $R_1^* := 1 - \lambda p/2 + (\lambda p)^2/8$, $R_2 := (\tilde{\lambda}/\lambda)^k (1 - \tilde{\lambda}^{-1} C_k^2 \tilde{p})$,

$$R_2^* := R_2 + \left(\frac{\tilde{\lambda}}{\lambda}\right)^k \big(C_k^4 + 3C_k^3\big) \max\left(\frac{\sum\limits_{i=1}^{n} \tilde{p}_i{}^3}{\tilde{\lambda}^3}, (\tilde{p}/\tilde{\lambda})^2 \right),$$

and, by definition, $C_k^m := k(k-1)\ldots(k-m+1)/m! = 0$ *for each* $k < m$.

Proof For every integer $k \le n$, we have

$$\frac{P(k)}{Q(k)} = k! \sum_{i_1 < \cdots < i_k \le n} \lambda^{-k} \tilde{p}_{i_1} \tilde{p}_{i_2} \ldots \tilde{p}_{i_k} (1 - p_1)(1 - p_2) \ldots (1 - p_n) e^{\lambda} = F_1 F_2, \qquad (1)$$

where $F_1 := (1 - p_1)(1 - p_2) \ldots (1 - p_n) e^{\lambda}$, $F_2(k) := \lambda^{-k} \sum\limits_{i_1 \ne \cdots \ne i_k \le n} \tilde{p}_{i_1} \tilde{p}_{i_2} \ldots \tilde{p}_{i_k}$. If we additionally put $F_2(0) := 1$ then the identity (1) holds for all $0 \le k \le n$.

We first evaluate F_1. By Taylor's formula we obtain that

$$\log(F_1) = \lambda + \sum_{i=1}^{n} \log(1 - p_i) = -\sum_{i=1}^{n} \frac{p_i^2}{2} - \sum_{i=1}^{n} \frac{p_i^3}{3} \frac{1}{(1 - \theta_i p_i)^3},$$

where $0 < \theta_i < 1$ for all i. Therefore,

$$-\sum_{i=1}^{n} \frac{p_i^2}{2} - \sum_{i=1}^{n} \frac{\tilde{p}_i^3}{3} \le \log(F_1) \le -\sum_{i=1}^{n} \frac{p_i^2}{2}.$$

Using the elementary two–sided inequality $1 - x \le e^{-x} \le 1 - x + x^2/2$ for $x \ge 0$ we deduce from here the two–sided estimate

$$1 - \frac{\lambda p}{2} - \sum_{i=1}^{n} \frac{\tilde{p}_i^3}{3} \le F_1 \le 1 - \frac{\lambda p}{2} + \frac{1}{8} (\lambda p)^2.$$

We now construct upper and lower bounds for $F_2(k)$ in (1) for all $k \le n$. Since $F_2(0) = 1$ and $F_2(1) = \tilde{\lambda}/\lambda$, i. e., for $k = 0, 1$, the estimates from the Lemma are valid, it suffices to study the case $2 \le k \le n$ only. Denote by $S := \{1, 2, \ldots, n\}^k$ the set of multi-indices $(i_1, i_2, \ldots, i_k)$ such that every coordinate i_m takes all the values from 1 to n. Introduce the following finite measure μ defined on subsets of S, by the formula

$$\mu(S') := \sum_{s \in S'} \tilde{p}(s),$$

where $S' \subseteq S$, $\tilde{p}(s) := \tilde{p}_{i_1}\tilde{p}_{i_2}\dots\tilde{p}_{i_k}$ if $s = (i_1, i_2, \dots, i_k)$; here, by definition, $\mu(\emptyset) = 0$. Then $F_2(k) = \lambda^{-k}\mu(S^*)$, where $S^* = \{(i_1, i_2, \dots, i_k) \in S : i_1 \neq i_2 \dots \neq i_k\}$. Put $S_{l,m} = \{(i_1, i_2, \dots, i_k) \in S,\ i_l = i_m\}$ for $0 < l < m \leq k$, and $S_{l,m} = \emptyset$ otherwise. It is clear that $S^* = S \setminus \cup_{l,m} S_{l,m}$. Therefore, by additivity of μ, the following inequalities are valid:

$$\mu(S^*) \geq \mu(S) - \sum_{l,m} \mu(S_{l,m}) = \tilde{\lambda}^k - C_k^2 \mu(S_{1,2}),$$

$$\mu(S^*) \leq \mu(S) - \sum_{l,m} \mu(S_{l,m}) + \frac{1}{2} \sum_{(l,m)\neq(i,j)} \mu(S_{l,m} \cap S_{i,j}) =: \mu^*. \tag{2}$$

Notice that, for $k = 2$, these two inequalities are reduced to the equalities since, in this case, the sum $\sum\limits_{(l,m)\neq(i,j)}$ vanishes (as a sum over the empty set of indices). Let $l < m$. Then

$$\mu(S_{l,m}) = \mu(S_{1,2}) = \sum_{i=1}^{n} \tilde{p}_i^2 \left(\sum_{i=1}^{n} \tilde{p}_i \right)^{k-2} = \frac{\sum\limits_{i=1}^{n} \tilde{p}_i^2}{\tilde{\lambda}^2} \tilde{\lambda}^k.$$

Thus, an lower bound for $F_2(k)$ has the form

$$F_2(k) = \mu(S^*)\lambda^{-k} \geq R_2.$$

In the case $k \geq 3$, to estimate the value μ^* in (2), calculating the measure μ of all pair–wise intersections of the subsets $S_{i,j}$ we note that, in the corresponding sum, the values of these measures are distinguished only in the two cases when either the pairs (i,j) and (l,m) have one identical coordinate or there are no such coordinates.

Respectively, in the first case, we have

$$\mu(S_{l,m} \cap S_{i,j}) = \mu(S_{1,2} \cap S_{2,3}) = \mu(S_{1,2} \cap S_{1,3}) = \mu(S_{1,3} \cap S_{2,3})$$

$$= \sum_{i=1}^{n} \tilde{p}_i^3 \left(\sum_{i=1}^{n} \tilde{p}_i \right)^{k-3} = \frac{\sum\limits_{i=1}^{n} \tilde{p}_i^3}{\tilde{\lambda}^3} \tilde{\lambda}^k, \tag{3}$$

and the number of such summands in the corresponding sum equals $6C_k^3$; in the second case,

$$\mu(S_{l,m} \cap S_{i,j}) = \mu(S_{1,2} \cap S_{3,4}) = \left(\sum_{i=1}^{n} \tilde{p}_i^2 \right)^2 \left(\sum_{i=1}^{n} \tilde{p}_i \right)^{k-4} = \left(\frac{\sum\limits_{i=1}^{n} \tilde{p}_i^2}{\tilde{\lambda}^2} \right)^2 \tilde{\lambda}^k \tag{4}$$

and the number of such summands in the sum above equals $2C_k^4$.

Denote $\alpha := \max \left(\tilde{\lambda}^{-3} \sum\limits_{i=1}^{n} \tilde{p}_i^3, \left(\tilde{\lambda}^{-1} \tilde{p} \right)^2 \right)$. From (2)–(4) we then finally obtain that

$$F_2(k) \leq \lambda^{-k}\mu^* = R_2 + 2^{-1}\lambda^{-k} \sum_{(l,m)\neq(i,j)} \mu(S_{l,m} \cap S_{i,j}) \leq R_2 + (\tilde{\lambda}/\lambda)^k \big(C_k^4 + 3C_k^3\big)\alpha = R_2^*.$$

Under the conditions of the Lemma, we have $R_1 \geq 0$, $R_1^* \geq 0$, $R_2^* \geq 0$. Although the value R_2 may be negative, the two–sided inequality $R_1R_2 \leq F_1F_2 \leq R_1^*R_2^*$ is valid. The Lemma is proven.

Lemma 3.2. *Let $\lambda p \to 0$. Then the following asymptotic relations are valid:*

1) $\sum\limits_{i=1}^{n} \tilde{p}_i^m/p = O((\lambda p)^{(m-2)/2})$ *for all* $m > 2$;

2) $\tilde{\lambda} \sim \lambda$, $\tilde{p} \sim p$;

3) $\sum \frac{p_i^2}{1-p_i} = \lambda p(1 + O((\lambda p)^{1/2}))$;

4) $\tilde{\lambda}\tilde{p} = \lambda p + O((\lambda p)^{3/2})$;

5) $\tilde{\lambda}^m - \lambda^m = m\lambda^m p(1 + O(m2^m(\lambda p)^{1/2}))$ *for each* $m \geq 1$.

Proof. Notice that, under the conditions of the Lemma, $\max\limits_{i \leq n} p_i \leq (\lambda p)^{1/2} \to 0$. It implies assertions 1–4.

To prove relation 5 we note that, by relation 3,

$$\tilde{\lambda} - \lambda = \sum_{i=1}^{n} (\tilde{p}_i - p_i) = \sum_{i=1}^{n} \left(\frac{p_i^2}{1-p_i} \right) = \lambda p(1 + O((\lambda p)^{1/2})) =: \delta.$$

Therefore, for every fixed $m \geq 2$,

$$\tilde{\lambda}^m - \lambda^m = (\lambda + \delta)^m - \lambda^m = \sum_{i=1}^{m} C_m^i \lambda^{m-i} \delta^i = \lambda^m \sum_{i=1}^{m} C_m^i p^i (1 + O(m(\lambda p)^{1/2}))$$

$$= m\lambda^m p(1 + O(m2^m(\lambda p)^{1/2}))$$

since

$$\sum_{i=2}^{m} C_m^i p^i \leq C_m^2 p^2 \sum_{i=2}^{m} C_{m-2}^{i-2} p^{i-2} = C_m^2 p^2 (1+p)^{m-2},$$

and, by the trivial estimates $p \leq \lambda$ and $p \leq 1$, we derive that, under the conditions of the Lemma, the right–hand side of the last inequality has the order $O(pm^2 2^m (\lambda p)^{1/2})$. The Lemma is proven.

Further we need the following simple properties of Poisson distributions.

Lemma 3.3 Let ζ_1 and ζ_2 be Poissonian random variables with arbitrary parameters μ_1 and μ_2 respectively. Then, for every function $f(k)$, $k \in \mathbf{Z}_+$ module of which increases faster than an exponential as $k \to \infty$, the following equality is valid:

$$\mathbf{E}(\mu_2/\mu_1)^{\zeta_1} f(\zeta_1) = e^{\mu_2 - \mu_1} \mathbf{E} f(\zeta_2).$$

Moreover, for every natural m,

$$\mathbf{E} C_{\zeta_2}^m = \frac{\mu_2^m}{m!}, \quad \mathbf{E}(C_{\zeta_2}^m)^2 = \frac{\mu_2^m}{(m!)^2} \mathbf{E}(\zeta_2 + 1) \ldots (\zeta_2 + m) \leq c(m) \mu_2^m (1 + \mu_2^m).$$

We now separately extract a fragment in the proof of Theorem 3.1.

Lemma 3.4. *Let the functions $A(k)$, $C(k)$, $S(k)$, and $B(k)$ defined on $\mathbf{Z}_+$ satisfy the following conditions:*

$$A(k)+C(k)\le S(k)\le A(k)+B(k),$$

$$\mathbf{E}(B(\eta))^2=o(p^2),\ \mathbf{E}(C(\eta))^2=o(p^2),\ \mathbf{E}|B(\eta)A(\eta)|=o(p^2),\ \mathbf{E}|C(\eta)A(\eta)|=o(p^2)$$

for $p\to 0$. Then $\mathbf{E}(S(\eta))^2=\mathbf{E}(A(\eta))^2+o(p^2)$.

Proof. Consider the function $D(k):=S(k)-A(k)$. It is clear that

$$\mathbf{E}(S(\eta))^2=\mathbf{E}(A(\eta))^2+2\mathbf{E}D(\eta)A(\eta)+\mathbf{E}(D(\eta))^2.$$

In conclusion, we use the estimate

$$|D(k)|\le|C(k)|+|B(k)|$$

and the conditions of the Lemma as well. The Lemma is proven.

We now begin to prove Theorem 3.1. First we note that, from the inequalities $p\le\lambda$ and $\lambda^7p\ge\lambda^bI(\lambda\ge 1)p$ for every $b\in[0,7]$, where $I(\cdot)$ is the indicator function, it follows the relation $\lambda^bp\to 0$ for each b from the interval above. In particular, the condition of the Theorem implies the condition of Lemma 3.2. Moreover, since $\max\limits_{i\le n}p_i\le(\lambda p)^{1/2}$ then $\sum\limits_{i=1}^{n}\tilde p_i^3\le(\lambda p)^{3/2}$. Therefore, if the value λp is small enough then the condition of Lemma 3.1 is fulfilled.

Introduce the following notation:

$$S(k):=\frac{P(k)}{Q(k)}-1,\quad A(k):=\left(\frac{\tilde\lambda}{\lambda}\right)^k\left(1-\frac{\lambda p}{2}-\tilde\lambda^{-1}C_k^2\tilde p\right)-1,\tag{5}$$

$$B(k):=R_1^*R_2^*-1-A(k),\quad C(k):=R_1R_2-1-A(k),$$

where R_1, R_2, R_1^*, and R_2^* are defined in Lemma 3.1. Notice that, under the above–introduced notation,

$$\chi^2(P,Q)=\mathbf{E}(S(\eta))^2.$$

Moreover, due to the Lemma 3.1, for λp small enough, the following two–sided estimate is valid:

$$A(k)+C(k)\le S(k)\le A(k)+B(k).$$

Put $\tilde\alpha:=\max\left(\tilde\lambda^{-1}\sum\limits_{i=1}^{n}\tilde p_i^3,\tilde p^2\right)$. Notice that, due to Lemma 3.2 and the conditions of the Theorem, $\tilde\alpha=o(p)$. It is easy to obtain that, for λp small enough, we have

$$B(k)^2\le\left(\frac{\tilde\lambda}{\lambda}\right)^{2k}c_0\big((\lambda p)^4+\big(C_k^2\tilde\lambda^{-1}p\big)^2(\lambda p)^2+\big(C_k^4+C_k^3\big)^2\tilde\lambda^{-4}\tilde\alpha^2\big)=:\left(\frac{\tilde\lambda}{\lambda}\right)^{2k}f(k).$$

Under the conditions of Lemma 3.3, we put $\zeta_1 := \eta$ (. . $\mu_1 := \lambda$) $\mu_2 := \tilde{\lambda}^2/\lambda$. Then

$$\mathbf{E}B(\eta)^2 \le \mathbf{E}\left(\frac{\tilde{\lambda}}{\lambda}\right)^{2\eta} f(\eta) = \exp\left(\frac{\tilde{\lambda}^2}{\lambda} - \lambda\right)\mathbf{E}f(\zeta_2). \tag{6}$$

Moreover,

$$\begin{aligned} \tilde{\lambda}^{-2}\mathbf{E}\big(C^2_{\zeta_2}\big)^2 &\le c_1(\tilde{\lambda}/\lambda)^2(1+(\tilde{\lambda}/\lambda)^4\lambda^2), \\ \tilde{\lambda}^{-4}\mathbf{E}\big(C^3_{\zeta_2}\big)^2 &\le c_1(\tilde{\lambda}/\lambda)^2(1+(\tilde{\lambda}/\lambda)^6\lambda^3)/\lambda, \\ \tilde{\lambda}^{-4}\mathbf{E}\big(C^4_{\zeta_2}\big)^2 &\le c_1(\tilde{\lambda}/\lambda)^4(1+(\tilde{\lambda}/\lambda)^8\lambda^4). \end{aligned} \tag{7}$$

We also note that the simple relation $\tilde{\alpha} = O(p \max_i p_i)$ and the estimate $\max_i p_i \le (\lambda p)^{1/2}$ imply the asymptotic relation $\tilde{\alpha}^2 = O(\lambda p^3)$. Therefore, under the conditions of the Theorem,

$$\begin{gathered} \tilde{\alpha}^2\tilde{\lambda}^{-4}\mathbf{E}\big(C^3_{\zeta_2}\big)^2 = O((\lambda p)^3) = o(p^2), \quad \tilde{\alpha}^2\tilde{\lambda}^{-4}\mathbf{E}\big(C^4_{\zeta_2}\big)^2 = O(\lambda^5 p^3) = o(p^2), \\ \tilde{\lambda}^{-2}\mathbf{E}\big(C^2_{\zeta_2}\big)^2 p^2(\lambda p)^2 = O(p^2((\lambda p)^2 + \lambda^4 p^2)) = o(p^2). \end{gathered} \tag{8}$$

Thus, we finally obtain the asymptotic relation

$$\mathbf{E}B(\eta)^2 = o(p^2). \tag{9}$$

For λp small enough, the estimates derived above imply the inequality (recall that $\tilde{\lambda} > \lambda$)

$$|B(k)A(k)| \le \left(\left(\frac{\tilde{\lambda}}{\lambda}\right)^k - 1 + \left(\frac{\tilde{\lambda}}{\lambda}\right)^k \left(\frac{\lambda p}{2} + C^2_k \tilde{p}/\tilde{\lambda}\right)\right)\left(\frac{\tilde{\lambda}}{\lambda}\right)^k \sqrt{f(k)} = B_1(k) + B_2(k) + B_3(k),$$

where

$$\begin{gathered} B_1(k) = (\lambda p/2)(\tilde{\lambda}/\lambda)^{2k}\sqrt{f(k)}, \quad B_2(k) = \tilde{p}(\tilde{\lambda}/\lambda)^{2k}\tilde{\lambda}^{-1}C^2_k\sqrt{f(k)}, \\ B_3(k) = ((\tilde{\lambda}/\lambda)^k - 1)(\tilde{\lambda}/\lambda)^k\sqrt{f(k)}. \end{gathered}$$

Hence,

$$\mathbf{E}|B(\eta)A(\eta)| \le \mathbf{E}B_1(\eta) + \mathbf{E}B_2(\eta) + \mathbf{E}B_3(\eta). \tag{10}$$

Evaluation of the three expectations on the right–hand side of (10) is carry out by Lemma 3.3 using an approach of the same kind. To prove the first summand on the right–hand side of (10) we use the inequality (6), where we substitute $\sqrt{f}$ for f, and further we apply the Cauchy — Bunyakovskii inequality. We then have

$$\mathbf{E}B_1(\eta) \le \frac{\lambda p}{2}\exp\left(\frac{\tilde{\lambda}^2}{\lambda} - \lambda\right)(\mathbf{E}f(\zeta_2))^{1/2}. \tag{11}$$

Using the asymptotic formulas in (8) (in terms of the symbol $O(\cdot)$) we finally obtain the following asymptotic relation for the expectation $\mathbf{E}f(\zeta_2)$:

$$\mathbf{E}B_1(\eta) = O((\lambda p)^{5/2} + \lambda^{7/2}p^{5/2} + \lambda^2 p^3) = O(p^2(\sqrt{\lambda^7 p} + \sqrt{\lambda^5 p} + \lambda^2 p)) = o(p^2). \tag{12}$$

Evaluation of the second summand on the right–hand side of (10) is absolutely the same. To do it we use inequality (6) in which we substitute $C_k^2\sqrt{f(k)}$ for $f(k)$ and apply the Cauchy — Bunyakovskii inequality and the asymptotic relations in (7) and (8) for expectations of the corresponding components.

To evaluate the last summand on the right–hand side of (10) we first note that, due to Lemma 3.3 and the Cauchy — Bunyakovskii inequality, we have

$$\mathbf{E}B_3(\eta) \le (\mathbf{E}((\tilde\lambda/\lambda)^\eta - 1)^2)^{1/2}(\mathbf{E}(\tilde\lambda/\lambda)^{2\eta}f(\eta))^{1/2}$$

$$= (\exp\{\tilde\lambda^2/\lambda - \lambda\} - 2\exp\{\tilde\lambda - \lambda\} + 1)^{1/2}(\mathbf{E}f(\zeta_2))^{1/2}\exp(1/2)(\tilde\lambda^2/\lambda - \lambda)\}. \qquad (13)$$

Consider the first factor on the right–hand side of (13). Using Taylor's expansions at zero for these two exponential up to terms of the second order as well as the relation

$$\tilde\lambda^2/\lambda - \lambda - 2(\tilde\lambda - \lambda) = (\tilde\lambda - \lambda)^2/\lambda = \lambda p^2(1 + o(1)),$$

which follows from Lemma 3.2, we obtain the following asymptotic representation:

$$(\exp\{\tilde\lambda^2/\lambda - \lambda\} - 2\exp\{\tilde\lambda - \lambda\} + 1)^{1/2} = O(\lambda^{1/2}p + \lambda p).$$

In other words, the estimating of (13) is reduced to (11) and (12).

So, $\mathbf{E}B_i(\eta) = o(p^2)$ for all $i = 1, 2, 3$, and hence,

$$\mathbf{E}|B(\eta)A(\eta)| = o(p^2). \qquad (14)$$

By analogy with the foregoing we prove the relations

$$\mathbf{E}C(\eta)^2 = o(p^2), \quad \mathbf{E}|C(\eta)A(\eta)| = o(p^2)$$

which together with (9), (14), and Lemma 3.4 imply

$$\lim \frac{\chi^2(P,Q)}{p^2} = \lim \frac{\mathbf{E}S(\eta)^2}{p^2} = \lim \frac{\mathbf{E}A(\eta)^2}{p^2}.$$

It remains to prove the limit relation $\lim \mathbf{E}A(\eta)^2/p^2 = 1/2$. To prove this fact we apply Lemmas 3.2 and 3.3 once more as well as the above–described technique of the asymptotic analysis. In contrast to the previous calculations, we use the upper bounds for the remainders in the corresponding assertions of Lemma 3.2 since we need an exact asymptotics of the expectation $\mathbf{E}A(\eta)^2$. For convenience of the reader we extract the assertions from Lemmas 3.2 and 3.3 we need:

$$\frac{\tilde\lambda^2}{\lambda} - \lambda = 2\lambda p(1 + O((\lambda p)^{1/2})), \quad \tilde\lambda - \lambda = \lambda p(1 + O((\lambda p)^{1/2})),$$
$$\tilde\lambda\tilde p = \lambda p(1 + O((\lambda p)^{1/2})), \quad \mathbf{E}\zeta(\zeta - 1) = \mu^2, \quad \mathbf{E}(\zeta(\zeta-1))^2 = \mu^4 + 4\mu^3 + 2\mu^2, \qquad (15)$$

where ζ is a Poissonian random variable with parameter μ. Notice that the last two formulas in (15) will be applied below in the two cases: For $\zeta = \zeta_2$ (i. e., for $\mu = \tilde{\lambda}^2/\lambda$) and for $\zeta = \tilde{\eta}$ (i. e., for $\mu = \tilde{\lambda}$).

First, by Lemma 3.3 we represent the second moment in such a way:

$$\mathbf{E}A(\eta)^2 = A_1 + A_2 + A_3,$$

where

$$A_1 := \exp(\tilde{\lambda}^2/\lambda - \lambda) - 2\exp(\tilde{\lambda} - \lambda) + 1,$$
$$A_2 := \exp(\tilde{\lambda} - \lambda)(p\lambda + \tilde{\lambda}\tilde{p}) - \exp(\tilde{\lambda}^2/\lambda - \lambda)(p\lambda + (\tilde{\lambda}^2/\lambda^2)\tilde{\lambda}\tilde{p}),$$
$$A_3 := \exp(\tilde{\lambda}^2/\lambda - \lambda)\mathbf{E}(\lambda p/2 + \tilde{\lambda}^{-1}\tilde{p}C^2_{\zeta_2})^2.$$

To evaluate A_1 we use Taylor's expansions at zero for both the exponentials up to terms of the third order and apply the relations from (15). We then have

$$A_1 = (\tilde{\lambda} - \lambda)^2/\lambda + (1/2)(\tilde{\lambda}^2/\lambda - \lambda) - (\tilde{\lambda} - \lambda)^2 = \lambda p^2 + (\lambda p)^2 + o(p^2).$$

By analogy, using the above–mentioned Taylor's expansions up to terms of the second order of smallness, we obtain the following asymptotic representation for A_2:

$$A_2 = -2(\lambda p^2 + (\lambda p)^2) + o(p^2).$$

Evaluate A_3. By (15) we have

$$A_3 = \exp\left(\frac{\tilde{\lambda}^2}{\lambda} - \lambda\right)\left(\frac{(\lambda p)^2}{4} + \frac{1}{2}\lambda p\tilde{\lambda}\tilde{p}\left(\frac{\tilde{\lambda}}{\lambda}\right)^2 + \frac{1}{4}\left(\frac{\tilde{\lambda}}{\lambda}\right)^4(\tilde{\lambda}\tilde{p})^2 + \left(\frac{\tilde{\lambda}}{\lambda}\right)^3\tilde{\lambda}\tilde{p}^2 + \frac{1}{2}\left(\frac{\tilde{\lambda}}{\lambda}\right)^2\tilde{p}^2\right)$$
$$= (\lambda p)^2 + \lambda p^2 + p^2/2 + o(p^2).$$

Thus, we finally obtain

$$\mathbf{E}A(\eta)^2 = p^2/2 + o(p^2).$$

Theorem 3.1 is proven.

4. Two–sided estimates of approximation to a classical binomial distribution.

In the sequel we put $p_i = p$, $i = 1, \ldots, n$. In this case, $\lambda = np$

$$P(k) = \frac{n!}{(n-k)!k!}(1-p)^{n-k}p^k, \quad Q(k) = \frac{\lambda^k}{k!}e^{-\lambda}.$$

The main result of this Section is as follows.

Theorem 4.1 *For arbitrary p and n, the following two–sided estimate is valid*:

$$\frac{p^2}{2}+\frac{2p^3}{3n}\le \chi^2(P,Q)\le \frac{p^2}{2}+\frac{2p^3}{3n}+\frac{p^4}{1-p}+\frac{p^8(23-20p)}{(1-p)^2}. \qquad (16)$$

Proof To construct the estimates from (16) we first reduce $\chi^2(P,Q)$-distance to the form $\sum\limits_{m\ge 0} B_m p^m$ and evaluate the coefficients B_m.

Lemma 4.1 *The following representation is valid*:

$$\chi^2(P,Q)+1=\sum_{m\ge 0}\frac{m!}{n^m}A_m^2 p^m, \qquad (17)$$

where $A_m := \sum\limits_{i=0}^{m}\frac{n^i(-1)^{m-i}}{i!}C_n^{m-i}$.

Proof We agree that if, for some summation indices, the range of summation is not indicated then the corresponding sum is taken over the set $\mathbf{Z}$ of all integers. Meanwhile, in the formulas below, we put for convenience that $(-k)! = \infty$ for all natural k that is equivalent to the above–mentioned in § 3 agreement on the equality $C_k^m = 0$ for all natural $k < m$ (of course, under the standard agreement concerning the arithmetical operations on the extended real line). Notice that the analogous interpretation of binomial coefficients is contained in [51].

Using the corresponding representation for $\chi^2(P,Q)$ from § 2 we have

$$\chi^2(P,Q)+1=\sum_{k=0}^{n}\frac{(n!)^2p^k}{((n-k)!)^2k!n^k}(1-p)^{n-k}(1-p)^{n-k}e^{np}$$

$$=\sum_{k=0}^{n}\frac{(n!)^2p^k}{((n-k)!)^2k!n^k}\sum_{i,j,l}\left((-1)^ip^i\frac{(n-k)!}{(n-k-i)!i!}\right)\left((-1)^jp^j\frac{(n-k)!}{(n-k-j)!j!}\right)\frac{n^lp^l}{l!}$$

$$=\sum_{k}\sum_{i,j,l}\frac{(n!)^2(-1)^{i+j}p^{i+j+l+k}n^l}{k!n^k(n-k-i)!i!(n-k-j)!j!l!}=: R_1.$$

In the last multiple sum, we change the four summation variables by the formulas $i = d + r - q$, $j = d$, $k = m - r - d$ and $l = q - d$. It is easy to verify that this linear transform of the variables is a bijection from $\mathbf{Z}^4$ to $\mathbf{Z}^4$. Then

$$R_1=\sum_{m,r,q,d}\frac{(n!)^2(-1)^{2d+r-q}p^mn^{r+q-m}}{(m-r-d)!(n-m+q)!(d+r-q)!(n-m+r)!d!(q-d)!}.$$

In the multiple sum of the last representation for R_1, we may consider the variables r and q to be not exceeding m, since, otherwise, at least one of the factorials in the fraction denominator equals $-\infty$.

Therefore, R_1 admits the following refinement:

$$R_1 = \sum_{m\geq 0}\sum_{r,q=0}^{m}\sum_{d}\frac{(n!)^2(-1)^{2d+r-q}p^m n^{r+q-m}}{(m-r-d)!(n-m+q)!(a+r-q)!(n-m+r)!d!(q-d)!}$$

$$= \sum_{m\geq 0}\sum_{r,q=0}^{m}\left(\frac{(n!)^2(-1)^{2d+r-q}p^m n^{r+q-m}m!}{(n-m+q)!(n-m+r)!r!q!(m-r)!(m-q)!}\sum_{d}\frac{C_q^d C_{m-q}^{m-r-d}}{C_m^r}\right);$$

meanwhile, due to the definition of a hypergeometric distribution,

$$\sum_{d}\frac{C_q^d C_{m-q}^{m-r-d}}{C_m^r} = 1.$$

Thus,

$$\chi^2(P,Q)+1 = \sum_{m\geq 0}p^m\frac{m!}{n^m}\sum_{r,q=0}^{m}\frac{(n!)^2(-1)^{r+q}n^{r+q}}{(n-m+q)!(n-m+r)!r!q!(m-r)!(m-q)!}$$

$$= \sum_{m\geq 0}p^m\frac{m!}{n^m}\left(\sum_{r=0}^{m}\frac{n!(-1)^r n^r}{(n-m+r)!r!(m-r)!}\right)\left(\sum_{q=0}^{m}\frac{n!(-1)^q n^q}{(n-m+q)!j!(m-q)!}\right)$$

$$= \sum_{m\geq 0}p^m\frac{m!}{n^m}A_m^2,$$

where the coefficients A_m are defined in (17). The assertion is proven.

The next Lemma describes a simple method to calculate the coefficients A_m.

Lemma 4.2 *For the coefficients A_m in* (17), *for each natural n, the following recurrent relation is valid*:

$$A_{m+2} = -\frac{n}{m+2}\sum_{i=0}^{m}A_i \quad \text{for all} m\geq 0.$$

Proof. First of all, we compute the generating function for the sequence $\{A_m;\ m\geq 0\}$ defined in (17) by the formula

$$A_m = \sum_{i=0}^{m}\frac{n^i}{i!}(-1)^{m-i}C_n^{m-i}.$$

By this formula, changing the summation order, we obtain the identity

$$\sum_{m=0}^{\infty}p^m A_m = \sum_{i=0}^{\infty}\frac{(np)^i}{i!}\sum_{m=i}^{n+i}(-p)^{m-i}C_n^{m-i} = (1-p)^n e^{np} \tag{18}$$

which is valid for all $p \in (0,1)$. We now differentiate in p both the parts of this equality (the termwise differentiation of the series in (18) is correct in a neighborhood of every point of the interval $(0,1)$). We then obtain the equality

$$\sum_{m=0}^{\infty} p^m A_{m+1}(m+1) = n(1-p)^n e^{np} - n(1-p)^{n-1}e^{np} = -n(1-p)^n e^{np}\frac{p}{1-p}$$

$$= -\sum_{m=0}^{\infty} p^m A_m \frac{np}{1-p} = -n\sum_{m=0}^{\infty} p^{m+1} A_m \frac{1}{1-p} = -n\sum_{m=1}^{\infty} p^m \left(\sum_{i=0}^{m-1} A_i\right)$$

which implies the relation we need:

$$A_{m+1}(m+1) = -n\sum_{i=0}^{m-1} A_i.$$

The lemma is proven.

Calculate a few first coefficients A_i:

$$\begin{gathered} A_0 = 1, \quad A_1 = 0, \quad A_2 = -\frac{n}{2}, \quad A_3 = -\frac{n}{3}, \quad A_4 = -\frac{n}{4} + \frac{n^2}{8}, \\ A_5 = -\frac{n}{5} + \frac{n^2}{6}, \quad A_6 = -\frac{n}{6} + \frac{13n^2}{72} - \frac{n^3}{48}, \quad A_7 = -\frac{n}{7} + \frac{11n^2}{60} - \frac{n^3}{24}. \end{gathered} \tag{19}$$

In particular, by (17) we obtain the following asymptotic expansion:

$$\chi^2(P,Q) = \frac{p^2}{2} + \frac{2p^3}{3n} + \sum_{m\ge 4} \frac{m!}{n^m} A_m^2 p^m. \tag{20}$$

¿From here the lower bound in (16) follows immediately. To obtain an upper bound for the remainder sum on the right–hand side of (20) we need the following assertion.

Lemma 4.3. *For every natural n and $k > 1$, the following inequality is valid:*

$$B_k := \frac{k!}{n^k} A_k^2 < 3k.$$

Proof. By Lemma 4.1, taking the relations $B_0 = 1$ and $B_1 = 0$ into account, we have

$$\sum_{i\ge 2} B_i p^i = \chi^2(P,Q).$$

Since $B_i \ge 0$ for each natural i then

$$B_k \le \chi^2(P,Q)p^{-k} \le p^{-k}\big(\sup_{0\le i\le n} (P(i)/Q(i)) - 1\big).$$

In [8] the following inequality is proved:

$$\sup_{0\le i\le n}(P(i)/Q(i)) \le \frac{1}{1-p}.$$

Therefore,

$$B_k \le \frac{p}{p^k(1-p)}. \tag{21}$$

Since B_k does not depend on p then, setting $p = 1 - 1/k$ in (21) (this is the point of the minimum of the right–hand side of (21)), we finally obtain

$$B_k \le \frac{k-1}{\left(1-\frac{1}{k}\right)^k} \le \frac{k-1}{1-\frac{k}{k}+\frac{k(k-1)}{2k^2}-\frac{k(k-1)(k-2)}{6k^3}} = \frac{3k^2}{k+1} < 3k.$$

The Lemma is proven.

We now easily derive the upper bound in (16). Taking the form of the coefficients A_i in (19) into account it is easy to verify that $\frac{m!}{n^m}A_m^2 \le 1$ for $m = 4, 5, 6, 7$. Therefore, by Lemma 4.3 and (20), we derive the following inequality:

$$\chi^2(P,Q) \le \frac{p^2}{2} + \frac{2p^3}{3n} + p^4 + p^5 + p^6 + p^7 + \sum_{k\ge 8} 3kp^k. \tag{22}$$

Further, it is easy to verify the identity

$$\sum_{k\ge 8} 3kp^k = p^8 \sum_{k\ge 0} p^k(3(k+1)+21) = p^8\left(\frac{3}{(1-p)^2} + \frac{21}{1-p}\right)$$

$$= \frac{p^8(23-20p)}{(1-p)^2} + \frac{p^8}{1-p} = \frac{p^8(23-20p)}{(1-p)^2} + \frac{p^4}{1-p} - p^4 - p^5 - p^6 - p^7.$$

Using this representation and (22) we obtain the upper bound in (16). The Theorem is proven.

Corollary For the total variation distance between the distributions P and Q, the following upper bound is valid:

$$\|P-Q\| \le p\left(\frac{1}{2} + \frac{2p}{3n} + \frac{p^2}{1-p} + \frac{p^6(23-20p)}{(1-p)^2}\right)^{1/2}. \tag{23}$$

5. Comparison of upper bounds for the total variation distance

The best known upper bound for the total variation distance between a binomial distribution and the accompanying Poisson law has the form (see [5, 8, 82])

$$\|P-Q\| \le \min(2p, 2np^2), \tag{24}$$

meanwhile, there are no restrictions on the parameters n and p. As follows from Theorem 4.1 (see the lower bound), χ^2-distance is too strong to obtain the estimates of order $O(np^2)$ for the total variation distance in whole range of the variables n and p. Therefore, in the sequel, we will consider only the case $\lambda = np \geq 1$. In this case, the right–hand side of (24) has the form $\min(2p, 2np^2) = 2p$.

As a consequence of evaluating the Kulback–Leibler distance in [57], the authors derived by Pinsker's inequality the upper bound

$$\|P - Q\| \leq p\left(-\frac{\log(1-p)+p}{p^2} + \frac{1+p}{2n(1-p)^3}\right)^{1/2}. \tag{25}$$

Denote by R_1^0 and R_2^0 the respective right–hand sides of inequalities (23) and (25).

Notice that, as $p \to 0$, uniformly over all $\lambda \geq 1$, the following asymptotic representations are valid:

$$R_1^0 = p\left(\frac{1}{2} + p^2\left(\frac{2}{3\lambda} + 1\right) + O(p^3)\right)^{1/2}, \quad R_2^0 = p\left(\frac{1}{2} + p\left(\frac{1}{2\lambda} + \frac{1}{3}\right) + O(p^2)\right)^{1/2}.$$

So, for p small enough, the upper bound in (23) is slightly better than (25) in spite of equivalence of the leading terms of the asymptotics of both the variables R_1^0 and R_2^0 as $p \to 0$: These parts are equal to $p/\sqrt{2}$. In this connection, we note that the constant $1/\sqrt{2} = 0,7071\cdots$ slightly exceeds the exact lower bound $3/2e = 0,5518\ldots$ from [9] for the constants c in upper bounds of the form cp for the total variation distance under consideration in the range $\lambda \geq 1$ and $p \leq 1/4$.

Notice that, under some restrictions on the parameters p and n, one can obtain stronger upper bounds than that in (24). For example, in [2] for $p \leq 1/4$ and $\lambda \geq 3$, the following inequality was proved:

$$\|P - Q\| \leq 1,64p.$$

For comparison, in a more broad range of the variables p and λ, namely, under the restrictions $p \leq 1/4$ and $\lambda \geq 1$, it is easy to show (in this case, the right–hand sides in (23) and (25) reach their maxima in the points $p = 1/4$ and $\lambda = 1$), so that $R_2^0 < 0,987p$ but the right–hand side of (23) admits the better estimate $R_1^0 < 0,796p$. Notice also that, for the estimates above, a restriction from below on λ is equivalent to a restriction from below on n. More precisely, if we consider restrictions of the type $p \leq p^*$ and, $\lambda \geq \lambda^*$ then the maxima of the factors of p in (23) and (25) coincide with those under the restrictions $p \leq p^*$ and $n \geq \lambda^*/p^*$, and these maxima are reached in the corresponding extreme points. As an example, we cite the uniform upper bounds for R_1^0 and R_2^0 in some typical ranges of the variables n and p:

1) In the range $p \leq 1/4$ and $n \geq 10$ we have $R_1^0 < 0,780p$ and $R_2^0 < 0,867p$;

2) In the range $p \leq 1/4, n \geq 100$ we have $R_1^0 < 0,770p$ $R_2^0 < 0,786p$.

Beginning with the next order $n \geq 1000$ the uniform over $p \leq 1/4$ upper bounds for the values under consideration do not change (up to the accuracy indicated above). For all such n, We may put that $R_1^0 < 0,769p$ and $R_2^0 < 0,778p$.

Remark In [105], for the total variation distance, the following upper bound is valid:

$$\|P - Q\| \le p\left(\frac{3}{2e} + \frac{14\sqrt{p}(3 - 2\sqrt{p})}{6(1 - \sqrt{p})^2}\right). \tag{26}$$

It is clear that the value $C(p) := 3/2e + O(\sqrt{p})$ in the brackets on the right–hand side of (26), for p small enough, is slightly less than the value

$$C_0(p, n) := \left(\frac{1}{2} + \frac{2p}{3n} + O(p^2)\right)^{1/2}$$

from (23). Since the function $C(p)$ increases on the interval $(0, 1/4)$, to obtain the corresponding upper bounds it suffices to calculate the values of this function in the corresponding extreme points of the intervals under consideration. For example, $C(1/4) = 9,885\ldots$, $C(0,1) = 4,288\ldots$, $C(0,01) = 1,358\ldots$, $C(0,001) = 0,782\ldots$, and finally, $C(0,0001) = 0,622\ldots$.

But as indicated above,

$$\max_{p\le 1/4,\, n\ge 10} C_0(p, n) = 0,780\ldots, \qquad \max_{p\le 1/4,\, n\ge 100} C_0(p, n) = 0,770\ldots.$$

Thus, the upper bounds in (26) will be better in comparison with (23) only for the values of p of the order 10^{-4}. But for such p, the problem of improvement of the multiplicative constant in upper bounds of the type cp for the total variation distance, with relation to some applications, seems empty.

6 Poissonian approximation of rescaled set-indexed empirical processes

Let $\{X_i;\ i \geq 1\}$ be independent identically distributed random variables in $\mathbf{R}^k$ possessing an absolutely continuous distribution with a density $f(t)$ continuous at zero. Introduce the rescaled empirical process

$$S_n(A) := \sum_{i \leq n} I\{n^{1/k} X_i \in A\}$$

indexed by the class $\mathcal{B}_0$ of all Borel subsets A of a bounded Borel set A_0 in $\mathbf{R}^k$. Denote by $\{\Pi(A);\ A \in \mathcal{B}_0\}$ the Poisson point process with mean measure $f(0)\lambda(\cdot)$, where $\lambda(\cdot)$ is the Lebesgue measure in $\mathbf{R}^k$.

Weak convergence of distributions of the point processes $\{S_n(A);\ A \in \mathcal{B}_0\}$ to that of $\{\Pi(A);\ A \in \mathcal{B}_0\}$ in a functional sample space has been studied by Ivanoff and Merzbach ([68]). Actually, they proved only convergence of finite-dimensional distributions of these point processes by using a theory of set-indexed martingales. The main goal of this chapter is to obtain a much stronger result by a simpler (but not traditional) technique of multivariate analysis. In the theorem below we prove that convergence of distributions of the processes actually holds in the total variation distance.

Theorem. *The point processes $S_n(\cdot)$ and $\Pi(\cdot)$ can be constructed on a common probability space such that*

$$\lim_{n\to\infty} \mathbf{P}\left(\sup_{A \in \mathcal{B}_0} |S_n(A) - \Pi(A)| > 0 \right) = 0. \tag{1}$$

Proof. It suffices to prove (1) only for the case in which $\mathcal{B}_0$ is a countable determining class for all discrete measures. More precisely, let $\mathcal{B}_0$ be the algebra generated by all balls with bounded rational diameters and rational centers in a bounded subset of $\mathbf{R}^k$. Consider the functional space

$$\mathbf{B} := \{f(A);\ A \in \mathcal{B}_0 :\ \sup_{A \in \mathcal{B}_0} |f(A)| < \infty\}$$

endowed with the norm $\|f\| := \sup_i \{\alpha_i |f(A_i)|\}$, where $\{A_i\} = \mathcal{B}_0$ and $\{\alpha_i\}$ is a sequence of positive numbers tending monotonically to zero. Note that the linear normed space $\{\mathbf{B},\ \|\cdot\|\}$ is separable (but not complete).

Recall that the generalized Poisson distribution $Pois(\mu)$ with the Lévy measure μ in a separable linear normed space $\mathbf{B}$ is defined by the relation

$$Pois(\mu) := e^{-\mu(\mathbf{B})} \sum_{k=0}^{\infty} \frac{\mu^{*k}}{k!}, \tag{2}$$

where μ^{*k} is the k-fold convolution of a finite measure μ with itself; μ^{*0} is the unit mass concentrated at zero. We say that $Pois(\mu)$ is the accompanying Poisson distribution for the nth convolution of $\mathcal{L}(X)$ (or for the nth sum of independent copies of a random variable X), if $\mu = n\mathcal{L}(X)$.

Put $\mu := n\mathcal{L}\{I(n^{1/k}X_1 \in A);\, A \in \mathcal{B}_0\}$ and consider the distribution $Pois(\mu)$. It is the accompanying Poisson distribution for $S_n(\cdot)$ in the space $\{\mathbf{B},\, \|\cdot\|\}$. The well-known Le Cam result (for the details, see Borisov ([18]) and Le Cam ([84])) yields the following upper bound.

Lemma 1. *For all n, the following inequality holds:*

$$V(\mathcal{L}(S_n),\, Pois(\mu)) \le np^2, \tag{3}$$

where $V(\cdot,\cdot)$ is the total variation distance between two finite measures in $\{\mathbf{B},\, \|\cdot\|\}$,

$$p := 1 - \mathbf{P}(I(n^{1/k}X_1 \in A) = 0,\, \forall A \in \mathcal{B}_0) = \mathbf{P}(n^{1/k}X_1 \in A_0).$$

Let Y be a random variable with the uniform distribution in the ball S_f in $\mathbf{R}^k$ with center zero and size $f(0)^{-1}$. Consider in $\{\mathbf{B},\, \|\cdot\|\}$ the generalized Poisson distribution $Pois(\mu_0)$ with the Lévy measure $\mu_0 := n\mathcal{L}\{I(n^{1/k}Y \in A);\, A \in \mathcal{B}_0\}$. It is easy to verify that, for sufficiently large n satisfying the condition $n^{-1/k}A_0 \subseteq S_f$, the equality $Pois(\mu_0) = \mathcal{L}(\Pi)$ holds, where the process $\Pi(\cdot)$ was introduced in the above theorem, because, in this case, $n\mathbf{E}I(n^{1/k}Y \in A) = f(0)\lambda(A)$ for all $A \in \mathcal{B}_0$.

Lemma 2. *For all n, the following inequality holds:*

$$V(Pois(\mu),\, Pois(\mu_0)) \le 2n \sup_{A \in \mathcal{B}_0} |\mathbf{P}(n^{1/k}X_1 \in A) - \mathbf{P}(n^{1/k}Y \in A)|. \tag{4}$$

Proof. First, we obtain from (2) the following simple upper bound:

$$V(Pois(\mu),\, Pois(\mu_0)) \le 2V(\mu,\, \mu_0). \tag{5}$$

Indeed, because of the symmetry, we can put $\mu(\mathbf{B}) \le \mu_0(\mathbf{B})$. Then

$$V(Pois(\mu),\, Pois(\mu_0)) \le (e^{-\mu(\mathbf{B})} - e^{-\mu_0(\mathbf{B})})e^{\mu(\mathbf{B})} + \\ e^{-\mu_0(\mathbf{B})} V\left(\sum_{k=0}^{\infty} \frac{\mu^{*k}}{k!},\, \sum_{k=0}^{\infty} \frac{\mu_0^{*k}}{k!}\right).$$

The upper bound follows from the above inequality and the two estimates:

$$e^{-\mu(\mathbf{B})} - e^{-\mu_0(\mathbf{B})} \le (\mu_0(\mathbf{B}) - \mu(\mathbf{B}))e^{-\mu(\mathbf{B})}, \\ V(\mu^{*k},\, \mu_0^{*k}) \le k\mu_0(\mathbf{B})^{k-1}V(\mu,\, \mu_0).$$

Because of the separability and the definition of the norm, the total variation distance on the right-hand side of (5) coincides with the supremum over all finite-dimensional cylindrical sets in $\{\mathbf{B},\, \|\cdot\|\}$. But, in this case, the right-hand side of (5) is estimated easily by the right-hand side of (4). The lemma is proved.

Finally, under the condition $n^{-1/k}A_0 \subseteq S_f$, we have

$$\sup_{A \in \mathcal{B}_0} |\mathbf{P}(n^{1/k}X_1 \in A) - \mathbf{P}(n^{1/k}Y \in A)| \leq \frac{1}{n} \int\limits_{A_0} |f(n^{-1/k}z) - f(0)|dz.$$

Moreover, we need the obvious representation

$$p = \frac{1}{n} \int\limits_{A_0} f(n^{-1/k}z)dz \sim \frac{1}{n} f(0)\lambda(A_0),$$

where the above integrals are understood to be k-fold. Relation (1) follows from the triangle inequality for $V(\cdot, \cdot)$, the above upper bounds and the classical Lebesgue theorem, and from a theorem of Dobrushin ([45]) which asserts: it is possible to construct two random variables in a separable metric space such that the probability of their noncoincidence equals the total variation distance between their distributions.

R e m a r k. If the density $f(t)$ satisfies the Lipschitz condition at zero, then $V(S_n, \Pi) = O(n^{-1/k})$. By the Dobrushin theorem, the analogous rate of convergence is valid for the probability in (1). Moreover, we can define $f(t)$ as the derivative of the distribution function with respect to the direction determined by the family of subsets

$$\{n^{-1/k}A_0;\, n \geq 1\}.$$

7 Poissonian approximation for expectations of functions of independent random variables

1. Statement of the main results.

Accuracy of Poisson approximation for sums of independent random variables (r.v.'s) has already been investigated for about half a century. One of the first results was obtained by Prohorov (1953), who estimated the total variation distance between a binomial distribution and the corresponding Poisson law. His upper bound is close to being unimprovable. Le Cam (1960) considerably generalized and strengthened the Prohorov estimate. He used the so-called operator technique to extend the estimate to the case of approximating distributions of sums of independent arbitrarily distributed r.v.'s by corresponding generalized (accompanying) Poisson laws. Take note also of the remarkable result due to Barbour and Hall (1984) which, in particular, proved unimprovability of the Prohorov–Le Cam estimate for the total variation distance between a binomial distribution and the corresponding Poisson law.

Approximation of the next orders in the Poisson limit theorem, called an asymptotic expansion, originates from Chen (1975). Note also the paper by Deheuvels and Pfeifer (1986), where the first term of the expansion was explicitly written out. Kruopis (1986a, b) simultaneously obtained a similar result. However, he represented the first term of the expansion implicitly as the Fourier transform of some signed measure. Borovkov (1988) proposed a new approach to derive terms of the expansion in a form similar to that of Kruopis (1986a, b). He used a combination of the operator technique and a coupling to derive complete asymptotic expansion in the Poisson theorem. As in Kruopis (1986a, b) the expansion was presented via the Fourier transforms and, moreover, a scheme to invert the transforms was discussed. However, in our opinion, the resulting explicit representation of the terms of the asymptotic expansion is resistant to analysis [in contrast, say, to Deheuvels and Pfeifer (1986)].

It is worth noting that in all of the above-mentioned papers remainders of the expansions are estimated in terms of the total variation distance. Thus, it is clear that these results can be reformulated in terms of moments of bounded functions with the corresponding upper bounds of remainders taken uniformly over all bounded (say, by 1) functions. There are also more publications where Poisson approximation of the moments is considered uniformly over special subclasses of bounded functions [e.g., see Roos (1995, 1998)].

However, the total variation distance, the point metric and some related distances become unsuitable for approximation of moments of unbounded functions of a binomial r.v. In this connection, investigations of the Poisson approximation for expectations of unbounded functions of sums of r.v.'s are to be distinguished. The papers which are to be particularly noted are those by Barbour (1987) and Barbour, Chen and Choi (1995). Barbour (1987) used the so-called Stein–Chen method to obtain complete asymptotic expansions in the Poisson approximation of at most polynomially growing functions of sums of independent arbitrarily distributed integer nonnegative r.v.'s. In the latter paper, a similar approach is used to obtain the first term of the expansion for expectation of some functions of a binomial r.v. under minimal moment restrictions on the functions under consideration. In particular, these restrictions allow the functions growing faster than exponential. The bounds on the remainders of the expansions in these two papers are not far from optimal.

The main goal of the chapter is to obtain complete asymptotic expansions of moments of unbounded functions of n independent r.v.'s in the case when each of these r.v.'s is equal to zero with high probability. The probabilities for these r.v.'s to be unequal to zero are considered as natural small parameters, sums of which powers are used to represent the asymptotic expansions. In the case of asymptotic expansions for bounded functions a similar formulation of the problem was given by Borovkov (1988). The particular case, wherein the functions depend only on the sum of the r.v.'s, is separately studied. In this case the estimate of the remainder of the expansion is unimprovable in some sense and improves the corresponding results of Barbour (1987) and Barbour, Chen and Choi (1995) in a broad range of change of the expansion parameters.

The approach used to derive the main results is based on the so-called Lindeberg method. This method was successfully employed in a great number of papers to study rates of convergence in the central limit theorem (the Gaussian case) under various settings including studying remainders in different asymptotic expansions. However, the method was rarely used in the Poisson approximation. One can find certain versions of the Lindeberg method in the Poisson approximation, for example, in the papers by Le Cam (1960), Deheuvels, Karr, Pfeifer and Serfling (1988) and Novak (1998).

In Section 5 it is shown that the problem of the Poisson approximation for expectation of a function of independent arbitrarily distributed r.v.'s can be reduced to the case of independent Bernoulli r.v.'s. Thus, in the chapter, this case is particularly studied.

Let ζ_1,...,ζ_n be independent Bernoulli r.v.'s with the success probabilities $p_j = \mathbf{P}(\zeta_j = 1)$, $j = 1, ..., n$. Let η_1,...,η_n be independent Poisson r.v.'s with parameters $p_1, ..., p_n$, respectively. Denote $\lambda_k = p_1 + \cdots + p_k$, $\lambda = \lambda_n$, and $\bar{\zeta} = (\zeta_1, ..., \zeta_n)$, $\bar{\eta} = (\eta_1, ..., \eta_n)$. Finally, let F be an arbitrary real function of n nonnegative integer variables.

Denote by $\bar{e_k}$ the n-dimensional vector which has the kth coordinate 1 and all the other coordinates 0. For any function G of n arguments, define the difference operator Δ_k:

$$\Delta_k G(\bar{a}) = G(\bar{a} + \bar{e_k}) - G(\bar{a}).$$

In the sequel we denote by Δ_k^r the corresponding operator power. In the one-dimensional case the subscript will be omitted.

The following theorem is the key result for deriving most of the subsequent statements.

THEOREM 1. *Suppose* $\mathbf{E}|F(\bar{\eta})| < \infty$. *Then*

$$\mathbf{E}F(\bar{\eta}) - \mathbf{E}F(\bar{\zeta}) = \sum_{k=1}^{n} \sum_{r=2}^{\infty} \frac{p_k^r}{r!} \mathbf{E}\Delta_k^r F(\bar{\phi}_k), \tag{1}$$

where $\bar\phi_k = (\zeta_1, ..., \zeta_{k-1}, 0, \eta_{k+1}, ..., \eta_n)$, *and, for each* k, *the corresponding series in* (1) *absolutely converges.*

Moreover, if $\mathbf{E}{\eta_k}^{s+1}|F(\bar\eta)| < \infty$ *for all* k *and some* $s \geq 1$, *then, first, the remainder of series in* (1) *can be estimated as follows*:

$$\left|\sum_{r=s+1}^{\infty} \frac{p_k^r}{r!}\mathbf{E}\Delta^r F(\bar\phi_k)\right| \leq e^{p_k}\frac{p_k^{s+1}}{(s+1)!}\mathbf{E}|\Delta_k^{s+1}F(\bar\psi_k)|, \tag{2}$$

where $\bar\psi_k = \bar\phi_k + \eta_k\bar e_k = (\zeta_1, ..., \zeta_{k-1}, \eta_k, ..., \eta_n)$ *and, second, another expansion of the difference* $\mathbf{E}F(\bar\eta) - \mathbf{E}F(\bar\zeta)$ *holds*:

$$\begin{aligned}&\left|\mathbf{E}F(\bar\eta) - \mathbf{E}F(\bar\zeta) - \sum_{r=2}^{s}(-1)^r(r-1)\sum_{k=1}^{n}\frac{p_k^r}{r!}\mathbf{E}\Delta_k^r F(\bar\psi_k)\right| \\ &\qquad \leq \frac{s}{(s+1)!}\sum_{k=1}^{n} e^{p_k}p_k^{s+1}\mathbf{E}|\Delta_k^{s+1}F(\bar\psi_k)|.\end{aligned} \tag{3}$$

Remark 1. Under the moment restrictions considered above expansion (3) cannot be represented as a converging series with an upper bound for its remainder like in (1) and (2), because this representation would require considerably stronger moment restrictions.

Remark 2. The right-hand sides of the inequalities (2) and (3) can be bounded through expectations of functions of $\bar\eta$ using the obvious unimprovable upper bound for the Radon–Nikodym derivative of the distribution of $\bar\psi$ with respect to the distribution of $\bar\eta$:

$$\mathbf{E}|\Delta_k^{s+1}F(\bar\psi_k)| \leq e^{\lambda_k}\mathbf{E}|\Delta_k^{s+1}F(\bar\eta)|.$$

The right-hand side of the inequality is finite if $\mathbf{E}\eta_k^{s+1}|F(\bar\eta)| < \infty$ for all k, since the following proposition is true:

Proposition 1. *Let* τ *be an arbitrary Poisson r.v. and* g *be an arbitrary real function. Then, for each* $l = 1, 2, ...$, *the following three conditions are equivalent*:

(a) $\mathbf{E}\tau^l|g(\tau)| < \infty$.
(b) $\mathbf{E}|g(\tau + l)| < \infty$.
(c) $\mathbf{E}|\Delta^l g(\tau)| < \infty$.

Corollary 1. *If, for all* k, $\mathbf{E}{\eta_k}^2|F(\bar\eta)| < \infty$, *then*

$$\begin{aligned}|\mathbf{E}F(\bar\zeta) - \mathbf{E}F(\bar\eta)| &\leq \tfrac{1}{2}\sum_{k=1}^{n} e^{p_k}p_k^2\mathbf{E}|\Delta_k^2 F(\bar\psi_k)| \\ &\leq \tfrac{1}{2}\sum_{k=1}^{n} e^{\lambda_k}p_k^2\mathbf{E}|\Delta_k^2 F(\bar\eta)|.\end{aligned}$$

COROLLARY 2. *If* $\mathbf{E}{\eta_k}^3|F(\bar\eta)|<\infty$, $\mathbf{E}{\eta_j}^2{\eta_k}^2|F(\bar\eta)|<\infty$ *for all* $j,\ k : j\neq k$, *then*

$$\left|\mathbf{E}F(\bar\zeta)-\mathbf{E}F(\bar\eta)+\tfrac{1}{2}\sum_{j=1}^{n}p_j^2\mathbf{E}\Delta_j^2F(\bar\eta)\right|$$
$$\le \tfrac{1}{4}\sum_{k=1}^{n}p_k^2\sum_{j=1}^{k-1}e^{p_j}p_j^2\mathbf{E}|\Delta_j^2\Delta_k^2F(\bar\psi_j)|+\tfrac{1}{3}\sum_{j=1}^{n}e^{p_j}p_j^3\mathbf{E}|\Delta_j^3F(\bar\psi_j)|.$$

Theorem 1 also allows us to obtain complete asymptotic expansions of $\mathbf{E}F(\bar\zeta)-\mathbf{E}F(\bar\eta)$, since the expectations $\mathbf{E}\Delta_k^rF(\bar\psi_k)$ in (3) can be subsequently approximated with expectations $\mathbf{E}\Delta_k^rF(\bar\eta_k)$ using (3).

COROLLARY 3. *Let* $l\ge 1$. *Suppose*

$$\mathbf{E}\eta_{k_1}^{r_1}\cdots\eta_{k_{s-1}}^{r_{s-1}}\eta_{k_s}^{l+s-r_1-\cdots-r_{s-1}}|F(\bar\eta)|<\infty$$

for all $1\le s\le l$, $n\ge k_1>k_2>\cdots>k_s\ge 1$, $2\le r_1\le l$,...,$2\le r_{s-1}\le l+s-2$. *Then*

$$\Bigg|\mathbf{E}F(\bar\zeta)-\mathbf{E}F(\bar\eta)-$$
$$\sum{}'(-1)^{s+r_1+\cdots+r_s}(r_1-1)\cdots(r_s-1)\frac{p_{k_1}^{r_1}\cdots p_{k_s}^{r_s}}{r_1!\cdots r_s!}\mathbf{E}\Delta_{k_1}^{r_1}\cdots\Delta_{k_s}^{r_s}F(\bar\eta)\Bigg|$$
$$\le\sum{}''e^{p_{k_s}}(r_1-1)\cdots(r_{s-1}-1)(l+s-r_1-...r_{s-1}-1)$$
$$\frac{p_{k_1}^{r_1}\cdots p_{k_{s-1}}^{r_{s-1}}p_{k_s}^{l+s-r_1-...-r_{s-1}}}{r_1!\cdots r_{s-1}!(l+s-r_1-...r_{s-1})!}\mathbf{E}|\Delta_{k_1}^{r_1}\cdots\Delta_{k_{s-1}}^{r_{s-1}}\Delta_{k_s}^{l+s-r_1-...r_{s-1}}F(\bar\psi_{k_s})|,$$

where $\sum'$ *and* $\sum''$ *denote the following sums*:

$$\sum{}'=\sum_{s=1}^{l-1}\sum_{k_1=1}^{n}\sum_{r_1=2}^{l}\sum_{k_2=1}^{k_1-1}\sum_{r_2=2}^{l+1-r_1}\cdots\sum_{k_s=1}^{k_{s-1}-1}\sum_{r_s=2}^{l+s-1-r_1-\cdots-r_{s-1}},$$

$$\sum{}''=\sum_{s=1}^{l}\sum_{k_1=1}^{n}\sum_{r_1=2}^{l}\sum_{k_2=1}^{k_1-1}\sum_{r_2=2}^{l+1-r_1}\cdots\sum_{k_{s-1}=1}^{k_{s-2}-1}\sum_{r_{s-1}=2}^{l+s-2-r_1-\cdots-r_{s-2}}\sum_{k_s=1}^{k_{s-1}-1}.$$

In the last sum we assume $k_0=n+1$ *if* $s=1$.

Now we consider the Poisson approximation for moments of functions of sums of independent Bernoulli r.v.'s. Put $S=\zeta_1+\cdots\zeta_n$, $Z=\eta_1+\cdots\eta_n$ and let h be an arbitrary function of nonnegative integer. Introduce the following notation: $\lambda=p_1+\cdots+p_n$, $\tilde p_k=\max\{p_1,...,p_k\}$, $\tilde p=\tilde p_n$.

Theorem 2. *Let $\mathbf{E}|h(Z)| < \infty$. Then*

$$\mathbf{E}h(Z) - \mathbf{E}h(S) = \sum_{k=1}^{n}\sum_{r=2}^{\infty} \frac{p_k^r}{r!}\mathbf{E}\Delta^r h(T_k), \tag{4}$$

where $T_k = \zeta_1 + \cdots + \zeta_{k-1} + \eta_{k+1} + \cdots + \eta_n$, and, for each k, corresponding series in (4) *absolutely converges.*

Moreover, if $\mathbf{E}Z^{s+1}|h(Z)| < \infty$ then, first,

$$\left|\sum_{r=s+1}^{\infty} \frac{p_k^r}{r!}\mathbf{E}\Delta^r h(T_k)\right| \le \frac{e^{p_k}}{(1-\tilde{p}_{k-1})^2}\frac{p_k^{s+1}}{(s+1)!}\mathbf{E}|\Delta^{s+1}h(Z)|, \qquad s \ge 1 \tag{5}$$

and, second, another expansion of the difference $\mathbf{E}h(Z) - \mathbf{E}h(S)$ holds:

$$\begin{aligned} &\left|\mathbf{E}h(Z) - \mathbf{E}h(S) - \sum_{r=2}^{s}(-1)^r(r-1)\sum_{k=1}^{n}\frac{p_k^r}{r!}\mathbf{E}\Delta^r h(Y_k)\right| \\ &\qquad \le \frac{s}{(s+1)!}\sum_{k=1}^{n}\frac{e^{p_k}}{(1-\tilde{p}_{k-1})^2}p_k^{s+1}\mathbf{E}|\Delta^{s+1}h(Z)|, \end{aligned} \tag{6}$$

where $Y_k = \zeta_1 + \cdots + \zeta_{k-1} + \eta_k + \cdots + \eta_n$.

Remark 3. The principal distinction between Theorem 1 and Theorem 2 is the appreciably sharper upper bound for the remainder in Theorem 2 which is obtained by the corresponding upper bound for the Radon–Nikodym derivative in Lemma 2 (see Section 2). Formal application of Theorem 1 to functions of sums of the arguments yields an upper bound for the remainder which is substantially rougher than that in Theorem 2 as $\lambda \to \infty$.

Remark 4. As noted in Proposition 1, the finiteness of $\mathbf{E}|\Delta^{s+1}h(Z)|$ is equivalent to finiteness of $\mathbf{E}|Z^{s+1}h(Z)|$. Nevertheless, the series (4) absolutely converges under weaker ($s = -1$) moment restrictions.

Corollary 4. *If $\mathbf{E}Z^2|h(Z)| < \infty$, then*

$$|\mathbf{E}h(S) - \mathbf{E}h(Z)| \le \frac{1}{2}\frac{e^{\tilde{p}}}{(1-\tilde{p})^2}\sum_{j=1}^{n} p_j^2\mathbf{E}|\Delta^2 h(Z)|.$$

Corollary 5. *If $\mathbf{E}Z^4|h(Z)| < \infty$, then*

$$\begin{aligned} &\left|\mathbf{E}h(S) - \mathbf{E}h(Z) + \frac{1}{2}\sum_{j=1}^{n}p_j^2\mathbf{E}\Delta^2 h(Z)\right| \\ &\qquad \le \frac{e^{\tilde{p}}}{(1-\tilde{p})^2}\left\{\frac{1}{3}\sum_{j=1}^{n}p_j^3\mathbf{E}|\Delta^3 h(Z)| + \frac{1}{8}\left(\sum_{j=1}^{n}p_j^2\right)^2\mathbf{E}|\Delta^4 h(Z)|\right\}. \end{aligned}$$

COROLLARY 6. *If* $\mathbf{E}Z^6|h(Z)| < \infty$*, then*

$$\left|\mathbf{E}h(S) - \mathbf{E}h(Z) + \frac{1}{2}\sum_{j=1}^{n} p_j^2 \mathbf{E}\Delta^2 h(Z) - \frac{1}{3}\sum_{j=1}^{n} p_j^3 \mathbf{E}\Delta^3 h(Z) - \frac{1}{8}\left(\sum_{j=1}^{n} p_j^2\right)^2 \mathbf{E}\Delta^4 h(Z)\right|$$

$$\leq \frac{e^{\tilde{p}}}{(1-\tilde{p})^2}\left\{\frac{1}{4}\sum_{j=1}^{n} p_j^4 \mathbf{E}|\Delta^4 h(Z)| + \frac{1}{6}\sum_{j=1}^{n} p_j^2 \sum_{k=1}^{n} p_k^3 \mathbf{E}|\Delta^5 h(Z)| + \frac{1}{48}\left(\sum_{j=1}^{n} p_j^2\right)^3 \mathbf{E}|\Delta^6 h(Z)|\right\}.$$

The complete asymptotic expansion for the sums can be written as follows:

COROLLARY 7. *If* $\mathbf{E}|\Delta^{2l} h(Z)| < \infty$*, then*

$$\Bigg|\mathbf{E}h(S) - \mathbf{E}h(Z) -$$

$$\sum{}' (-1)^{s+r_1+\dots+r_s}(r_1-1)\cdots(r_s-1)\frac{p_{k_1}^{r_1}\cdots p_{k_s}^{r_s}}{r_1!\cdots r_s!}\mathbf{E}\Delta^{r_1+\dots+r_s} h(Z)\Bigg|$$

$$\leq \sum{}'' \frac{e^{p_{k_s}}}{(1-\tilde{p}_{k_s})^2}(r_1-1)\cdots(r_{s-1}-1)(l+s-r_1-\dots r_{s-1}-1)$$

$$\frac{p_{k_1}^{r_1}\cdots p_{k_{s-1}}^{r_{s-1}} p_{k_s}^{l+s-r_1-\dots-r_{s-1}}}{r_1!\cdots r_{s-1}!(l+s-r_1-\dots r_{s-1})!}\mathbf{E}|\Delta^{l+s} h(Z)|,$$

where $\sum'$ *and* $\sum''$ *denote the following sums*:

$$\sum{}' = \sum_{s=1}^{l-1}\sum_{k_1=1}^{n}\sum_{r_1=2}^{l}\sum_{k_2=1}^{k_1-1}\sum_{r_2=2}^{l+1-r_1}\cdots\sum_{k_s=1}^{k_{s-1}-1}\sum_{r_s=2}^{l+s-1-r_1-\dots-r_{s-1}},$$

$$\sum{}'' = \sum_{s=1}^{l}\sum_{k_1=1}^{n}\sum_{r_1=2}^{l}\sum_{k_2=1}^{k_1-1}\sum_{r_2=2}^{l+1-r_1}\cdots\sum_{k_{s-1}=1}^{k_{s-2}-1}\sum_{r_{s-1}=2}^{l+s-2-r_1-\dots-r_{s-2}}\sum_{k_s=1}^{k_{s-1}-1}.$$

In the last sum we suppose $k_0 = n+1$ *if* $s = 1$.

2. Preliminary results.

In this section we prove the following three lemmas, which are also of independent interest.

LEMMA 1. *Let* $p_1 = \cdots = p_n = p$. *Then*

$$\sup_j \frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)} \le \frac{1}{1-p} \tag{7}$$

and moreover,

$$\frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)} \le 1$$

if $j \le \max\{0, \lambda - \sqrt{2n(-\log(1-p))} + 1\}$ *or* $j \ge \lambda + \min\{\lambda, \sqrt{2\lambda}\} + 1$.

PROOF. For $j \le n$,

$$\begin{aligned}\frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)} &= \frac{n(n-1)\cdots(n-j+1)}{n^j(1-p)^j}(1-p)^n e^{np} = \\ &\exp\left\{n(p+\log(1-p)) - j\log(1-p) + \sum_{i=0}^{j-1}\log\left(1-\frac{i}{n}\right)\right\} \le\end{aligned}$$

$$\exp\Bigg\{ -\log(1-p) + n(p+\log(1-p)) - (j-1)\log(1-p)$$

$$+n\int_0^{(j-1)/n}\log(1-x)dx\Bigg\} \le \exp\left\{-\log(1-p) - nH_p\left(\frac{j-1}{n}\right)\right\}.$$

where $H_p(x) = -p + x + (1-x)\log((1-x)/(1-p))$. The following properties of H_p are obvious: $H_p(x) \ge 0$ if $x \le 1$ [hence (7) is true], $H_p(1) = 1-p$, $H_p(p) = 0$, $\frac{d}{dx}H_p(p) = 0$, $\frac{d^2}{dx^2}H_p(x) = 1/(1-x)$, that implies

$$H_p(x) \ge \frac{(x-p)^2}{2(1-p)} \qquad \text{if } p \le x \le 1.$$

Hence, for $j \ge \lambda + 1$,

$$\frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)} \le \exp\left\{-\log(1-p) - \frac{(j-1-np)^2}{2n(1-p)}\right\} \le 1 \qquad \text{if } j \ge \lambda + \sqrt{2\lambda} + 1.$$

Analogously,

$$H_p(x) \ge \frac{(x-p)^2}{2} \qquad \text{if } x \le p.$$

And hence, for $j \le \lambda + 1$,

$$\frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)} \le \exp\left\{-\log(1-p) - \frac{(j-1-np)^2}{2n}\right\} \le 1 \qquad \text{if } j \le \lambda - \sqrt{2n(-\log(1-p))} + 1.$$

We also have

$$\frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)} = \exp\left\{n(p+\log(1-p)) - j\log(1-p) + \sum_{i=0}^{j-1}\log\left(1-\frac{i}{n}\right)\right\} \le$$

$$\exp\{n(p+\log(1-p))-j\log(1-p)-1/n-2/n-\cdots-(j-1)/n\}\le$$
$$\exp\{jn^{-1}(np-(j-1)/2)\}\le 1 \qquad \text{if } j\ge 2\lambda+1.$$

The lemma is proved. □

LEMMA 2. *In the case of arbitrary p_j, the following inequality holds:*

$$\sup_j \frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)}\le \frac{1}{(1-\tilde p)^2}.$$

PROOF. Denote by $S(a_1,...,a_m)$ the sum of m independent Bernoulli r.v.'s the jth of which is equal to 1 with probability a_j. Denote also by $Z(b)$ a Poisson r.v. with parameter b. Let g be an arbitrary real function. The proof of Corollary 2.1 in Hoeffding (1956) does not have to be changed when it is applied to the similar statement concerning

$$\sup\{\mathbf{E}g(S(a_1,...,a_n)):\ 0\le a_1\le\tilde p,...,0\le a_n\le\tilde p,\ a_1+\cdots+a_n=\lambda\}.$$

This sumpremum is attained with such $a_1,...,a_n$ that $a_1=a_2=\cdots=a_m=a$, $a_{m+1}=...=a_k=\tilde p$, $a_{k+1}=\cdots=a_n=0$ for some a,m,k, $0<a\le\tilde p$, $1\le m\le k\le n$.

Now, for each $j=0,1,...,n$, we set $g(y)=I(y=j)$ and find the corresponding values $a=a(j)$, $m=m(j)$, $k=k(j)$. Let $a_1(j)=\cdots=a_{m(j)}(j)=a(j)$, $a_{m(j)+1}(j)=\cdots=a_{k(j)}(j)=\tilde p$, $a_{k(j)+1}(j)=\cdots=a_n(j)=0$.

In the case $m(j)<k(j)$ we have

$$\frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)}\le\frac{\mathbf{P}(S(a_1(j),...,a_n(j))=j)}{\mathbf{P}(Z=j)}=$$
$$\frac{\mathbf{P}(S_1(a_1(j),...,a_{m(j)}(j))+S_2(a_{m(j)+1}(j),...,a_{k(j)}(j))=j)}{\mathbf{P}(Z_1(m(j)a(j))+Z_2((k(j)-m(j))\tilde p)=j)}\le$$
$$\sup_{i\ge0}\frac{\mathbf{P}(S(a_1(j),...,a_{m(j)}(j))=i)}{\mathbf{P}(Z(m(j)a(j))=i)}\sup_{i\ge0}\frac{\mathbf{P}(S(a_{m(j)+1}(j),...,a_{k(j)}(j))=i)}{\mathbf{P}(Z((k(j)-m(j))\tilde p)=i)}\le$$
$$\frac{1}{(1-\tilde p)^2},$$

where $S_1(\cdot)$, $S_2(\cdot)$ and $Z_1(\cdot)$, $Z_2(\cdot)$ denote pairs of independent r.v.'s with the corresponding distributions; the last inequality follows from Lemma 1.

In the case $k(j)=m(j)$ the inequality

$$\frac{\mathbf{P}(S=j)}{\mathbf{P}(Z=j)}\le\frac{1}{(1-\tilde p)^2}$$

is also true. The lemma is proved. □

LEMMA 3. *Let Z' be a Poisson r.v. with parameter $\lambda-\delta$, where $0<\delta<\lambda$. Then*

$$\left|\mathbf{E}g(Z')-e^{\delta}\sum_{j=0}^{m}\frac{(-\delta)^j}{j!}\mathbf{E}g(Z+j)\right|\le\frac{e^{\delta}\delta^{m+1}}{(m+1)!}\mathbf{E}|g(Z+m+1)|,$$

if $\mathbf{E}|g(Z+m+1)|<\infty$.

PROOF. We have

$$\mathbf{E}g(Z')=e^{\delta}\mathbf{E}g(Z)(1-\delta/\lambda)^{Z}=$$
$$e^{\delta}\mathbf{E}g(Z)(1-\delta/\lambda)^{Z}I(Z\le m)+e^{\delta}\mathbf{E}g(Z)(1-\delta/\lambda)^{Z}I(Z>m).$$

We assume binomial coefficient C_l^j to be zero if $j>l$. Consider the first term on the right-hand side of the last relation. We have

$$\begin{aligned}\mathbf{E}g(Z)(1-\delta/\lambda)^{Z}I(Z\le m)&=\mathbf{E}g(Z)\sum_{j=0}^{m}C_Z^j(-\delta/\lambda)^jI(Z\le m)\\ &=\sum_{j=0}^{m}(-\delta/\lambda)^j\mathbf{E}C_Z^jg(Z)I(Z\le m)\\ &=\sum_{j=0}^{m}\frac{(-\delta)^j}{j!}\mathbf{E}g(Z+j)I(Z\le m),\end{aligned}\tag{8}$$

where the last equality is true because of the identity

$$\mathbf{E}Z(Z-1)\cdots(Z-j)g(Z)=\lambda^j\mathbf{E}g(Z+j).$$

On the subset of elementary events $\{Z>m\}$, using Taylor's formula we have

$$(1-\delta/\lambda)^{Z}=\sum_{j=0}^{m}C_Z^j(-\delta/\lambda)^j+C_Z^{m+1}(\delta/\lambda)^{m+1}\theta^{m+1},$$

where θ is a function of δ, λ, Z and m, such that $|\theta|\le 1$. Hence

$$\begin{aligned}\mathbf{E}g(Z)(1-\delta/\lambda)^{Z}I(Z>m)=\sum_{j=0}^{m}\frac{(-\delta)^j}{j!}&\mathbf{E}g(Z+j)I(Z>m)\\ &+\mathbf{E}C_Z^{m+1}(\delta/\lambda)^{m+1}\theta^{m+1}g(Z)I(Z>m),\end{aligned}\tag{9}$$

where the last summand can be easily estimated:

$$\begin{aligned}|\mathbf{E}C_Z^{m+1}(\delta/\lambda)^{m+1}\theta^{m+1}g(Z)I(Z>m)|&\le\mathbf{E}C_Z^{m+1}(\delta/\lambda)^{m+1}|g(Z)I(Z>m)|\\ &=\frac{\delta^{m+1}}{(m+1)!}\mathbf{E}|g(Z+m+1)|.\end{aligned}\tag{10}$$

Combining relations (8)–(10) we complete the proof. □

3. Proofs of the main results.

Proof of Proposition 1. First show that (a) implies (b):

$$\mathbf{E}|g(\tau+l)| = \sum_{j=l}^{\infty} e^{-\mu}\frac{\mu^{k-l}}{(k-l)!}|g(k)| \le \mu^{-l}e^{-\mu}\sum_{k=l}^{\infty}\frac{\mu^k k^l}{k!}|g(k)| \le \mu^{-l}e^{-\mu}\mathbf{E}\tau^l|g(\tau)|.$$

Analogously, (b) implies (a). It is also clear that (b) implies (c), since $\mathbf{E}|g(\tau+l)| < \infty$ implies $\mathbf{E}|g(\tau+k)| < \infty$ for all $k \le l$, and hence

$$\mathbf{E}|\Delta^l g(\tau)| = \mathbf{E}\left|\sum_{k=0}^{l}(-1)^{l-k}C_l^k g(\tau+k)\right| \le \mathbf{E}\sum_{k=0}^{l} C_l^k|g(\tau+k)| < \infty.$$

Now we show that (c) implies (b) if $l = 1$. We have

$$\mathbf{E}|g(\tau+1)| = \mathbf{E}\left|g(0) + \sum_{j=0}^{\tau}\Delta g(j)\right| \le |g(0)| + \mathbf{E}\sum_{j=0}^{\tau}|\Delta g(j)| =$$

$$|g(0)| + \sum_{k=0}^{\infty} e^{-\mu}\frac{\mu^k}{k!}\sum_{j=0}^{k}|\Delta g(j)| = |g(0)| + e^{-\mu}\sum_{j=0}^{\infty}\sum_{k=j}^{\infty}\frac{\mu^k}{k!}|\Delta g(j)| \le$$

$$|g(0)| + e^{\mu}\mathbf{E}|\Delta g(\tau)|.$$

Finally, it is easy to prove equivalence of (b) and (c) using induction on l. □

Proof of Theorem 1. The proof is substantially based on the Lindeberg method which is contained in the following identity:

$$\mathbf{E}F(\bar\eta) - \mathbf{E}F(\bar\zeta) = \sum_{k=1}^{n}(\mathbf{E}F(\bar\phi_k + \eta_k\bar e_k) - \mathbf{E}F(\bar\phi_k + \zeta_k\bar e_k)). \tag{11}$$

We have

$$\mathbf{E}F(\bar\phi_k + \zeta_k\bar e_k) = \mathbf{E}F(\bar\phi_k) + p_k\mathbf{E}\Delta_k F(\bar\phi_k).$$

For any function g, the following equality is well known:

$$\Delta^r g(y) = \sum_{j=0}^{r}(-1)^{r-j}\frac{r!}{(r-j)!j!}g(y+j).$$

Thus

$$\mathbf{E}F(\bar\phi_k + \eta_k\bar e_k) = \sum_{j=0}^{\infty} e^{-p_k}\frac{p_k^j}{j!}\mathbf{E}F(\bar\phi_k + j\bar e_k) =$$

$$\sum_{j=0}^{\infty}\sum_{t=0}^{\infty}\frac{(-1)^t p_k^{j+t}}{t!j!}\mathbf{E}F(\bar\phi_k + j\bar e_k) = \sum_{r=0}^{\infty}\sum_{j=0}^{r} p_k^r\frac{(-1)^{r-j}}{j!(r-j)!}\mathbf{E}F(\bar\phi_k + j\bar e_k) =$$

$$\sum_{r=0}^{\infty} \frac{p_k^r}{r!} \mathbf{E}\Delta_k^r F(\bar{\phi}_k), \tag{12}$$

where the order of summing was changed by Fubini's theorem.

Therefore,

$$\mathbf{E}F(\bar{\phi}_k + \eta_k \bar{e_k}) - \mathbf{E}F(\bar{\phi}_k + \zeta_k \bar{e_k}) = \sum_{r=2}^{\infty} \frac{p_k^r}{r!} \mathbf{E}\Delta_k^r F(\bar{\phi}_k). \tag{13}$$

The last equality together with (11) proves relation (1).

Now we prove (2). Let $\mathbf{E}|\Delta_k^{s+1} F(\bar{\eta})| < \infty$ and hence $\mathbf{E}|\Delta_k^{s+1} F(\bar{\phi}_k)| < \infty$. Thus we have

$$\left| \sum_{r=s+1}^{\infty} \frac{p_k^r}{r!} \mathbf{E}\Delta_k^r F(\bar{\phi}_k + j\bar{e_k}) \right| = \left| p_k^{s+1} \sum_{r=0}^{\infty} \sum_{j=0}^{r} \frac{p_k^r}{(r+s+1)!} \frac{(-1)^{r-j} r!}{j!(r-j)!} \mathbf{E}\Delta_k^{s+1} F(\bar{\phi}_k + j\bar{e_k}) \right| =$$

$$\left| p_k^{s+1} \sum_{j=0}^{\infty} \sum_{t=0}^{\infty} p_k^{j+t} \frac{(-1)^t (j+t)!}{(j+t+s+1)! j! t!} \mathbf{E}\Delta_k^{s+1} F(\bar{\phi}_k + j\bar{e_k}) \right| \leq$$

$$\frac{p_k^{s+1}}{(s+1)!} \sum_{j=0}^{\infty} \frac{p_k^j}{j!} \left| \sum_{t=0}^{\infty} \frac{(s+1)!(j+t)!}{(j+t+s+1)!} \frac{(-p_k)^t}{t!} \right| |\mathbf{E}\Delta_k^{s+1} F(\bar{\phi}_k + j\bar{e_k})| \leq$$

$$\frac{p_k^{s+1}}{(s+1)!} \sum_{j=0}^{\infty} \frac{p_k^j}{j!} \mathbf{E}|\Delta_k^{s+1} F(\bar{\phi}_k + j\bar{e_k})| = e^{p_k} \frac{p_k^{s+1}}{(s+1)!} \mathbf{E}|\Delta_k^{s+1} F(\bar{\phi}_k + \eta_k \bar{e_k})|.$$

So, (2) is true.

Now we proceed to proving (3). If $s = 1$ then, by (1) and (2), the relation is true. Consider the case $s \geq 2$. Set

$$f(j) = f_k(j) = \mathbf{E}F(\bar{\phi}_k + j\bar{e_k}).$$

To prove (3) it suffices to show that, for each k,

$$\left| \mathbf{E}f(\eta_k) - \mathbf{E}f(\zeta_k) - \sum_{r=2}^{s} (-1)^r (r-1) \frac{p_k^r}{r!} \mathbf{E}\Delta^r f(\eta_k) \right| \leq \frac{s}{(s+1)!} e^{p_k} p_k^{s+1} \mathbf{E}|\Delta^{s+1} f(\eta_k)|. \tag{14}$$

In order to prove the last relation we need the expression

$$1 - \sum_{j=2}^{m} (-1)^j (j-1) C_r^j, \qquad m \geq 2,$$

to be calculated. In order to do it we use the identity

$$\sum_{j=0}^{t} (-1)^j C_i^j = (-1)^t C_{i-1}^t$$

and derive that

$$\begin{aligned}\sum_{j=2}^{m}(-1)^j C_r^j &= (-1)^m C_{r-1}^m - 1 + r,\\ -\sum_{j=2}^{m}(-1)^j j C_r^j &= r\sum_{j=1}^{m-1}(-1)^j C_{r-1}^j = -(-1)^m m C_{r-2}^{m-1} - r.\end{aligned}$$

Thus, for $m \geq 2$,

$$1 - \sum_{j=2}^{m}(-1)^j (j-1) C_r^j = (-1)^m \left(C_{r-1}^m - m C_{r-2}^{m-1}\right)$$

$$= -(-1)^m \left(\frac{(m-1)(r-1)!}{m!(r-1-m)!} + \frac{(r-2)!}{(m-1)!(r-1-m)!}\right). \tag{15}$$

To prove (14) we show by induction on m that, for all $m = 2, 3, ..., s$, the following relation holds:

$$\mathbf{E}f(\eta_k) - \mathbf{E}f(\zeta_k) - \sum_{r=2}^{m}(-1)^r (r-1)\frac{p_k^r}{r!}\mathbf{E}\Delta^r f(\eta_k)$$

$$= \sum_{r=m+1}^{\infty}\left(1 - \sum_{j=2}^{m}(-1)^j (j-1) C_r^j\right)\frac{p_k^r}{r!}\Delta^r f(0). \tag{16}$$

For $m = 2$, by (12) and (13), the equality is true. Now suppose that (16) is valid for some $m \geq 2$. Then, by (12) and (15),

$$\sum_{r=m+1}^{\infty}\left(1 - \sum_{j=2}^{m}(-1)^j (j-1) C_r^j\right)\frac{p_k^r}{r!}\Delta^r f(0) =$$

$$(-1)^m m \frac{p_k^{m+1}}{(m+1)!}\Delta^{m+1} f(0) + \sum_{r=m+2}^{\infty}\left(1 - \sum_{j=2}^{m}(-1)^j (j-1) C_r^j\right)\frac{p_k^r}{r!}\Delta^r f(0) =$$

$$(-1)^m m \frac{p_k^{m+1}}{(m+1)!}\mathbf{E}\Delta^{m+1} f(\eta_k) + \sum_{r=m+2}^{\infty}\left(1 - \sum_{j=2}^{m+1}(-1)^j (j-1) C_r^j\right)\frac{p_k^r}{r!}\Delta^r f(0),$$

and hence (16) is true for $m+1$. Thus (16) is valid for $m = s$.

Finally, for (14) to be proved, it remains only to estimate the right-hand side of the equality (16) for $m = s$. Because of (15) we have

$$\left|\sum_{r=s+1}^{\infty}\left(1 - \sum_{j=2}^{s}(-1)^j (j-1) C_r^j\right)\frac{p_k^r}{r!}\Delta^r f(0)\right| =$$

$$\left| \sum_{r=s+1}^{\infty} \left(\frac{(s-1)(r-1)!}{s!(r-1-s)!} + \frac{(r-2)!}{(s-1)!(r-1-s)!} \right) \frac{p_k^r}{r!} \Delta^r f(0) \right| =$$

$$\left| p_k^{s+1} \sum_{r=0}^{\infty} \frac{p_k^r}{(r+s+1)!} \left(\frac{(s-1)(r+s)!}{s!r!} + \frac{(r+s-1)!}{(s-1)!r!} \right) \sum_{j=0}^{r} \frac{(-1)^{r-j} r!}{j!(r-j)!} \Delta^{s+1} f(j) \right| =$$

$$\left| p_k^{s+1} \sum_{j=0}^{\infty} \sum_{t=0}^{\infty} \frac{p_k^j}{j!} (-1)^t p_k^t \left(\frac{(s-1)(t+j+s)!}{s!(t+j+s+1)!t!} + \frac{(t+j+s-1)!}{(s-1)!(t+j+s+1)!t!} \right) \Delta^{s+1} f(j) \right|, \tag{17}$$

where the last expression was derived by changing the order of summing and substituting the variable $t = r - j$. To estimate the expression (17) it suffices to note that

$$\left| \sum_{t=0}^{\infty} (-1)^t p_k^t \left(\frac{(s-1)(t+j+s)!}{s!(t+j+s+1)!t!} + \frac{(t+j+s-1)!}{(s-1)!(t+j+s+1)!t!} \right) \right| \le \frac{s}{(s+1)!},$$

and hence the expression (17) is not greater than

$$\frac{s}{(s+1)!} p_k^{s+1} \sum_{j=0}^{\infty} \frac{p_k^j}{j!} |\Delta^{s+1} f(j)| = \frac{s}{(s+1)!} e^{p_k} p_k^{s+1} \mathbf{E} |\Delta^{s+1} f(\eta_k)|.$$

Therefore, (14) is true and hence (3) holds for $s \ge 2$. The theorem is proved. □

Proof of Corollary 2. The assertion follows from the two inequalities below which are easy consequences of Theorem 1:

$$\left| \mathbf{E} F(\bar{\zeta}) - \mathbf{E} F(\bar{\eta}) + \frac{1}{2} \sum_{j=1}^{n} p_j^2 \mathbf{E} \Delta_j^2 F(\bar{\psi}_j) \right| \le \sum_{j=1}^{n} e^{p_j} \frac{p_j^3}{3} \mathbf{E} |\Delta_j^3 F(\bar{\psi}_j)|,$$

$$\left| \sum_{k=1}^{n} p_k^2 \mathbf{E} \Delta_k^2 F(\bar{\psi}_k) - \sum_{k=1}^{n} p_k^2 \mathbf{E} \Delta_k^2 F(\bar{\eta}) \right| \le \sum_{k=1}^{n} p_k^2 \frac{1}{2} \sum_{j=1}^{k-1} p_j^2 \mathbf{E} |\Delta_j^2 \Delta_k^2 F(\bar{\psi}_j)|.$$

Corollary 3 is proven by repeated application of relation (3). Theorem 2 is the immediate consequence of Theorem 1 and Lemma 2. The proof of Corollary 5 is analogous to that of Corollary 2. Corollary 7 is proven by repeated application of relation (6). Corollary 6 follows from Corollary 7 for $l = 3$. □

4. Comparison with predecessors' results.

We compare the results of the chapter with the corresponding results of Barbour (1987) and Barbour, Chen and Choi (1995). Corollary 4 will be compared with the following theorem due to Barbour, Chen and Choi (1995):

THEOREM A. *Let* $\mathbf{E}Z^2|h(Z)| < \infty$. *Then*

$$|\mathbf{E}h(S) - \mathbf{E}h(Z)| \le \tfrac{1}{2}C\left(\sum_{j=1}^{n} p_j^2\right)\left(4\min\{1,\lambda^{-1}\}\mathbf{E}|h(Z+1)| + \mathbf{E}\Delta^2|h(Z)|\right),$$

where $C = \max_k \sup_{j\ge 0} \frac{\mathbf{P}(S-\zeta_k=j)}{\mathbf{P}(Z=j)}$.

Barbour, Chen and Choi (1995) obtained an upper bound for C which implies, in particular, that $C \le 2e^{13/12}\sqrt{\pi}$ for $\tilde{p} \le 1/2$.

Corollary 5 will be compared with the following theorem in the same paper.

THEOREM B. *Let* $\mathbf{E}Z^4|h(Z)| < \infty$. *Then*

$$|\mathbf{E}h(S) - \mathbf{E}h(Z) + \tfrac{1}{2}\sum_{j=1}^{n} p_j^2 \mathbf{E}\Delta^2 h(Z)| \le$$

$$C\left\{\left(\sum_{j=1}^{n} p_j^2\right)^2 R_1 + \left(\sum_{j=1}^{n} p_j^3\right) R_2\right\},$$

where

$$R_1 = 12\min\{1,\lambda^{-2}\}\mathbf{E}|h(Z+2)| +$$
$$\tfrac{1}{3}\min\{1,\lambda^{-1}\}(5\mathbf{E}\Delta^2|h(Z+1)| + \mathbf{E}\Delta^2|h(Z)|) +$$
$$\tfrac{1}{8}\mathbf{E}\Delta^4|h(Z)|,$$
$$R_2 = 2\min\{1,\lambda^{-1}\}(\mathbf{E}|h(Z+2)| + \mathbf{E}|h(Z+1)|) +$$
$$\tfrac{1}{3}(\mathbf{E}\Delta^2|h(Z+1)| + 2\mathbf{E}\Delta^2|h(Z)|),$$

and the constant C *is defined in Theorem A.*

For functions of at most polynomial growth complete asymptotic expansions were obtained by Barbour (1987). The following statement follows from Theorem 2, Remark 3 on it and equality (2.13) in Barbour (1987):

THEOREM C. *Let* $l \ge 1$, $H \ge 0$ *and* $t \ge 0$. *Let* h *be a real function of integer argument. Suppose that* $\lambda \ge 1$ *and, for all* y, $|\Delta^l h(y)| \le H(1 + \lambda^{-t/2}|y - [\lambda]|^t)$. *Then*

$$\left|\mathbf{E}h(S) - \mathbf{E}h(Z) + \sum_{s=1}^{l-1}\sum_{[s]}\prod_{j=1}^{s}\frac{1}{r_j!}\left(\frac{(-1)^j\sum_{i=1}^{n} p_i^{j+1}}{j+1}\right)^{r_j}\mathbf{E}\Delta^{r_1+\cdots+r_s+s}h(Z)\right|$$

$$\le KH\max_{(s)}\left\{\lambda^{-k/2}\prod_{j=1}^{k}\left(\sum_{i=1}^{n} p_i^{s_j+1}\right)\right\} \le KH\lambda^{l/2-1}\sum_{i=1}^{n} p_i^{l+1}, \tag{18}$$

where $\sum_{[s]}$ *denotes the sum over all* $(r_1, ..., r_s) \in (\mathbf{Z}^+)^s$ *such that* $\sum_{j=1}^{s} jr_j = s$*;* $\max_{(s)}$ *is taken over*

$$\left\{k \geq 1;\; s_j \geq 1, 1 \leq j \leq k;\; \sum_{j=1}^{k} s_j = l\right\};$$

K *is some constant depending only on* l *and* t.

But if $\lambda \leq 1$ *and* $|\Delta^l h(y)| \leq H(1+y^t)$*, then*

$$\left|\mathbf{E}h(S) - \mathbf{E}h(Z) + \sum_{s=1}^{l-1} \sum_{[s]} \prod_{j=1}^{s} \frac{1}{r_j!} \left(\frac{(-1)^j \sum_{i=1}^{n} p_i^{j+1}}{j+1}\right)^{r_j} \mathbf{E}\Delta^{r_1+\dots+r_s+s} h(Z)\right|$$

$$\leq KH \max_{(s)} \left\{\prod_{j=1}^{k} \left(\sum_{i=1}^{n} p_i^{s_j+1}\right)\right\}.$$

We shall compare the above-mentioned results in the case $\lambda \to \infty$, $\tilde{p} \to 0$, $h(y)$ being an arbitrary polynomial of order $m \geq 3$.

Define coefficients K_h^j by the relation

$$h(y) = \sum_{j=0}^{m} K_h^j y_{[j]}, \tag{19}$$

where $y_{[j]}$ denotes the so-called jth factorial power of y: $y_{[j]} = y(y-1)\cdots(y-j+1)$. By these moments we can obtain the simple representation for $\Delta^k h$:

$$\Delta^k h(y) = \sum_{j=k}^{m} j(j-1)\cdots(j-k+1) K_h^j y_{[j-k]}. \tag{20}$$

In particular,

$$\mathbf{E}\Delta^2 h(Z) = \mathbf{E}\sum_{j=2}^{m} j(j-1) K_h^j Z_{[j-2]} = \sum_{j=2}^{m} j(j-1) K_h^j \lambda^{j-2}.$$

It is easy to see that

$$\mathbf{E}|h(Z+1)| \sim |K_h^m| \lambda^m,$$

$$\mathbf{E}|\Delta^2 h(Z)| \sim \mathbf{E}\Delta^2 |h(Z)| \sim |\mathbf{E}\Delta^2 h(Z)| \sim |K_h^m| m(m-1)\lambda^{m-2} \tag{21}$$

as $\lambda \to \infty$. Now we compare the following resulting estimates for $|\mathbf{E}h(S) - \mathbf{E}h(Z)|$ given by the above-listed results:

Corollary 4: $K_2 \lambda^{m-2} \sum_{j=1}^{n} p_j^2$;

Theorem A: $K_3 \lambda^{m-1} \sum_{j=1}^{n} p_j^2$;

Theorem C: $K_1 \lambda^{m-3/2} \sum_{j=1}^{n} p_j^2$;

where K_1, K_2, K_3 are some positive constants which depend only on h. Note that, in the case under consideration, the constant H in Theorem C must be of order λ^{m-1}. We see that in this case the upper bound in Theorem A is rougher than that in Theorem C and that in Corollary 4.

Comparison of Corollary 5, Theorem B and Theorem C can be done analogously. We get the following bounds for $|\mathbf{E}h(S) - \mathbf{E}h(Z) - \frac{1}{2}\sum_{j=1}^n p_j^2 \mathbf{E}\Delta^2 h(Z)|$:

Corollary 5: $K_4 \lambda^{m-3} \sum_{j=1}^n p_j^3$;

Theorem B: $K_5 \lambda^{m-1} \sum_{j=1}^n p_j^3$;

Theorem C: $K_6 (\lambda^{m-3} (\sum_{j=1}^n p_j^2)^2 + \lambda^{m-5/2} \sum_{j=1}^n p_j^3)$;

where K_4, K_5, K_6 are constants depending only on h. To derive the first two of these three estimates the following inequality was used:

$$\left(\sum_{j=1}^n p_j^2\right)^2 \le \lambda \sum_{j=1}^n p_j^3.$$

Proposition 2. *Let $p_1 = \cdots = p_n = p$ and $h(y)$ be a polynomial of order $m \ge 2$. Suppose $\lambda \to \infty$ and $p \to 0$. Then*

$$|\mathbf{E}h(S) - \mathbf{E}h(Z)| \sim \tfrac{1}{2} np^2 \mathbf{E}|\Delta^2 h(Z)|,$$

$$\left|\mathbf{E}h(S) - \mathbf{E}h(Z) + \tfrac{1}{2} np^2 \mathbf{E}\Delta^2 h(Z)\right| \sim \tfrac{1}{3} np^3 \mathbf{E}|\Delta^3 h(Z)| + \tfrac{1}{8} n^2 p^4 \mathbf{E}|\Delta^4 h(Z)|,$$

$$\left|\mathbf{E}h(S) - \mathbf{E}h(Z) + \tfrac{1}{2} np^2 \mathbf{E}\Delta^2 h(Z) - \tfrac{1}{3} np^3 \mathbf{E}\Delta^3 h(Z) - \tfrac{1}{8} n^2 p^4 \mathbf{E}\Delta^4 h(Z)\right|$$
$$\sim \tfrac{1}{4} np^4 \mathbf{E}|\Delta^4 h(Z)| + \tfrac{1}{6} n^2 p^5 \mathbf{E}|\Delta^5 h(Z)| + \tfrac{1}{48} n^3 p^6 \mathbf{E}|\Delta^6 h(Z)|.$$

Thus the bounds in Corollaries 4, 5 *and* 6 *are asymptotically precise.*

Proof of Proposition 2. Calculating $\mathbf{E}h(S)$ and $\mathbf{E}h(Z)$ is very simple:

$$\mathbf{E}h(Z) = \sum_{j=0}^m K_h^j \lambda^j, \quad \mathbf{E}S^m = \sum_{j=0}^m K_h^j n_{[j]} p^j,$$

where coefficients K_h^j are defined by (19). Thus

$$\mathbf{E}h(Z) - \mathbf{E}h(S) \sim K_h^m \frac{m(m-1)}{2} \lambda^{m-1} p.$$

At the same time the following relation was already noted in (21):

$$\mathbf{E}|\Delta^2 h(Z)| \sim |K_h^m| m(m-1) \lambda^{m-2}.$$

Hence the estimate of Corollary 4 is asymptotically precise.

Now we proceed to proving the exactness of Corollary 5. First, consider the case $h(y) = y_{[m]}$. We have

$$\mathbf{E}h(Z) = \lambda^m,$$

$$\mathbf{E}h(S) = n_{[m]}p =$$

$$\lambda^m - \frac{m(m-1)}{2}\lambda^{m-1}p + \left(\sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1} ij\right)\lambda^{m-2}p^2 + O(\lambda^{m-3}p^3),$$

$$\frac{1}{2}np^2\mathbf{E}\Delta^2 h(Z) = \frac{m(m-1)}{2}\lambda^{m-1}p.$$

Hence

$$\left|\mathbf{E}h(Z) - \mathbf{E}h(S) - \tfrac{1}{2}np^2\mathbf{E}\Delta^2 h(Z)\right| \sim \left(\sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1} ij\right)\lambda^{m-2}p^2 =$$

$$\tfrac{1}{24}m(m-1)(m-2)(3m-1)\lambda^{m-2}p^2.$$

On the other hand, by (20) we have

$$\tfrac{1}{8}n^2p^4\mathbf{E}|\Delta^4 h(Z)| + \tfrac{1}{3}np^3\mathbf{E}|\Delta^3 h(Z)| =$$

$$\left(\tfrac{1}{8}m(m-1)(m-2)(m-3) + \tfrac{1}{3}m(m-1)(m-2)\right)\lambda^{m-2}p^2 =$$

$$\tfrac{1}{24}m(m-1)(m-2)(3m-1)\lambda^{m-2}p^2.$$

So, the assertion is asymptotically precise for $h(y) = y_{[m]}$. It is easy to understand that, because of (19), this assertion is also asymptotically precise for any polynomial of order m.

The proof of the exactness of Corollary 6 can be conducted analogously using the following identity:

$$\sum_{i=1}^{m-3}\sum_{j=i+1}^{m-2}\sum_{k=j+1}^{m-1} ijk = \tfrac{1}{48}m^2(m-1)^2(m-2)(m-3) = \tfrac{1}{4}m_{[4]} + \tfrac{1}{6}m_{[5]} + \tfrac{1}{48}m_{[6]}.$$

5. The approximation for arbitrary distributions.

The content of this section is based on and to a considerable extent repeats the idea of Kchinchine (1933) [cf. Borovkov (1988), Borisov (1993, 1996)]. We apply the results of Section 1 to approximation of vectors of r.v.'s with arbitrary, not necessarily Bernoulli, distributions. Let ξ_1,...,ξ_n be independent r.v.'s in an arbitrary measurable Abelian group $\mathcal{A}$ with distributions $Q_1, ..., Q_n$, respectively. The "+" operation in $\mathcal{A}$ is assumed to be measurable. Let $P_1, ..., P_n$ be the accompanying Poisson distributions for $Q_1, ..., Q_n$, respectively, and let $\beta_1, ..., \beta_n$ be independent r.v.'s with distributions $P_1, ..., P_n$, respectively. Finally, let $G(y_1, ..., y_n)$ be an arbitrary measurable function of n arguments in $\mathcal{A}$. We evaluate the difference

$$\mathbf{E}G(\xi_1, ..., \xi_n) - \mathbf{E}G(\beta_1, ..., \beta_n)$$

when both of the expectations exist.

Denote by F the following expectation:

$$F(k_1, ..., k_n) = \mathbf{E}G(\tau_1^{*k_1}, ..., \tau_n^{*k_n}),$$

where $\tau_j^{*k} = \tau_j^{(1)} + \cdots + \tau_j^{(k)}$ is the sum of k independent r.v.'s such that each of them has the distribution equal to the conditional distribution of ξ_j under the condition $\xi_j \neq 0$. All the r.v.'s $\tau_1^{(1)}$, $\tau_1^{(2)}$, ..., $\tau_n^{(1)}$, $\tau_n^{(2)}$, ... are supposed to be independent. Let $p_j = \mathbf{P}(\xi_j \neq 0)$, $j = 1, ..., n$. As in Section 1, $\zeta_1, ..., \zeta_n$ denote independent Bernoulli r.v.'s with the success probabilities $p_j = \mathbf{P}(\zeta_j = 1)$, and $\eta_1, ..., \eta_n$ denote independent Poisson r.v.'s with parameters $p_1, ..., p_n$, respectively. We have

$$\mathbf{E}G(\xi_1, ..., \xi_n) = \mathbf{E}F(\zeta_1, ..., \zeta_n), \qquad \mathbf{E}G(\beta_1, ..., \beta_n) = \mathbf{E}F(\eta_1, ..., \eta_n). \tag{22}$$

Actually, these identities can be easily deduced from the corresponding results in Khintchine (1933) [cf. Borovkov (1988)]. These relations allow us to apply Theorem 1 and its Corollaries to the approximation for vectors of independent arbitrarily distributed r.v.'s in a measurable Abelian group.

In the case when, for each j, the conditional distribution of ξ_j under the condition $\xi_j \neq 0$ coincides with some distribution Q independent of j, and

$$G(y_1, ..., y_n) = g(y_1 + \cdots + y_n)$$

Theorem 2 and its corollaries can be used. Denote by $\tau_1, \tau_2, ...$ i.i.d r.v.'s with distribution Q. For

$$h(k) = \mathbf{E}g(\tau_1 + \cdots + \tau_k)$$

then the equalities

$$\mathbf{E}G(\xi_1, ..., \xi_n) = \mathbf{E}h(S), \qquad \mathbf{E}G(\beta_1, ..., \beta_n) = \mathbf{E}h(Z)$$

hold where $S = \zeta_1 + \cdots + \zeta_n$, $Z = \eta_1 + \cdots + \eta_n$. In fact, these relations were obtained by Khintchine (1933) [cf. Borisov (1993, 1996)]. It is clear that these representations are equivalent to (22). They reduce the problem of Poisson approximation in an abstract sample space to investigation of closeness of a binomial and the corresponding accompanying Poisson distributions.

Example. Let $\xi_1, ..., \xi_n$ be arbitrary r.v.'s on the real line. Suppose that, for all j, the conditional distributions of ξ_j under the condition $\xi_j \neq 0$ coincide, and

$$G(y_1, ..., y_n) = (y_1 + \cdots + y_n)^l.$$

Also suppose that $\mathbf{E}(\tau_1)^l < \infty$. We have

$$h(k) = \mathbf{E}(\tau_1 + \cdots + \tau_k)^l = (\mathbf{E}\tau_1)^l k_{[l]} + B_{l-1}(k) + \mathbf{E}\tau_1^l k,$$

where $B_{j-1}(k)$ is a polynomial of k of order $\leq j-1$ whose coefficients depend only on expectations $\mathbf{E}\tau_1, ..., \mathbf{E}\tau_1^{l-1}$. Hence

$$\Delta^2 h(k) = (\mathbf{E}\tau_1)^l k_{[l-2]} + B'_{l-3}(k),$$

where $B'_{l-3}(k)$ is a polynomial of k of order $\leq l-3$ whose coefficients depend only on $\mathbf{E}\tau_1, ..., \mathbf{E}\tau_1^{l-1}$. Therefore, because of Corollary 4,

$$(23)\quad |\mathbf{E}(\xi_1 + \cdots + \xi_n)^l - \mathbf{E}(\beta_1 + \cdots + \beta_n)^l| \leq \frac{1}{2}\frac{e^{\tilde{p}}}{(1-\tilde{p})^2}\sum_{j=1}^{n} p_j^2 \left((\mathbf{E}\tau_1)^{l-2}\lambda^{l-2} + B''_{l-3}(\lambda)\right),$$

where $B''_{l-3}(\lambda)$ is a polynomial of λ of order $\leq l-3$ with coefficients depending only on $\mathbf{E}\tau_1, ..., \mathbf{E}\tau_1^{l-1}$.

Barbour (1987) obtained complete asymptotic expansions for $\mathbf{E}g(\xi_1+\cdots+\xi_n) - \mathbf{E}g(\beta_1 + \cdots + \beta_n)$ in the case when $\xi_1, ..., \xi_n$ are nonnegative integer r.v.'s and the function g is of at most polynomial growth. But in the case under consideration, when $g(k) = k^l$ and, for all j, the conditional distributions of ξ_j under the condition $\xi_j \neq 0$ coincide, these expansions, in general, don't allow us to separate the parameters $p_1, ..., p_n$ and moments of τ_1 as it is done in (23). Such separation in Barbour (1987) is possible only for some simplest classes of distributions but not for arbitrary.

6. Expansions under lesser moment restrictions.

In this section, under lesser moment restrictions than those in the theorems and corollaries in Section 1, the asymptotic expansions are studied. However, these expansions may appear inconvenient in case of nonidentically distributed r.v.'s. We use the notations p_k, λ_k, λ, $\bar{\zeta}$, $\bar{\eta}$, $\bar{e_k}$, Δ_k that were defined in Section 1. In this section, for the sake of convenience, we also suppose that $p_k \neq 0$ for all k.

At first we give complete asymptotic expansions for $\mathbf{E}F(\bar{\zeta}) - \mathbf{E}F(\bar{\eta})$.

COROLLARY 8. *Let $\mathbf{E}{\eta_k}^{l+1}|F(\bar{\eta})| < \infty$ for all k and some $l \geq 1$. Then*

$$\left|\mathbf{E}F(\bar{\zeta}) - \mathbf{E}F(\bar{\eta}) + \sum\nolimits' (-1)^s \frac{p_{k_1}^{r_1}\cdots p_{k_s}^{r_s}}{r_1!\cdots r_s!}\mathbf{E}\Delta_{k_1}^{r_1}\cdots\Delta_{k_s}^{r_s}F(\bar{\eta}^{(k_1,...,k_s)})\right|$$

$$\leq \sum\nolimits'' e^{p_{k_s}} \frac{p_{k_1}^{r_1}\cdots p_{k_{s-1}}^{r_{s-1}} p_{k_s}^{l+s-r_1-\cdots-r_{s-1}}}{r_1!\cdots r_{s-1}!(l+s-r_1-...r_{s-1})!}$$
$$\times \mathbf{E}|\Delta_{k_1}^{r_1}\cdots\Delta_{k_{s-1}}^{r_{s-1}}\Delta_{k_s}^{l+s-r_1-...r_{s-1}}F(\bar{\phi}_{k_s}^{(k_1,...,k_{s-1})} + \eta_{k_s}\bar{e_{k_s}})|,$$

where the right-hand side of the inequality is finite, and $\bar{\eta}^{(k_1,...,k_s)} = \bar{\eta} - \eta_{k_1}\bar{e_{k_1}} - \cdots - \eta_{k_s}\bar{e_{k_s}}$, $\bar{\phi}_{k_s}^{(k_1,...,k_s)} = \bar{\phi}_{k_s} - \eta_{k_1}\bar{e_{k_1}} - \cdots - \eta_{k_s}\bar{e_{k_s}}$; $\sum'$ and $\sum''$ denote the following sums:

$$\sum\nolimits' = \sum_{s=1}^{l-1}\sum_{k_1=1}^{n}\sum_{r_1=2}^{l}\sum_{k_2=1}^{k_1-1}\sum_{r_2=2}^{l+1-r_1}\cdots\sum_{k_s=1}^{k_{s-1}-1}\sum_{r_s=2}^{l+s-1-r_1-\cdots-r_{s-1}},$$

$$\sum\nolimits'' = \sum_{s=1}^{l}\sum_{k_1=1}^{n}\sum_{r_1=2}^{l}\sum_{k_2=1}^{k_1-1}\sum_{r_2=2}^{l+1-r_1}\cdots\sum_{k_{s-1}=1}^{k_{s-2}-1}\sum_{r_{s-1}=2}^{l+s-2-r_1-\cdots-r_{s-2}}\sum_{k_s=1}^{k_{s-1}-1}.$$

In the last sum we put $k_0 = n+1$ *if* $s = 1$.

This corollary is proven through subsequent application of relations (1) and (2) of Theorem 1.

Further we consider the sums $S = \zeta_1 + \dots + \zeta_n$ and $Z = \eta_1 + \dots + \eta_n$. As in Section 1, put $\tilde{p}_k = \max\{p_1, ..., p_k\}$, $\tilde{p} = \tilde{p_n}$ and let h be an arbitrary real function of integer nonnegative argument. For the sake of convenience we put $\tilde{p}_0 = 0$. The following corollaries are proved by using relations (4) and (5) of Theorem 2:

Corollary 9. *Let* $\mathbf{E}|\Delta^3 h(Z)| < \infty$. *Then*

$$|\mathbf{E}h(S) - \mathbf{E}h(Z) + \frac{1}{2}\sum_{j=1}^{n} p_j^2 \mathbf{E}\Delta^2 h(Z)| \le$$

$$\frac{e^{\tilde{p}}}{(1-\tilde{p})^2}\left\{\frac{1}{4}\sum_{k=1}^{n} p_k^2 \sum_{j=1}^{k-1} p_j^2 \mathbf{E}|\Delta^4 h(Z^{(k)})| + \frac{2}{3}(\sum_{k=1}^{n} p_k^3)\mathbf{E}|\Delta^3 h(Z)|\right\},$$

where $Z^{(k)}$ *is a Poisson r.v. with parameter* $\lambda - p_k$.

Remark 5. Because of the obvious upper bound for the corresponding Radon–Nykodim derivative the inequality

$$\mathbf{E}|\Delta^4 h(Z^{(k)})| \le e^{p_k}\mathbf{E}|\Delta^4 h(Z)|$$

holds. The right-hand side of the inequality may be infinite while the left-hand side is finite if $\mathbf{E}|h(Z)| < \infty$ and $p_k \ne 0$.

Proof of Corollary 9. Because of relations (4) and (5) of Theorem 2, the following inequalities hold:

$$\left|\mathbf{E}h(Z) - \mathbf{E}h(S) - \frac{1}{2}\sum_{k=1}^{n} p_k^2 \mathbf{E}\Delta^2 h(T_k)\right| \le \frac{e^{\tilde{p}}}{(1-\tilde{p})^2}\sum_{k=1}^{n}\frac{p_k^3}{6}\mathbf{E}|\Delta^3 h(Z)|,$$

$$\left|\sum_{k=1}^{n} p_k^2 \mathbf{E}\Delta^2 h(T_k) - \sum_{k=1}^{n} p_k^2 \mathbf{E}\Delta^2 h(Z^{(k)})\right| \le \sum_{k=1}^{n} p_k^2 \frac{1}{2}\frac{e^{\tilde{p}}}{(1-\tilde{p})^2}\sum_{j=1}^{k-1} p_j^2 \mathbf{E}|\Delta^4 h(Z^{(k)})|,$$

$$\left|\sum_{k=1}^{n} p_k^2 \mathbf{E}\Delta^2 h(Z) - \sum_{k=1}^{n} p_k^2 \mathbf{E}\Delta^2 h(Z^{(k)})\right| = \left|\sum_{k=1}^{n} p_k^2 \sum_{r=1}^{\infty}\frac{p_k^r}{r!}\mathbf{E}\Delta^{r+2} h(Z^{(k)})\right| \le$$

$$\sum_{k=1}^{n} p_k^2 e^{p_k} p_k \mathbf{E}|\Delta^3 h(Z)|,$$

where the proof of the last inequality is analogous to that of inequality (2) in Theorem 1. The above three inequalities immediately imply the assertion. □

We see that the corollary contains lesser restrictions on moments than those in Corollary 5 or Theorem B.

Corollary 10. *Let* $\mathbf{E}|\Delta^{l+1}h(Z)|<\infty$ *for some* $l\geq 1$. *Then*

$$\left|\mathbf{E}h(S)-\mathbf{E}h(Z)+\sum{}'(-1)^s\frac{p_{k_1}^{r_1}\cdots p_{k_s}^{r_s}}{r_1!\cdots r_s!}\mathbf{E}\Delta^{r_1+\cdots+r_s}h(Z^{(k_1,\ldots,k_s)})\right|$$

$$\leq \sum{}''\frac{e^{\tilde p_{k_s}}}{(1-\tilde p_{k_s})^2}\frac{p_{k_1}^{r_1}\cdots p_{k_{s-1}}^{r_{s-1}}p_{k_s}^{l+s-r_1-\cdots-r_{s-1}}}{r_1!\cdots r_{s-1}!(l+s-r_1-...r_{s-1})!}\mathbf{E}|\Delta^{l+s}h(Z^{(k_1,\ldots,k_{s-1})})|,$$

where $Z^{(k_1,\ldots,k_s)}=Z-\eta_{k_1}-\cdots-\eta_{k_s}$, $\sum'$ *and* $\sum''$ *denote the following sums*:

$$\sum{}'=\sum_{s=1}^{l-1}\sum_{k_1=1}^{n}\sum_{r_1=2}^{l}\sum_{k_2=1}^{k_1-1}\sum_{r_2=2}^{l+1-r_1}\cdots\sum_{k_s=1}^{k_{s-1}-1}\sum_{r_s=2}^{l+s-1-r_1-\cdots-r_{s-1}},$$

$$\sum{}''=\sum_{s=1}^{l}\sum_{k_1=1}^{n}\sum_{r_1=2}^{l}\sum_{k_2=1}^{k_1-1}\sum_{r_2=2}^{l+1-r_1}\cdots\sum_{k_{s-1}=1}^{k_{s-2}-1}\sum_{r_{s-1}=2}^{l+s-2-r_1-\cdots-r_{s-2}}\sum_{k_s=1}^{k_{s-1}-1}.$$

In the last sum we suppose $k_0=n+1$ *if* $s=1$.

This corollary is proven by subsequent application of relations (4) and (5).

It was already noted that the right-hand side of the above inequality is finite since, for any $s\geq 1$,

$$\mathbf{E}|\Delta^{l+s}h(Z^{(k_1,\ldots,k_{s-1})})|<\infty$$

if $\mathbf{E}|\Delta^{l+1}h(Z)|<\infty$.

By Lemma 3, the expectations of functions of $Z^{(k_1,\ldots,k_s)}$ can be expressed through expectations of functions of Z. However, application of Lemma 3 leads to necessity for enforcing restrictions on the moments.

8 Asymptotic expansion for expectations of smooth functions in the central limit theorem

1. Preliminaries

We denote by ξ_j, $j = \overline{1,n}$, independent identically distributed RVs with $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$, and set $S_n = n^{-1/2}\sum_{j=1}^n \xi_j$. Throughout the sequel, we denote by C, c, C_j, and c_j positive constants which are independent of n and the probability characteristics of the RV ξ_1, and denote by N and N_j independent RVs with the standard normal distribution.

The method of the chapter is based on the following inequality:

$$\mathbf{E}f(S_n) - \mathbf{E}f(N) = \int (\mathbf{P}(N < x) - \mathbf{P}(S_n < x))\, df(x), \tag{1}$$

where existence of the corresponding integrals is certainly presumed. It then seems natural to use the classical expansion for the distribution function of S_n [93, p. 197]. However, the estimate for the remainder term in this expansion is meaningful only if, in addition to the natural moment constraints, the well-known Cramér condition holds:

$$\lim_{|t|\to\infty} \sup |h(t)| < 1, \tag{2}$$

where $h(t)$ is the characteristic function of the RV ξ_1. In this case, by (1) we can easily obtain the complete asymptotic expansion for $\mathbf{E}f(S_n)$ as a consequence of the above-mentioned results under the minimal (of the known) constraints on the function f.

Theorem 1. *If $\mathbf{E}|\xi_1|^k < \infty$ and $\int \frac{1}{1+|x|^k}|df(x)| < \infty$ for some integer $k \geq 4$ then under condition (2)*

$$\left|\mathbf{E}f(S_n) - \mathbf{E}f(N) - \sum_{j=1}^{k-3} \frac{1}{n^{j/2}} \int Q_j(x)\, df(x)\right| \leq C\frac{\mathbf{E}|\xi_1|^k}{n^{(k-2)/2}}, \tag{3}$$

where $Q_j(x)$ are the standard terms of the expansion for the distribution function of S_n which depend only on the moments of ξ_1 up to the order $j+2$ inclusively (see [93, p. 171]*).*

For $k = 3$ (in this case the sum on the left-hand side of (3) is absent), estimate (3) is valid without condition (2).

To reject the excessively rigid constraint (2), we use some special smoothing and the ideas of Lindeberg's operator method.

2. The Main Result

Theorem 2. *Suppose that $f \in C^2(R)$ and*

$$|f''(x) - f''(y)| \leq C|x-y|^\varepsilon (1 + |x|^{2-\varepsilon} + |y|^{2-\varepsilon})$$

for all $x, y \in R$, *with* $0 < \varepsilon \le 1$. *If* $E|\xi_1|^4 < \infty$ *then*

$$\mathbf{E}f(S_n) = \mathbf{E}f(N) + \frac{1}{\sqrt{n}} \int f(x)\, dQ_1(x) + \frac{1}{n} H, \tag{4}$$

where

$$Q_1(x) = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{x^2-1}{6} \mathbf{E}\xi_1^3, \quad |H| \le C_1 (\mathbf{E}|\xi_1|^4)^2.$$

R e m a r k. If f is an even function or $\mathbf{E}\xi_1^3 = 0$ (for instance, the RV ξ_1 is symmetrically distributed) then the second summand on the right-hand side of (4) vanishes and the estimate for the convergence rate improves:

$$|\mathbf{E}f(S_n) - \mathbf{E}f(N)| \le \frac{|H|}{n}.$$

The classical expansion by von Bahr for $\mathbf{E}|S_n|^p$, $p > 0$, under the condition of existence of $\mathbf{E}\xi_1^4$ makes it possible to obtain an estimate of order $O(n^{-(p^*+1)/2})$ for the convergence rate, where $p^* = \min(p, 1)$. It is easy to see that for $p \ge 2$ an analogous result follows from Theorem 2. It is worth nothing that, with Cramér's condition (2) holding and the fourth moment of ξ_1 existent, Theorem 1 enables us to obtain the convergence rate $O(n^{-1})$ for the power functions $f(x) = |x|^p$ for every $p \in [0, 4]$. At the same time, by von Bahr's result this estimate is valid only if $p \in [1, 4]$. It can be shown that, for a lattice distribution (i.e., without condition (2)) of ξ_1, the order $n^{-(p+1)/2}$ of the estimate is unimprovable for $p < 1$.

Note that von Bahr obtained the complete asymptotic expansion. However, the version of the Fourier method that he used is inapplicable to a function class wider than $|x|^p$, $p > 0$.

A. Barbour [6], using the Stein method, obtained the complete asymptotic expansions for moments of smooth functions of sums of independent random variables. However, the relations he established between the order of the highest moment and the order of growth of the highest derivative are not optimal. For example, under the conditions of Theorem 2 on the function f, Barbour's result allows us to obtain an estimate of order $O(n^{-1})$ for the corresponding remainder term only on the condition of existence of the sixth (not fourth!) moments of summands.

The method to be exposed allows us to distinguish also the third term of the asymptotic expansion, provided that the fifth moments of the summands exist (to this end, the seventh moment is required in [6]). However, in this case the order of smallness of the remainder term is slightly worse than the optimal one.

3. Proof of Theorem 2

To ensure the use of the optimal growth rate for the function $f(x)$ and its derivatives in Theorem 2, we apply the technique of truncations. Denote

$$\xi_{jn} = \xi_j \mathrm{I}\{|\xi_j| \le \sqrt{n}\,\} - \mathbf{E}\xi_j \mathrm{I}\{|\xi_j| \le \sqrt{n}\,\},\ j=\overline{1,n},\ \sigma^2 = \mathbf{E}\xi_{1n}^2,\ S_n^* = \frac{1}{\sqrt{n}}\sum_{j=1}^n \xi_{jn}.$$

Observe that $\sigma \le 1$ and $1-\sigma^2 \le \frac{2}{n}\mathbf{E}|\xi_1|^4$. Furthermore, it is easy to show that $\sigma^2 \ge 11/16$ for $n \ge 16(\mathbf{E}|\xi_1|^3)^2$.

We now use Lindeberg's operator method to demonstrate that $\mathbf{E}f(S_n)$ and $\mathbf{E}f\big(S_n^*\big)$ are close enough.

Lemma 1. *Let f satisfy the conditions of Theorem 2. Then for every $m \ge 4$*

$$\big|\mathbf{E}f(S_n) - \mathbf{E}f\big(S_n^*\big)\big| \le Cn^{-(m-2)/2}\mathbf{E}|\xi_1|^m \mathbf{E}|\xi_1|^3,$$

where the constant C depends only on f.

Proof. We have

$$\mathbf{E}f(S_n) - \mathbf{E}f\big(S_n^*\big) = \sum_{j=1}^n \mathbf{E}\int \left(f\left(x + \frac{1}{\sqrt{n}}\xi_j\right) - f\left(x + \frac{1}{\sqrt{n}}\xi_{jn}\right)\right) d\mathbf{P}(A_{jn} < x), \quad (5)$$

where

$$A_{jn} = \frac{1}{\sqrt{n}}\sum_{i=1}^{j-1}\xi_i + \frac{1}{\sqrt{n}}\sum_{i=j+1}^{n}\xi_{in}, \quad j=\overline{1,n}, \quad \sum_{i\in\emptyset} = 0.$$

We need the consequence of Rosenthal's inequality [108]:

$$\mathbf{E}|A_{jn}|^3 \le C_1\mathbf{E}|\xi_1|^3, \quad j=\overline{1,n},$$

and the estimates

$$|f(x)-f(y)| \le C_0|x-y|(1+|x|^3+|y|^3),$$
$$|\mathbf{E}\xi_1\mathrm{I}\{|\xi_1| \le \sqrt{n}\,\}| = |\mathbf{E}\xi_1\mathrm{I}\{|\xi_1| > \sqrt{n}\,\}| \le n^{-(m-1)/2}\mathbf{E}|\xi_1|^m$$

which also follow from the conditions of Theorem 2. Using these inequalities, we estimate the modulus of the right-hand side of (6) from above as follows:

$$\sum_{j=1}^n \mathbf{E}\int C_0\left|\frac{1}{\sqrt{n}}\xi_j - \frac{1}{\sqrt{n}}\xi_{jn}\right|\left(1+|x|^3+\left|\frac{1}{\sqrt{n}}\xi_j\right|^3+\left|\frac{1}{\sqrt{n}}\xi_{jn}\right|^3\right)d\mathbf{P}(A_{jn}<x)$$

$$\le \sum_{j=1}^n C_0\left[\frac{1}{\sqrt{n}}\mathbf{E}|\xi_j\mathrm{I}\{|\xi_j| > \sqrt{n}\,\} - \mathbf{E}\xi_j\mathrm{I}\{|\xi_j| > \sqrt{n}\,\}|(\mathbf{E}|A_{jn}|^3+1)\right.$$

$$+\frac{1}{n^2}\mathbf{E}|\xi_j \mathrm{I}\{|\xi_j| > \sqrt{n}\} - \mathbf{E}\xi_j \mathrm{I}\{|\xi_j| > \sqrt{n}\}|(|\xi_j|^3 + |\xi_{jn}|^3)\Big] \le C_2 n^{-(m-2)/2}\mathbf{E}|\xi_1|^m \mathbf{E}|\xi_1|^3.$$

Lemma 1 is proven.

Denote

$$\Delta_n(f) = \mathbf{E}f(S_n) - \mathbf{E}f(N), \quad \Delta_{n1}(f) = \mathbf{E}f\left(S_{n1} + \sqrt{\frac{b\log n}{n}}\sigma N_0\right) - \mathbf{E}f(N),$$

$$\Delta_{n2}(f) = \mathbf{E}f(S_n^*) - \mathbf{E}f\left(S_{n1} + \sqrt{\frac{b\log n}{n}}\sigma N_0\right),$$

where $S_n^* = S_{n1} + S_{n2}$, S_{n1} and S_{n2} are independent, $\mathbf{E}S_{n2}^2 \sim \frac{b\log n}{n}\sigma^2$, and b is a positive constant (independent of n) to be specified below. For $m = 4$, Lemma 1 implies the estimate

$$|\Delta_n(f) - \Delta_{n1}(f) - \Delta_{n2}(f)| \le \frac{c}{n}\mathbf{E}|\xi_1|^3 \cdot \mathbf{E}|\xi_1|^4.$$

Lemma 2. *If* $\mathbf{E}|\xi_1|^4 < \infty$, $f \in C^1(R)$, *and* $|f'(x)| \le C_1(1 + |x|^3)$ *for all* $x \in R$, *then*

$$\left|\Delta_{n1}(f) - \frac{1}{\sqrt{n}}\int f(x)\,dQ_1(x)\right| \le \frac{C_2}{n}(\mathbf{E}|\xi_1|^4)^2. \tag{6}$$

Proof. Denote $m = n - [b\log n]$. Represent $S_{n1} + \sqrt{\frac{b\log n}{n}}\sigma N_0$ as a sum of independent identically distributed RVs:

$$Y_m = \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\nu_{jn},$$

where

$$\nu_{jn} = \sqrt{\frac{m}{n}}\xi_{jn} + \sqrt{\frac{b\log n}{n}}\sigma N_j, \quad j = \overline{1,m}.$$

Note that $\mathbf{E}\nu_{1n} = 0$ and $\mathbf{E}\nu_{1n}^2 = \sigma^2$. Under the conditions of the lemma, $\mathbf{P}(|Y_m| > x) = o(x^{-N})$ as $x \to \infty$ for every $N > 0$. Moreover, $|f'(x)| \le c(1 + |x|^3)$; hence, the integral on the right-hand side of equality (1) is well defined with Y_m substituted for S_n. Now, we can apply the classical expansion of the distribution function of sums of independent identically distributed RVs:

$$\left|\mathbf{P}\left(\frac{1}{\sigma}Y_m < x\right) - \mathbf{P}(N < x) - \frac{Q_{1n}(x)}{\sqrt{m}}\right| \le$$

$$\frac{|Q_{2n}x)|}{m} + \frac{C}{\sigma^4(1+|x|)^4 m}\int\limits_{|y|\ge\sigma\sqrt{m}(1+|x|)}|y|^4\,dV(y)$$

$$+\frac{C}{\sigma^5(1+|x|)^5 m^{3/2}}\int\limits_{|y|<\sigma\sqrt{m}(1+|x|)}|y|^5\,dV(y) + \left(\sup_{|t|\ge\delta}|v(t)| + \frac{1}{2m}\right)^m \frac{Cm^{10}}{(1+|x|)^5}, \tag{7}$$

where $V(y)$ is the distribution function and $v(t)$ is the characteristic function of the RV ν_{1n} and

$$\delta = \frac{\sigma^2}{12\mathbf{E}|\nu_{1n}|^3}, \quad Q_{1n}(x) = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{x^2-1}{6} \mathbf{E}\nu_{1n}^3,$$

$$Q_{2n}(x) = -\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \left[\frac{x^5+15x-10x^3}{72} \left(\mathbf{E}\nu_{1n}^3\right)^2 + \frac{x^3-3x}{24} \left(\frac{1}{3} + \mathbf{E}\nu_{1n}^4 \right) \right].$$

The main difficulty in use of this expansion consists in estimating the last summand on the right-hand side of (7). Smoothing enables us to do this without imposing additional constraints on the distribution of ξ_1. We have

$$|v(t)| \leq \left| \mathbf{E} \exp \left\{ \mathrm{it} \sqrt{\frac{b \log n}{n}} \sigma N_1 \right\} \right| = \exp \left\{ -t^2 \sigma^2 \frac{b \log n}{n} \right\}.$$

Then

$$\left(\sup_{|t| \geq \delta} |v(t)| + \frac{1}{2m} \right)^m \leq C_3 \left(1 - C_4 \delta^2 \frac{b \log n}{n} \sigma^2 \right)^n \leq$$

$$C_3 \exp\{-C_4 \delta^2 \sigma^2 b \log n\} = C_3 n^{-C_4 \delta^2 \sigma^2 b}.$$

It is clear that, by choice of the constant b, we can make $C_4 \delta^2 \sigma^2 b$ arbitrarily large. It suffices that $b \geq C_5 (\mathbf{E}|\nu_{1n}|^3)^2 / \sigma^4$. The definition of ν_{1n} implies that this inequality holds, for instance, with $b = C_6 (\mathbf{E}|\xi_1|^3)^2$ for all n such that $n / \log n \geq b^{2/3}$.

Thus, we first prove (6) for all n satisfying the inequality

$$n \geq C_6 (\mathbf{E}|\xi_1|^3)^2. \tag{8}$$

Moreover, we can assume that $n / \log n \geq C_6 (\mathbf{E}|\xi_1|^3)^{4/3}$. Now, we successively estimate the integrals with respect to $df(x)$ separately for each summands on the right-hand side of (7).

From the definition of $Q_{2n}(x)$ and in view of (8), we have

$$\left| \int \frac{|Q_{2n}(x)|}{m} \, df(x) \right| \leq \frac{C_7}{n} \mathbf{E}|\xi_1|^4.$$

Furthermore, the estimate

$$\mathbf{P}(|\nu_{1n}| \geq t) \leq \mathbf{P} \left(\left| \sqrt{\frac{b \log n}{n}} \sigma N_1 \right| > t - \sqrt{m} \right)$$

is valid for every t. Using this inequality together with Chebyshev's inequality, we estimate the integral of the second summand on the right-hand side of (7) as follows:

$$\left| \int_R \frac{df(x)}{\sigma^4 (1+|x|)^4 m} \int_{|y| \geq \sigma \sqrt{m}(1+|x|)} |y|^4 \, dV(y) \right|$$

$$\leq \frac{C_8}{n}\left|\int\limits_0^\infty \frac{dx}{1+x}\int\limits_{\sigma\sqrt{m}(1+x)}^\infty y^4\, d\mathbf{P}(|\nu_{1n}|>y)\right| \leq \frac{C_8}{n}\int\limits_0^{1/\sigma}\frac{dx}{1+x}\mathbf{E}|\xi_1|^4$$

$$+\frac{C_9}{n}\int\limits_{1/\sigma}^\infty \frac{dx}{\sigma\sqrt{m}x^2}\int\limits_{\sigma\sqrt{m}x}^\infty y^4\mathbf{P}\left(\sqrt{\frac{b\log n}{n}}\sigma|N_1|>y-\sqrt{m}\right)dy \leq \frac{C_{10}}{n}\mathbf{E}|\xi_1|^4.$$

Using Fubini's theorem, we obtain the following estimate for the integral of the third summand on the right-hand side of (7):

$$\left|\int\limits_R \frac{df(x)}{\sigma^5(1+|x|)^5 m^{3/2}}\int\limits_{|y|<\sigma\sqrt{m}(1+|x|)}|y|^5\, dV(y)\right|$$

$$\leq \frac{C_{11}}{n}\left(\int\limits_0^\infty \frac{dx}{\sqrt{m}\,(1+x)^2}\int\limits_0^{\sigma\sqrt{m}(1+x)} y^4\mathbf{P}(|\nu_{1n}|>y)\,dy+\mathbf{E}|\xi_1|^4\right)$$

$$=\frac{C_{11}}{n}\left(\int\limits_0^{\sigma\sqrt{m}}\frac{y^4}{\sqrt{m}}\mathbf{P}(|\nu_{1n}|>y)\,dy\int\limits_0 \frac{dx}{(1+x)^2}\right.$$

$$\left.+\frac{1}{\sqrt{m}}\int\limits_{\sigma\sqrt{m}}^\infty y^4\mathbf{P}(|\nu_{1n}|>y)\int\limits_{y/\sigma\sqrt{m}-1}^\infty \frac{dx}{(1+x)^2}\,dy+\mathbf{E}|\xi_1|^4\right) \leq \frac{C_{12}}{n}\mathbf{E}|\xi_1|^4.$$

The estimate for the integral of the last summand is obvious.

Applying (1), Rosenthal's inequality [108], and the fact that the function $f(x)$ is Lipschitz continuous, we obtain

$$\left|\int\left(\mathbf{P}\left(\frac{1}{\sigma}Y_m<x\right)-\mathbf{P}(Y_m<x)\right)df(x)\right| \leq C_{13}\mathbf{E}\left|\left(\frac{1}{\sigma}-1\right)Y_m\right||Y_m|^3$$

$$\leq C_{14}(1-\sigma^2)\mathbf{E}|Y_m|^4 \leq \frac{C_{15}}{n}(\mathbf{E}|\xi_1|^4)^2.$$

From the definition of $Q_{1n}(x)$ and the simple inequality

$$\left|\mathbf{E}\nu_{1n}^3-\mathbf{E}\xi_1^3\right| \leq \frac{C_{16}}{\sqrt{n}}(\mathbf{E}|\xi_1|^3+\mathbf{E}|\xi_1|^4)$$

we infer that

$$\left|\int\left(\frac{Q_{1n}(x)}{\sqrt{m}}-\frac{Q_1(x)}{\sqrt{n}}\right)df(x)\right| \leq \frac{C_{17}}{n}(\mathbf{E}|\xi_1|^3+\mathbf{E}|\xi_1|^4).$$

In conclusion, note that (6) is also valid for $n \leq C_6(\mathbf{E}|\xi_1|^3)^2$. This ensues from the rather elementary estimates

$$|\Delta_{n1}(f)| \leq C_{18}\frac{(\mathbf{E}|\xi_1|^3)^2}{n}(1+\mathbf{E}|\xi_1|^4),$$

$$\left|\frac{1}{\sqrt{n}}\int f(x)\,dQ_1(x)\right| \le C_{19}\frac{\mathbf{E}|\xi_1|^3}{n}(1+\mathbf{E}|\xi_1|^3), (\mathbf{E}|\xi_1|^3)^2 = (\mathbf{E}|\xi_1|\xi_1^2)^2 \le \mathbf{E}\xi_1^4.$$

Lemma 2 is proven.

Lemma 3. *If $\mathbf{E}|\xi_1|^4 < \infty$, $f \in C^2(R)$, and the inequality $|f''(x) - f''(y)| \le C|x - y|^\varepsilon(1 + |x|^{2-\varepsilon} + |y|^{2-\varepsilon})$ holds for all x, $y \in R$ and some $0 < \varepsilon \le 1$, then*

$$|\Delta_{n2}(f)| \le \frac{c}{n}(\mathbf{E}|\xi_1|^3)^4. \tag{9}$$

Proof. We use the Taylor formula

$$f(x+\lambda) = f(x) + \lambda f'(x) + \frac{\lambda^2}{2}f''(x) + \lambda^2\int_0^1 (1-\theta)(f''(x+\theta\lambda) - f''(x))\,d\theta.$$

By successively setting $x = S_{n1}$, $\lambda = S_{n2}$ or $\lambda = \sigma\sqrt{\frac{b\log n}{n}}N_0$, we obtain

$$|\Delta_{n2}(f)| \le C_1\mathbf{E}\left\{|S_{n2}|^{2+\varepsilon}\int_0^1(1-\theta)\theta^\varepsilon(1 + |S_{n1}+\theta S_{n2}|^{2-\varepsilon} + |S_{n1}|^{2-\varepsilon})\,d\theta\right\}$$

$$+C_2\mathbf{E}\left\{\left|\sigma\sqrt{\frac{b\log n}{n}}N_0\right|^{2+\varepsilon}\int_0^1(1-\theta)\theta^\varepsilon\left(1 + \left|S_{n1}+\theta\sigma\sqrt{\frac{b\log n}{n}}N_0\right|^{2-\varepsilon} + |S_{n1}|^{2-\varepsilon}\right)d\theta\right\}$$

$$\le C_3\mathbf{E}|S_{n2}|^{2+\varepsilon} + C_4\left(\frac{b\log n}{n}\right)^{(2+\varepsilon)/2}.$$

Rosenthal's inequality and the definition of b imply the estimate

$$\mathbf{E}|S_{n2}|^{2+\nu} \le C_5\left(\frac{b\log n}{n}\right)^{1+\nu/2} \le \frac{c}{n}(\mathbf{E}|\xi_1|^3)^{2+\nu}$$

for every $\nu \in (0, 2]$. The last estimate implies (9). Lemma 3 and Theorem 2 therewith are proven.

9 Minimal smoothness conditions for asymptotic expansions of moments in the central limit theorem

1. Statement of the main results

The topic of this article was motivated by the results of [6] and [56] where asymptotic expansions were obtained under smoothness conditions close to optimal. These results are formulated in Theorems GH and B below.

The accuracy of approximation for expectations of functions in the CLT depends both on the smoothness of distributions of the summands and on the smoothness of the functions. It is interesting to note that, in the first case, to obtain the complete asymptotic expansions of the moments we need no additional smoothness conditions on the functions. An illustration of the fact is contained in Theorem P below.

Denote by ξ_i, $i = 1, \ldots, n$, some independent identically distributed (i.i.d.) random variables (r.v.'s) satisfying the conditions $\mathbf{E}\,\xi_1 = 0$ and $\mathbf{E}\,\xi_1^2 = 1$. Put $S_n = n^{-1/2} \sum_{i=1}^n \xi_i$ and denote by N an r.v.with standard Gaussian distribution. Introduce the classical Cramér regularity condition of the distributions:

$$\limsup_{|t|\to\infty} |h(t)| < 1, \tag{1.1}$$

where $h(t)$ is the characteristic function of ξ_1. Under the minimal (known) constraints on a function f, we can obtain complete asymptotic expansions of $\mathbf{E}\,f(S_n)$.

Theorem P. (see [93: p.171]) *If* $\mathbf{E}\,|\xi_1|^k < \infty$ *and* $\int \frac{1}{1+|x|^k} |df(x)| < \infty$ *for some integer* $k \ge 4$ *then, under Cramér's condition (1.1),*

$$\left| \mathbf{E}\,f(S_n) - \mathbf{E}\,f(N) - \sum_{i=1}^{k-3} \frac{1}{n^{i/2}} \int_{-\infty}^{\infty} Q_i(x) df(x) \right| \le C(f) \frac{\mathbf{E}\,|\xi_1|^k}{n^{(k-2)/2}}, \tag{1.2}$$

where the constant $C(f)$ *depends only on* f*,* $Q_i(x)$ *are the standard expansion members of the distribution function of* S_n*, which depend on the moments of* ξ_1 *of order up to* $i+2$*.*

If $k = 3$ then the above estimate (in this case, we omit the sum on the left-hand side of the inequality) holds without Cramér's condition.

This result follows immediately from the classical asymptotic expansions for the distribution function of S_n (see [93]) and the following simple representation:

$$\mathbf{E}\,f(S_n) - \mathbf{E}\,f(N) = \int_{-\infty}^{\infty} \big(\mathbf{P}\,(N < x) - \mathbf{P}\,(S_n < x)\big)\, df(x),$$

where both sides of the identity are well defined under the conditions of the theorem.

To obtain complete asymptotic expansions under the fixed moment constraints without Cramér's condition, we need stronger smoothness of $f(x)$ than in Theorem P. Roughly speaking, to obtain k members of the expansion (under the corresponding moment condition) we require $f(x)$ to be k times continuously differentiable. It is interesting to note that, for obtaining the analogous result by the well-known Lindeberg operator method, we need at least $3k$ derivatives of the functions $\big($see [10]$\big)$.

The following result was proven in [56].

Theorem GH. *Assume that, for some integer $k \geq 3$, the following conditions are fulfilled:*

$$\mathbf{E}\,|\xi_1|^k < \infty, \quad f \in C^{k-2}, \quad \sup_x \frac{|f^{(k-2)}(x)|}{1+|x|^2} < \infty.$$

Then expansion (1.2) holds with the following estimate of the remainder term η:

$$|\eta| = O\big(n^{-(k-2)/2}\big).$$

R e m a r k. If f is an even function or $\mathbf{E}\,\xi_1^3 = 0$ (for instance, if the r.v. ξ_1 is symmetrically distributed) then the first summand of the sum on the left-hand side of (1.2) vanishes and the estimate for the convergence rate can be improved (if the fourth moment of the summands is finite):

$$\big|\mathbf{E}\,f(S_n) - \mathbf{E}\,f(N)\big| \leq C_1(f)\frac{\mathbf{E}\,\boldsymbol{\xi}_1^4}{n}.$$

The classical expansion by von Bahr for $\mathbf{E}\,|S_n|^p$, $p > 0$, under the condition that $\mathbf{E}\,\xi_1^4$ exists, makes it possible to obtain an estimate of order $O\big(n^{-(p^*+1)/2}\big)$ for the convergence rate, where $p^* = \min(p,1)$. It is easy to see that, for $p \geq 2$, an analogous result follows from Theorem GH. It is worth noting that, whenever Cramér's condition (1.1) holds and the fourth moment of ξ_1 exists, Theorem P enables us to obtain the convergence rate $O(n^{-1})$ for the power functions $f(x) = |x|^p$ for every $p \in [0,4]$. At the same time, by von Bahr's result, this estimate is valid only if $p \in [1,4]$. In Theorem 1 below we show that, for lattice distributions of ξ_1 $\big($i.e; without condition (1.1)$\big)$, the order $n^{-(p+1)/2}$ of the estimate is unimprovable for $p < 1$.

Theorem 1. *Let ξ_i, $i \geq 1$, be a sequence of i.i.dinteger-valued r.v.'s satisfying the conditions: $\mathbf{E}\,\xi_1 = 0$, $\mathbf{E}\,\xi_1^2 = 1$, and $\mathbf{E}\,\xi_1^4 < \infty$. Let $f(x) = |x|^p h(x)$, where $0 < p < 1$ and $h(x)$ is an arbitrary twice continuously differentiable function with the following properties:*

$$h(x) = h(-x), \quad h(0) > 0,$$

$$|h(x)| < M, \quad |h''(x)| < M_1\big(1 + |x|^m\big),$$

where M and M_1 are constants and m is a natural. Then

$$\liminf_{n\to\infty} n^{(p+1)/2}\big|\mathbf{E}\,f(S_n) - \mathbf{E}\,f(N)\big| > 0.$$

The following result formulated in Theorem 2 gives a lower bound of the smoothness which still allows us to obtain an optimal estimate of the remainder term in the asymptotic expansions. In particular, we show that the smoothness conditions on the function $f(x)$ in Theorem GH cannot be improved.

Theorem 2. *Let a sequence $\varphi(n)$ be such that $\lim_{n\to\infty}\varphi(n)=\infty$. Then there exists a function $f\in C^{k+\alpha}(R)$, $k\in\{1,2,\dots\}$, $\alpha\in(0,1)$, such that, for any sequence ξ_i of i.i.dinteger-valued r.v.'s with the properties $\mathbf{E}\,\xi_i=0$, $\mathbf{E}\,\xi_i^2=1$, and $\mathbf{E}\,|\xi_i|^{k+3}<\infty$, the following relation holds:*

$$\limsup_{n\to\infty}\ \varphi(n)n^{\frac{k+\alpha}{2}}\left|\mathbf{E}\,f(S_n)-\mathbf{E}\,f(N)\quad-\frac{1}{n^{1/2}}\int_{-\infty}^{\infty}f(x)dQ_1(x)-\cdots\right.$$
$$\left.-\frac{1}{n^{k/2}}\int_{-\infty}^{\infty}f(x)dQ_k(x)\right|=\infty,$$

where $Q_i(x)$ are defined in Theorem P.

Now, we consider the case of nonidentically distributed summands. In this case, the most universal result was obtained in [6]. First, we introduce an array of row-wise independent centered r.v.'s $\xi_{n,i}$, $i=1,\dots,n$, satisfying the condition $\sum_{i=1}^n\mathbf{E}\,\xi_{n,i}^2=1$. Denote

$$S_n=\sum_{i=1}^n\xi_{n,i},\quad L_k=\sum_{i=1}^n\mathbf{E}\,|\xi_{n,i}|^k.$$

Theorem B. *For some integer $k\geq 2$, real $0\leq\alpha\leq 1$, and $p\geq 0$, let one of the following two conditions be fulfilled:*

1) $f\in C^{k-2}$ and

$$\sup_{x\neq y}\frac{\left|f^{(k-2)}(x)-f^{(k-2)}(y)\right|}{|x-y|^\alpha\left(1+|x|^p+|y|^p\right)}\leq H_1;$$

or

2) $f\in C^{k-1}$ and

$$\sup_{x\neq y}\frac{\left|f^{(k-1)}(x)-f^{(k-1)}(y)\right|}{|x-y|^\alpha\left(1+|x|^{p+1}+|y|^{p+1}\right)}\leq H_2.$$

Then, under the conditions $\mathbf{E}\,|\xi_{n,i}|^{k+p+\alpha}<\infty$, $i=1,\dots,n$, the following asymptotic expansion holds:

$$\mathbf{E}\,f(S_n)=\mathbf{E}\,f(N)+\sum_{(k-2)}(-1)^r\prod_{j=1}^r\frac{k_{s_j+2}(S_n)}{(s_j+1)!}\,\mathbf{E}\prod_{j=1}^r(D_{s_j+1}\Theta)f(N)+\eta,\tag{1.3}$$

where $k_t(S_n)$ is the cumulant of order t for S_n, $(D_{s_j+1}\Theta)f(N)$ is a function depending only on f, the symbol $\sum_{(k-2)}$ denotes the summation over the subsets of indices $\big\{r\geq 1,\ s_j\geq 1\ (1\leq j\leq r):\sum_{j=1}^r s_j\leq k-2\big\}$, and

$$|\eta|\leq C_jH_j\big(L_{k+\alpha}+L_{k+p+\alpha}\big),$$

$j=1,2$, with the constants C_j depending on p, k, and α.

Note that in Theorem B the correlation between the order of growth of the highest derivative and the order of the highest moments is not optimal. For example, to obtain an upper bound $O(n^{-1})$ for the third summand of the asymptotic expansion under the conditions on f imposed in Theorem GH, we must require existence of the 6 th moments of the summands (instead of the 4 th ones). The following assertion improves Theorem B.

Theorem 3. *Let $f \in C^{k-2}$, $k \geq 3$, and let, for some $0 \leq \alpha \leq 1$ and $p \geq 0$, the following relation hold:*

$$\sup_{x\neq y} \frac{\left|f^{(k-2)}(x) - f^{(k-2)}(y)\right|}{|x-y|^{\alpha}\left(1+|x|^p+|y|^p\right)} \leq H.$$

Then the expansion (1.3) holds with the upper bound

$$|\eta| \leq C\left(L_{k+\alpha} + L_{k+\alpha+\sigma}\right)$$

for the remainder term, where $\sigma = \max\{0, p-2\}$ and the constant C depends only on p, k, α, and H.

2. Proof of Theorems 1 and 2

Proof of Theorem 1. Put $\varepsilon = 1/\sqrt{n}$. By the definition of expectation, we have

$$\mathbf{E}\, f(S_n) = \sum_{i=-\infty}^{\infty} f(i\varepsilon)\mathbf{P}(S_n = i\varepsilon),$$

where the local probabilities satisfy the equality

$$\left(1+|x|^{k+3}\right)\left(\sqrt{n}\,\mathbf{P}(S_n = i\varepsilon) - \eta(x)\left(1+\varepsilon q_1(x) + \cdots + \varepsilon^{k+1} q_{k+1}(x)\right)\right) = o\left(\varepsilon^{k+1}\right)$$

uniformly in i (see [93: p.255]). Here $x = i\varepsilon$, $\eta(x) = e^{-x^2/2}/\sqrt{2\pi}$, $q_m(x)$ are polynomials, and $q_m(x)\eta(x) = \frac{dQ_m(x)}{dx}$. Therefore,

$$\mathbf{E}\, f(S_n) = \sum_{i=-\infty}^{\infty} f(i\varepsilon)\eta(i\varepsilon)\left(1+\varepsilon q_1(i\varepsilon) + \varepsilon^2 q_2(i\varepsilon)\right)\varepsilon$$

$$+\varepsilon_n \varepsilon^2 \sum_{i=-\infty}^{\infty} f(i\varepsilon)\left(1+(i\varepsilon)^4\right)^{-1}\varepsilon,$$

where $\varepsilon_n \to 0$. By estimating $|f(x)| \leq M|x|^p$, we derive that the second sum on the right-hand side of the relation has order $o(n^{-1})$, since the function $|x|^p/(1+x^4)$ is integrable. Under the conditions of the theorem, $f(x)$ is an even function and $q_1(x)$ is an odd function. Hence,

$$\sum_{i=-\infty}^{\infty} f(i\varepsilon)\eta(i\varepsilon)q_1(i\varepsilon)\varepsilon = 0.$$

Now, we will show that

$$|\mathbf{E}\, f(N) - \sum_{i=-\infty}^{\infty} f(i\varepsilon)\eta(i\varepsilon)\varepsilon| > K\varepsilon^{p+1}$$

whenever n is large enough. We have

$$\begin{aligned}
&\mathbf{E}\, f(N) - \sum_{i=-\infty}^{\infty} f(i\varepsilon)\eta(i\varepsilon)\varepsilon \\
&= 2\int_0^{\varepsilon/2} f(x)\eta(x)dx + 2\sum_{i\geq 1}\int_{(i-1/2)\varepsilon}^{(i+1/2)\varepsilon} \big[f(x)\eta(x) - f(i\varepsilon)\eta(i\varepsilon)\big]dx \\
&= 2\int_0^{\varepsilon/2} f(x)\eta(x)dx + 2\sum_{i\geq 1}(f\eta)''(\alpha_i)\varepsilon^3/24\,,
\end{aligned}$$

where $\alpha_i \in \big[(i-1/2)\varepsilon,\ (i+1/2)\varepsilon\big]$. The second derivative of the product $f\eta$ is as follows:

$$(f\eta)''(x) = p(p-1)|x|^{p-2}(h\eta)(x) + 2p|x|^{p-1}(h\eta)'(x) + |x|^p(h\eta)''(x).$$

Since the functions $|x|^{p-1}(h\eta)'(x)$ and $|x|^p(h\eta)''(x)$ are integrable, for all n we have:

$$\left|\sum_{i\geq 1} 2p|\alpha_i|^{p-1}(h\eta)'(\alpha_i)\varepsilon\right| < C_1\,,$$

$$\left|\sum_{i\geq 1} |\alpha_i|^{p}(h\eta)''(\alpha_i)\varepsilon\right| < C_2\,.$$

The last sums are bounded; therefore,

$$\begin{aligned}
&\liminf_{n\to\infty} \varepsilon^{-(p+1)}\big|\mathbf{E}\, f(S_n) - \mathbf{E}\, f(N)\big| \\
&\geq \liminf_{n\to\infty} \varepsilon^{-(p+1)}\left|2\int_0^{\varepsilon/2} f(x)\eta(x)dx + 2\sum_{i\geq 1} p(p-1)|\alpha_i|^{p-2}(h\eta)(\alpha_i)\varepsilon^3/24\right|.
\end{aligned}$$

It now remains to prove that the right-hand side of this inequality is strictly greater than zero. Using the mean value theorem, we can write down the following equalities:

$$\int_0^{\varepsilon/2} |x|^p h(x)e^{-x^2/2}/\sqrt{2\pi}\, dx = \frac{\beta_n}{\sqrt{2\pi}}h(0)\frac{1}{(p+1)2^{p+1}}\varepsilon^{p+1}, \quad \beta_n \to 1.$$

Next,

$$\sum_{i\geq 1} \frac{1}{24}p(p-1)|\alpha_i|^{p-2}(h\eta)(\alpha_i)\varepsilon^3$$

$$= \sum_{i\geq 1} \frac{1}{24} p(p-1) \frac{|\widetilde{\alpha}_i|^{p-2}}{(2\sqrt{n})^{p-2}} (h\eta)(\alpha_i)\varepsilon^3$$

$$= \sum_{i\geq 1} \frac{1}{3} \frac{p(p-1)}{2^{p+1}} |\widetilde{\alpha}_i|^{p-2} (h\eta)(\alpha_i)\varepsilon^{p+1}$$

$$= \frac{\varepsilon^{p+1}}{2^{p+1}} \sum_{i\geq 1} \frac{1}{3} p(p-1)|\widetilde{\alpha}_i|^{p-2}(h\eta)(\alpha_i),$$

where $\widetilde{\alpha}_i = 2\sqrt{n}\,\alpha_i \in [2i-1, 2i+1]$. Then

$$\varepsilon^{-(p+1)} \left| \int_0^{\varepsilon/2} f(x)\eta(x)dx + \sum_{i\geq 1} \frac{p(p-1)}{24} |\alpha_i|^{p-2}(h\eta)(\alpha_i)\varepsilon^3 \right|$$

$$= \left| \frac{\beta_n}{\sqrt{2\pi}} h(0) \frac{1}{(p+1)2^{p+1}} + \sum_{i\geq 1} \frac{1}{3} \frac{p(p-1)}{2^{p+1}} |\widetilde{\alpha}_i|^{p-2}(h\eta)(\alpha_i) \right|,$$

$$\left| \sum_{i\geq 1} \frac{1}{3} \frac{p(p-1)}{2^{p+1}} |\widetilde{\alpha}_i|^{p-2}(h\eta)(\alpha_i) \right|$$

$$\leq \frac{p(1-p)}{2^{p+1}3} \max_{x\in[0,\delta]} h(x)\eta(x) \sum_{1\leq i\leq \delta\sqrt{n}} \frac{1}{(2i-1)^{2-p}}$$

$$+ \frac{p(1-p)}{2^{p+1}3} \max_{x\in\mathbf{R}} h(x)\eta(x) \sum_{i\geq \delta\sqrt{n}} \frac{1}{(2i-1)^{2-p}}$$

$$\leq \frac{p(1-p)}{2^{p+1}3} \frac{h(0)}{\sqrt{2\pi}} \beta_\delta \left(1 + \frac{2^{p-1}}{1-p}\right) + \Delta(\delta, n),$$

where $\beta_\delta \to 1$ as $\delta \to 0$, and $\Delta(\delta, n) \to 0$ as $n \to \infty$ for all fixed δ. Moreover, for all δ, the following inequality holds:

$$\varepsilon^{-(p+1)} \left| \int_0^{\varepsilon/2} f(x)\eta(x)dx + \sum_{i\geq 1} \frac{p(p-1)}{24} |\alpha_i|^{p-2}(h\eta)(\alpha_i)\varepsilon^3 \right|$$

$$\geq \frac{\beta_n}{\sqrt{2\pi}} h(0) \frac{1}{(p+1)2^{p+1}} - \frac{1}{3} \frac{p(1-p)}{2^{p+1}} \frac{h(0)}{\sqrt{2\pi}} \beta_\delta \left(1 + \frac{2^{p-1}}{1-p}\right) - \Delta(\delta, n).$$

Taking the lower limits as $n \to \infty$ on both sides of this inequality and passing to the limit as $\delta \to 0$ on the right-hand side, we obtain:

$$\liminf_{n\to\infty} \varepsilon^{-(p+1)} \left| \int_0^{\varepsilon/2} f(x)\eta(x)dx + \sum_{i\geq 1} \frac{p(p-1)}{24} |\alpha_i|^{p-2}(h\eta)(\alpha_i)\varepsilon^3 \right|$$

$$\geq \frac{1}{\sqrt{2\pi}} h(0) \frac{1}{(p+1)2^{p+1}} - \frac{1}{3} \frac{p(1-p)}{2^{p+1}} \frac{h(0)}{\sqrt{2\pi}} \left(1 + \frac{2^{p-1}}{1-p}\right)$$

$$= \frac{h(0)}{\sqrt{2\pi}} \frac{1}{2^{p+1}} \left(\frac{1}{p+1} - \frac{1}{3} p(1-p) - \frac{1}{3} p 2^{p-1} \right)$$

$$\geq \frac{1}{12} \frac{h(0)}{\sqrt{2\pi}} \frac{1}{2^{p+1}}.$$

Theorem 1 is proven.

Proof of Theorem 2 reduces to an integral approximation problem by the classical expressions for the local probabilities of sums of independent lattice-valued r.v.'s. The key point of the proof (Lemma 2.2) is an application of the classical Banach–Steinhaus theorem (the boundedness principle for linear operators in a Banach space).

First, we formulate the following auxiliary result.

Lemma 2.1. *Let $f \in C^r[0,1]$ and $f(0) = f'(0) = \dots = f^{r-1}(0) = f(1) = f'(1) = \dots = f^{r-1}(1) = 0$. For $r = 1$, these constraints are not required. Then*

$$\int_0^1 f(x)dx - \sum_{k=1}^{n} f\left(\frac{k}{n}\right)\frac{1}{n} = O(n^{-r}), \quad n \to \infty.$$

This lemma follows readily from the Euler–Maclaurin summation formula (see [11: p.271]).

Before formulating the main auxiliary lemma, we introduce some notation. Let $0 < \alpha < 1$. Consider the Banach space

$$C^{k+\alpha}[0,1] = \left\{ f \in C^k[0,1] : \sum_{i=0}^{k} \sup_{x \in [0,1]} \left| f^{(k)}(x) \right| \right.$$

$$\left. + \sup_{x,y \in [0,1]} \left| f^{(k)}(x) - f^{(k)}(y) \right| / |x-y|^{\alpha} < \infty \right\},$$

where the expression in brackets determines the norm. We select in this space a fundamental subspace:

$$C_0^{k+\alpha}[0,1] = \Big\{ f \in C^{k+\alpha}[0,1] : \ f(0) = f'(0) = \dots = f^k(0)$$

$$= f(1) = f'(1) = \dots = f^k(1) = 0 \Big\}.$$

Lemma 2.2. *For any sequence $\varphi(n)$ satisfying the condition* $\lim_{n\to\infty} \varphi(n) = \infty$, *there exists a function $f \in C_0^{k+\alpha}[0,1]$ such that*

$$\limsup_{n\to\infty} \varphi(n) n^{k+\alpha} \left| \int_0^1 f(x)dx - \sum_{k=1}^{n} f\left(\frac{k}{n}\right)\frac{1}{n} \right| = \infty.$$

Proof. Consider the following family of continuous linear functionals on $C^{k+\alpha}[0,1]$:

$$F_n(g) = \varphi(n) n^{k+\alpha} \left[\int_0^1 g(x)dx - \sum_{k=1}^{n} g\left(\frac{k}{n}\right)\frac{1}{n} \right].$$

We prove that $\sup_n \|F_n\|^* = \infty$, where $\|\cdot\|^*$ is the norm in the corresponding dual space. For this purpose, we construct a sequence of functions $f_n(x)$ as follows: $f_n(x) = x^{k+\alpha}$ if $x \in [0, 1/4n]$, $f_n(x) = |x - 1/n|^{k+\alpha}$ if $x \in [3/4n, 1/n]$. Inside the interval $[1/4n, 3/4n]$, we define $f_n(x)$ by the equality $f_n(x) = P'(x)$, where $P(x)$ is Hermit's interpolating polynomial defined by the following conditions (put $h = 1/4n$):

$$P(h) = 0, \qquad P'(h) = h^{k+\alpha},$$

$$P''(h) = (k+\alpha)h^{k+\alpha-1}, \ \ldots, \ P^{(k+1)}(h) = (k+\alpha)\cdots(1+\alpha)h^{\alpha},$$

$$P(3h) = 0, \quad P'(3h) = h^{k+\alpha},$$

$$P''(3h) = (-1)(k+\alpha)h^{k+\alpha-1}, \ \ldots, \ P^{(k+1)}(3h) = (-1)^k(k+\alpha)\cdots(1+\alpha)h^{\alpha}.$$

This polynomial is constructed as follows: $P(x) = P_1(x) + P_2(x)$, $P_1(x) = \sum_{l=0}^{k+1} P^{(l)}(h)P_{1l}(x)$, $P_2(x) = \sum_{l=0}^{k+1} P^{(l)}(3h)P_{2l}(x)$, where

$$P_{1l}(x) = (x-3h)^{k+2}\frac{(x-h)^l}{l!}\left\{\frac{1}{(x-3h)^{k+2}}\right\}_{(h)}^{(k+1-l)},$$

$$P_{2l}(x) = (x-h)^{k+2}\frac{(x-3h)^l}{l!}\left\{\frac{1}{(x-h)^{k+2}}\right\}_{(3h)}^{(k+1-l)}.$$

The expression $F(x)_{(a)}^{(\lambda)}$ denotes the sum of the members in Taylor's expansion of $F(x)$ in the neighborhood of $x = a$ containing all degrees of $(x-a)$ which are less than a natural λ. Outside the interval $[0, 1/n]$, we continue $f_n(x)$ as a periodic function, i.e., $f_n(x + k/n) = f_n(x)$, $x \in [0, 1/n]$, $0 \le k \le n-1$. We now prove that the norm of the function $f_n(x)$ in the space $C^{k+\alpha}[0,1]$ is bounded by some constant not depending on n. For proving this, it suffices to estimate this norm on the interval $[0, 1/n]$. The main problem here is to estimate $\max_{x\in[1/4n,3/4n]} \left|P^{(l+1)}(x)\right|$, where $0 \le l \le k+1$. To this end, consider a separate member, in particular, in $P_1(x)$:

$$S = P^{(m)}(h)(x-3h)^{k+2}\frac{(x-h)^m}{m!}\,\frac{(-k-2)\cdots(-k-r-1)}{(-2h)^{k+2+r}}(x-h)^r,$$

$$0 \le m \le k+1, \quad 0 \le r \le k+1-m.$$

The derivative of order $l+1$ of S is a sum of members of the following type:

$$S_i = C(k,m,r,l,i)\frac{h^{k+\alpha+1-m}}{h^{k+2+r}}(x-3h)^{k+2-i}(x-h)^{m+r-(l+1-i)},$$

$$1 \le i \le l-1.$$

Since the inequalities $|x-h| \le 2h$ and $|x-3h| \le 2h$ take place for $x \in [h, 3h]$, we have $|S_i| \le \widetilde{C}(k,m,r,l,i)(1/n)^{k+\alpha-l}$, and $\widetilde{C}$ does not depend on n. Note that the number of all such members does not depend on n and depends only on k. Therefore, $\sup_{x\in[h,3h]} |f^{(l)}(x)| \le C_l(1/n)^{k+\alpha-l}$, $0 \le l \le k+1$, and, moreover,

$$\sup_{x\in[0,1/n]} \left|f^{(l)}(x)\right| \le \max\left(\sup_{x\in[0,h]} \left|f^{(l)}(x)\right|, \sup_{x\in[h,3h]} \left|f^{(l)}(x)\right|\right)$$

$$\le \max\left((k+\alpha)\cdots(k+\alpha+1\_l)\left(\frac{1}{4n}\right)^{k+\alpha-l},\ C_l\left(\frac{1}{n}\right)^{k+\alpha-l}\right)$$

$$\le K_l h^{k+\alpha-l},\quad 0\le l\le k.$$

It remains to show that $\sup_{x,y\in[0,1/n]}\big|f_n^{(k)}(x)-f_n^{(k)}(y)\big|/|x-y|^\alpha\le C$, where C is a constant not depending on n. If $x,y\in[0,1/4n]$ then $\big|f_n^{(k)}(x)-f_n^{(k)}(y)\big|/|x-y|^\alpha\le(k+\alpha)\cdots(1+\alpha)$ because $|x^\alpha-y^\alpha|\le|x-y|^\alpha$. If $x,y\in[1/4n,3/4n]$ then

$$\frac{\big|f_n^{(k)}(x)-f_n^{(k)}(y)\big|}{|x-y|^\alpha}\le\sup_{t\in[1/4n,3/4n]}\big|f_n^{(k+1)}(t)\big|\big|x-y\big|^{1-\alpha}$$

$$\le C_{k+1}\left(\frac{1}{n}\right)^{\alpha-1}\left(\frac{1}{n}\right)^{1-\alpha}=C_{k+1}.$$

If $x\in[0,1/4n)$ and $y\in(1/4n,3/4n]$ then

$$\frac{|f_n^{(k)}(x)-f_n^{(k)}(y)|}{|x-y|^\alpha}\le\frac{|f_n^{(k)}(1/4n)-f_n^{(k)}(x)|}{|x-1/4n|^\alpha}+\frac{|f_n^{(k)}(1/4n)-f_n^{(k)}(y)|}{|1/4n-y|^\alpha}$$

$$\le(k+\alpha)\cdots(1+\alpha)+C_{k+1}.$$

The cases of other x and y are treated by analogy. So, we have proven that

$$\|f_n(x)\|_{C^{k+\alpha}}\le K.$$

$$F_n(f_n)=2\varphi(n)n^{1+\alpha}n\int_0^{1/4n}x^{k+\alpha}dx=\varphi(n)\frac{2}{(k+\alpha+1)4^{k+\alpha+1}},$$

$$\|F_n\|^*\ge|F_n(f_n)|\,/\,\|f_n\|\ge\varphi(n)\frac{2}{K(k+\alpha+1)4^{k+\alpha+1}},$$

which means that $\sup_n\|F_n\|^*=\infty$. By the Banach–Steinhaus theorem $\big($[80: p.107]$\big)$ the following two conditions are equivalent:

1) $\sup_n\|F_n\|^*<\infty$,

2) $\sup_n|F_n(f)|<\infty$ for all $f\in C_0^{k+\alpha}[0,1]$.

Since $\sup_n\|F_n\|^*=\infty$, there exists a function $f\in C_0^{k+\alpha}[0,1]$ such that $\sup_n|F_n(f)|=\infty$. Lemma 2.2 is proven.

We now turn directly to the proof of the main assertion. Define a function $f(x)$ on the interval $[0,1]$ by the equality

$$f(x)=\sqrt{2\pi}\,e^{x^2/2}f_0(x),$$

where $f_0(x)$ is the function in Lemma 2.1. Let $f(x)=0$ for $x\in[1,\infty)$ and extend f onto the negative half-line by putting $f(x)=f(-x)$. We have (see the proof of Theorem 1):

$$\mathbf{E}\,f(S_n)=\sum_{i=-\infty}^{\infty}f(i\varepsilon)\eta(i\varepsilon)\big(1+\varepsilon q_1(i\varepsilon)+\cdots+\varepsilon^{k+1}q_{k+1}(i\varepsilon)\big)\varepsilon$$

$$+\sum_{i=-\infty}^{\infty} f(i\varepsilon)(1+\varepsilon^{k+3})\varepsilon\, o(\varepsilon^{k+1}).$$

Let n be such that $\sqrt{n} = \varepsilon^{-1}$ is an integer. Then

$$\left|\mathbf{E}\, f(N) - \sum_{i=-\infty}^{\infty} f(i\varepsilon)\eta(i\varepsilon)\varepsilon\right|$$

$$= 2\left|\int_0^1 f(x)\eta(x)dx - \sum_{i=0}^{\sqrt{n}} f(i\varepsilon)\eta(i\varepsilon)\varepsilon\right|$$

$$= 2\left|\int_0^1 f_0(x)dx - \sum_{i=0}^{\sqrt{n}} f_0(i\varepsilon)\varepsilon\right|.$$

Since $f_0(x)q_l(x)$, $1 \le l \le k+1$, satisfy the conditions of Lemma 2.1, we have

$$\varepsilon^l\left|\int_{-\infty}^{\infty} f(x)dQ_l(x) - \sum_{i=-\infty}^{\infty} f(i\varepsilon)\eta(i\varepsilon)q_l(i\varepsilon)\varepsilon\right|$$

$$\le 2\varepsilon^l\left|\int_0^1 f_0(x)q_l(x)dx - \sum_{i=0}^{\sqrt{n}} f_0(i\varepsilon)q_l(i\varepsilon)\varepsilon\right|$$

$$= O(\varepsilon^{k+l}),$$

$$\varphi(n)\varepsilon^{-(k+\alpha)}\bigg|\mathbf{E}\, f(S_n) - \mathbf{E}\, f(N)$$

$$-\varepsilon\int_{-\infty}^{\infty} f(x)dQ_1(x) - \cdots - \varepsilon^{k+1}\int_{-\infty}^{\infty} f(x)dQ_{k+1}(x)\bigg|$$

$$\ge 2\varphi(n)\varepsilon^{-(k+\alpha)}\left|\int_0^1 f_0(x)dx - \sum_{i=0}^{\sqrt{n}} f_0(i\varepsilon)\varepsilon\right| - C\varphi(n)\varepsilon^{1-\alpha}.$$

It is clear that, without loss of generality, we may assume that $\limsup_{n\to\infty}\varphi(n)\varepsilon^{1-\alpha} < \infty$. The function $f_0(x)$ satisfies the conditions of Lemma 2.1, which implies the assertion of Theorem 2.

3. Proof of Theorem 3

We use the classical truncation technique. Denote

$$\widehat{\xi}_{n,i} = \xi_{n,i} I\big\{|\xi_{n,i}| \le 1\big\} - \mathbf{E}\,\xi_{n,i} I\big\{|\xi_{n,i}| \le 1\big\}, \quad i = 1, \dots, n,$$

$$B_n^2=\sum_{i=1}^n D\widehat{\xi}_{n,i}, \quad \bar{\xi}_{n,i}=\frac{1}{B_n}\widehat{\xi}_{n,i}, \quad \overline{S}_n=\sum_{i=1}^n \bar{\xi}_{n,i}, \quad \overline{L}_t=\sum_{i=1}^n \mathbf{E}\,|\bar{\xi}_{n,i}|^t.$$

It is clear that $\bar{\xi}_{n,i}$, $i=1,\dots,n$, satisfy the conditions of Theorem B, and, in addition, condition 1) for f in Theorem B is fulfilled. Hence, the assertion of Theorem B for the sum $\overline{S}_n$ holds with the following upper bound for the remainder term:

$$|\,\overline{\eta}| \le CH\big(\overline{L}_{k+\alpha}+\overline{L}_{k+p+\alpha}\big).$$

By equality $\mathbf{E}\,\xi_{n,i}=0$, it follows:

$$|\mathbf{E}\,\xi_{n,i}I\big\{|\xi_{n,i}|\le 1\big\}|=|\mathbf{E}\,\xi_{n,i}I\big\{|\xi_{n,i}|>1\big\}|\le \mathbf{E}\,|\xi_{n,i}|^t$$

for all $t\ge 1$ and every $i=1,\dots,n$. It is clear that

$$1-B_n^2=\sum_{i=1}^n\Big(\mathbf{E}\,\xi_{n,i}^2I\big\{|\xi_{n,i}|>1\big\}+\big(\mathbf{E}\,\xi_{n,i}I\big\{|\xi_{n,i}|\le 1\big\}\big)^2\Big)\le 2L_t, \quad t\ge 2.$$

We take $t=k+\alpha$. Without loss of generality we may assume that $L_{k+\alpha}\le 3/8$. Otherwise, the estimate in Theorem 3 becomes trivial (use the well-known Rosenthal inequality). Thus, hereafter we assume that $B_n\ge 1/2$. It is easy to see that

$$ll\mathbf{E}\,|\bar{\xi}_{n,i}|^{k+\alpha}\le C(k)\mathbf{E}\,|\xi_{n,i}|^{k+\alpha}, \mathbf{E}\,|\bar{\xi}_{n,i}|^{k+p+\alpha}\le C(k)\mathbf{E}\,|\xi_{n,i}|^{k+\alpha}, \qquad i=1,\dots,n. \quad (3.1)$$

Hence, we can obtain the estimate $|\,\overline{\eta}|\le C(k)HL_{k+\alpha}$. Fro proving Theorem 3, we need to estimate proximity of $\mathbf{E}\,f(S_n)$ and $\mathbf{E}\,f(\overline{S}_n)$ (Lemma 3.1), as well as proximity of the corresponding cumulants of S_n and $\overline{S}_n$ (Lemma 3.2).

Lemma 3.1. *Under the conditions of Theorem 3, the following estimate holds:*

$$|\mathbf{E}\,f(S_n)-\mathbf{E}\,f(\overline{S}_n)|\le CL_{k+\alpha+\sigma},$$

where $\sigma=\max\{0,p-2\}$ *and the constant* C *depends only on* p, k, α, *and* H.

Proof. We use the standard arguments of Lindeberg's operator method. Denote

$$T_1=\sum_{i=2}^n\xi_{n,i}, \quad T_n=\sum_{i=1}^{n-1}\bar{\xi}_{n,i}, T_j=\sum_{i=1}^{j-1}\bar{\xi}_{n,i}+\sum_{i=j+1}^n\xi_{n,i}, \quad 2\le j\le n-1.$$

Note that T_j does not depend on $\xi_{n,j}$ and $\bar{\xi}_{n,j}$. The well-known Marcinkiewicz–Zygmund inequality provides the estimate $\mathbf{E}\,|T_j|^t\le C(t)L_t$, $j=1,\dots,n$. Moreover, the function f satisfies the Lipschitz condition

$$|f(x)-f(y)|\le C(H)|x-y|\big(1+|x|^{k-3+p+\alpha}+|y|^{k-3+p+\alpha}\big).$$

Using these arguments, we obtain the following estimate:

$$|\mathbf{E}\,f(S_n)-\mathbf{E}\,f(\overline{S}_n)|$$

$$= \left| \sum_{j=1}^{n} \mathbf{E} \Big\{ \int_{-\infty}^{\infty} \big(f(y+\xi_{n,j}) - f(y+\bar{\xi}_{n,j}) \big) d\,\mathbf{P}(T_j < y) \Big\} \right|$$

$$\le C(H) \sum_{j=1}^{n} \mathbf{E} \Big\{ \int_{-\infty}^{\infty} |\xi_{n,j} - \bar{\xi}_{n,j}| \Big(1 + |y|^{k-3+p+\alpha} + |\xi_{n,j}|^{k-3+p+\alpha}$$

$$+ |\bar{\xi}_{n,j}|^{k-3+p+\alpha} \Big) d\,\mathbf{P}(T_j < y) \Big\}.$$

Obviously,

$$\mathbf{E}\,|\xi_{n,i} - \bar{\xi}_{n,i}| \le \mathbf{E}\,|\xi_{n,i} - \widehat{\xi}_{n,i}| + \frac{1-B_n^2}{B_n(1+B_n)} \mathbf{E}\,|\widehat{\xi}_{n,i}|\,,$$

$$\mathbf{E}\,|\xi_{n,j} - \widehat{\xi}_{n,j}| \le 2\mathbf{E}\,|\xi_{n,j} I\big\{|\xi_{n,j}| > 1\big\}|,$$

$$\int_{-\infty}^{\infty} |y|^{k-3+p+\alpha} d\,\mathbf{P}(T_j < y) \le \int_{|y|\le 1} d\,\mathbf{P}(T_j < y) + \int_{|y|>1} |y|^{k+\alpha+\sigma} d\,\mathbf{P}(T_j < y)$$

$$\le 1 + \mathbf{E}\,|T_j|^{k+\alpha+\sigma}$$

for all $1 \le j \le n$. Taking into account the inequalities of type (3.1), we obtain the statement of Lemma 3.1.

Denote by $K_r(Z)$ the rth cumulant of an r.v˙ Z.

Lemma 3.2. *Under the conditions of Theorem 3, the following relation holds:*

$$\left| \sum_{(k-2)} \prod_{j=1}^{r} K_{s_j+2}(S_n) - \sum_{(k-2)} \prod_{j=1}^{r} K_{s_j+2}(\overline{S}_n) \right| \le C L_{k+\alpha}, \tag{3.2}$$

where $\sum_{(k-2)}$ *is defined in Theorem* B*, and the constant* C *depends on* k *and* α *only.*

Proof. Denote $A_l = \mathbf{E}\,\xi_{n,1}^{\,l}$ and $B_l = \mathbf{E}\,\widehat{\xi}_{n,1}^{\,l}$. By definition, we have

$$K_d(\xi_{n,1}) - K_d(\widehat{\xi}_{n,1}) = \sum_{[d]} C_1(d) \Big\{ \prod_{l=1}^{d} A_l^{m_l} - \prod_{l=1}^{d} B_l^{m_l} \Big\}$$

$$= \sum_{[d]} C_1(d) \Big\{ (A_1^{m_1} - B_1^{m_1}) \prod_{i=2}^{d} A_i^{m_i} + B_1^{m_1} (A_2^{m_2} - B_2^{m_2}) \prod_{i=2}^{d} A_i^{m_i}$$

$$+ \cdots + (A_d^{m_d} - B_d^{m_d}) \prod_{i=1}^{d-1} B_i^{m_i} \Big\},$$

where $\sum_{[d]}$ denotes the summation over all integer nonnegative numbers $(m_1, \ldots, m_d)$ such that $m_1 + 2m_2 + \cdots + dm_d = d$, $d \le k$. If $m_l = 0$ then, obviously, $A_l^{m_l} - B_l^{m_l} = 0$. If $m_l \ge 1$ then, by (3.1), we can obtain the inequality

$$|A_l^{m_l} - B_l^{m_l}| \le C_2(l) |A_l - B_l| |A_l|^{m_l - 1} \le C_3(l) \mathbf{E}\,|\xi_1|^{k+\alpha}.$$

Using this inequality and the properties of the cumulants, we obtain

$$|K_d(\xi_{n,1}) - K_d(\bar{\xi}_{n,1})| \le |K_d(\xi_{n,1}) - K_d(\widehat{\xi}_{n,1})| + \left(\frac{1}{B_n^d} - 1\right) K_d(\widehat{\xi}_{n,1})$$

$$\le C_4(d)\mathbf{E}\,|\xi_{n,1}|^{k+\alpha}.$$

The analogous inequality is true for all r.v.'s $\xi_{n,i}$ and $\bar{\xi}_{n,i}$, $i = 1, \ldots, n$. Since the summands are independent, the following relation holds:

$$|K_d(S_n) - K_d(\bar{S}_n)| = \left| \sum_{i=1}^{n} \big(K_d(\xi_{n,i}) - K_d(\bar{\xi}_{n,i})\big) \right| \le C_5(d) L_{k+\alpha}.$$

Therefore, as an upper bound for the left-hand side of inequality (3.2), we can take the sum

$$\sum_{(k-2)} \big\{C_5(s_1+2) L_{k+\alpha} L_{s_1+2} \cdots L_{s_r+2} + \cdots + C_5(s_r+2) L_{s_1+2} \cdots L_{s_r+2} L_{k+\alpha}\big\}$$

$$\le C L_{k+\alpha} \sum_{(k-2)} \prod_{i=1}^{r} L_{s_i+2}$$

We recall that, without loss of generality, we assume $L_{k+\alpha} \le 3/8$. Hence,

$$\prod_{i=1}^{r} L_{s_i+2} \le \prod_{i=1}^{r} L_{k+\alpha}^{\frac{s_i+2}{k+\alpha}} \le C L_{k+\alpha}.$$

Lemma 3.2 and Theorem 3 are proven.

10 The central limit theorem for generalized canonical Von Mises statistics

1. Basic definitions and statement of the main result

In this chapter we study the limit behavior of a wide class of von Mises statistics defined by an arbitrary array of degenerate (canonical) kernel functions. The impetus for our investigation is the paper by P. Major [87] where limit theorems for weighted U-statistics are obtained. The proof of the main result of this chapter is based on a representation of generalized von Mises statistics as multiple stochastic integrals with respect to some empirical measures different from those previously considered in investigations of the limit behavior of classical von Mises statistics.

Let $X_1, X_2, \ldots$ be independent identically distributed random variables taking values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$ with distribution P on $\mathcal{B}$. We consider statistics of the following type:

$$M_n^{(m)} := n^{-\frac{m}{2}} \sum_{1 \le i_1, \ldots, i_m \le n} f_n\left(\frac{i_1}{n}, \ldots, \frac{i_m}{n},\ X_{i_1}, \ldots, X_{i_m}\right), \quad n = 1, 2, \ldots, \tag{1}$$

which are called *generalized von Mises statistics*, with the kernel functions

$$f_n\, [0,1]^m \times \mathcal{X}^m \to \mathbf{R} \tag{2}$$

canonical (or degenerate), i.e.,

$$\mathbf{E} f_n(t_1, \ldots, t_m,\ x_1, \ldots, x_{k-1},\ X_k,\ x_{k+1}, \ldots, x_m) = 0 \tag{3}$$

for all $t_1, \ldots, t_m \in [0,1]$, $x_1, \ldots, x_m \in \mathcal{X}$, and $k = 1, \ldots, m$.

It is clear that we could consider an equivalent form of the statistics $M_n^{(m)}$ based on an array of kernel functions $\{f_{n, i_1, \ldots, i_m}\}$ without any normalization. In this case, each kernel function would depend only on the sample $\{X_i\}$. Possibly, such a form of notation is more natural than that in (1) but, for our purpose, it is more convenient to emphasize dependence on the multi-index and n in the arguments of the functions f_n; otherwise, we need some additional notations.

We would like to say a few words about the history of the objects under study. The theory of U-statistics appeared in the late 40s when W. Hoeffding [59] and R. von Mises [89] began to investigate both U-statistics of the form

$$U_n := \frac{1}{C_n^m} \sum_{1 \le i_1 < \cdots < i_m \le n} g(X_{i_1}, \ldots, X_{i_m}),$$

where g is a symmetric function with respect to all permutations of arguments, and the so-called von Mises statistics

$$M_n := n^{-m} \sum_{1 \le i_1, \ldots, i_m \le n} g(X_{i_1}, \ldots, X_{i_m}).$$

As a matter of fact, the asymptotic behavior, as $n \to \infty$, of statistics M_n and U_n is the same, since they differ from each other only by presence (or absence) of elements whose multi-index contains at least two equal coordinates. The number of elements of such a type in these multiple sums is much less than that of other elements. Essential differences appear only in the study of their moments. Note that we do not need the symmetry property for the kernel functions in von Mises statistics. On the other hand, if the distribution of X_i is continuous then we can interpret U-statistics as a particular case of von Mises statistics with symmetric kernel functions which vanish on all subspaces containing the main diagonal of the product sample space.

Conversely, it is easy to see that every von Mises statistic can be represented as a finite sum of U-statistics of decreasing dimension. Moreover, every U-statistic can be represented as a finite sum of degenerate U-statistics (see [61]). Thus, we only study the canonical case, although a great number of papers is dedicated to the nondegenerate case (for example, see the references in [77]).

In 1962, A. A. Filippova [52] obtained limit theorems for multiple stochastic integrals based on the classical empirical measure on the real line. In other words, she obtained the corresponding limit theorems for degenerate von Mises statistics M_n which admit the representation mentioned above. We would like yet to select two remarkable papers by M. Arcones and E. Giné [3, 4], where probability inequalities, the central limit theorem, and the law of the iterated logarithm are studied for the canonical case including the multivariate (functional) setting (the so-called *uniform* limit theorems for U-processes).

Since the 70s, the generalized U-statistics of the form

$$\widetilde{U}_n := \frac{1}{C_n^m} \sum_{1 \le i_1 < \cdots < i_m \le n} g_{i_1,\ldots,i_m}(X_{i_1}, \ldots, X_{i_m})$$

are studied under some properties of the whole array of symmetric kernel functions $g_{i_1,\ldots,i_m}$.

In particular, in the generality mentioned, V. H. de la Peña [91] obtained moment and probability inequalities for $\widetilde{U}_n$ (see also [92]). In 1994, P. Major [87] proved limit theorems for the following weighted U-statistics:

$$\overline{U}_n := n^{-\frac{m}{2}} \sum_{1 \le i_1 < \cdots < i_m \le n} a(i_1, \ldots, i_m) g(X_{i_1}, \ldots, X_{i_m}),$$

where a and g are symmetric functions, g is canonical, and X_1 has uniform distribution on the interval $[0,1]$ (the case of an arbitrary distribution on $\mathbf{R}$ can be reduced to the above-mentioned case by the standard quantile transformation). In other words, an important particular case of the problem is studied in [87].

Let $\mathcal{A}$ be the σ-algebra of all Borel subsets of the interval $[0,1]$, let $\Lambda(\cdot)$ be the Lebesgue measure on $[0,1]$, and let $\{\mathbf{K}_P(A,B);\ A \in \mathcal{A},\ B \in \mathcal{B}\}$ be an elementary centered Gaussian stochastic measure defined on the semiring of all rectangles $A \times B$ of the Cartesian product $[0,1] \times \mathcal{X}$ and having the covariance

$$\mathbf{E}\,\mathbf{K}_P(A,B)\mathbf{K}_P(A',B') = \Lambda(A \cap A')\big(P(B \cap B') - P(B)P(B')\big).$$

Note that existence of such a kind random process $\{\mathbf{K}_P(A,B);\ A\in\mathcal{A},\ \mathcal{B}\in\mathcal{B}\}$ is provided by the classical Kolmogorov theorem. In particular, from here it follows that if some sets A and A' are disjoint then, for all subsets B and B', the random variables $\mathbf{K}_P(A,B)$ and $\mathbf{K}_P(A',B')$ are independent. Additivity (with probability 1) of the stochastic measure with respect to each argument A or B (it is equivalent to additivity of the measure on the semiring mentioned) is rather easily verified. Say, let B and B' be disjoint Borel subsets of the real line. Then

$$\mathbf{E}\big(\mathbf{K}_P(A,B\cup B')-\mathbf{K}_P(A,B)-\mathbf{K}_P(A,B')\big)^2$$

$$\begin{aligned}=\Lambda(A)\Big\{&P(B\cap B')\big(1-P(B\cap B')\big)-2P(B)\big(1-P(B\cap B')\big)\\&-2P(B')\big(1-P(B\cap B')\big)+P(B)\big(1-P(B)\big)\\&+P(B')\big(1-P(B')\big)-2P(B)P(B')\Big\}=0.\end{aligned}$$

Additivity in the first argument A is verified similarly. If the arguments A and B are changed only within the subclass of intervals $\{[0,s];\ s\in[0,1]\}$, then we deal with the so-called Kiefer process ("Brownian pocket" on the plane).

Let $\{\mathbf{B}_P(A,B);\ A\in\mathcal{A},\ \mathcal{B}\in\mathcal{B}\}$ be a Wiener process on the Cartesian product $[0,1]\times\mathcal{X}$, i.eȧn elementary centered Gaussian stochastic measure on the above-introduced semiring having covariance $\mathbf{E}\,\mathbf{B}_P(A,B)\mathbf{B}_P(A',B')=\Lambda(A\cap A')P(B\cap B')$. If the rectangles $A\times B$ and $A'\times B'$ are disjoint, then the corresponding values of the Wiener process are independent. For $A,B\in\{[0,s];\ s\in[0,1]\}$, the corresponding Gaussian process with two-dimensional time parameter is called a "Wiener sheet."

Hereafter we use the following two L_2-norms $\|\cdot\|_*$ and $\|\cdot\|_n$:

$$\|f\|_*^2:=\sum_{i_1,\dots,i_m\le m}\ \sum_{j_1,\dots,j_m\le m}\mathbf{E}f^2(\omega_{i_1},\dots,\omega_{i_m},\,X_{j_1},\dots,X_{j_m}),\tag{4}$$

$$\|f\|_n^2:=\sum_{i_1,\dots,i_m\le m}\ \sum_{j_1,\dots,j_m\le m}\mathbf{E}f^2\big(\omega_{i_1}^{(n)},\dots,\omega_{i_m}^{(n)},\,X_{j_1},\dots,X_{j_m}\big),\tag{5}$$

where $\omega_1,\dots,\omega_m$ are independent uniformly distributed on $[0,1]$ random variables which do not depend on the sample $\{X_j\}$ on an extended probability space, and $\omega_k^{(n)}=[n\omega_k]/n$ for all k. As a rule, norms of the type (4) are introduced in the theory of von Mises statistics (see [16, 52]). But for asymptotic analysis of distributions of generalized U-statistics, with account taken of the above remark, it is sufficient to define only the moments

$$\mathbf{E}f^2_{i_1,\dots,i_m}(X_1,\dots,X_m).$$

Introduction of the second norm (5) is explained by necessity of approximation to the initial kernel f_n of statistic $M_n^{(m)}$ by a new kernel independent of n. In the generality considered, we cannot manage only with the norm $\|\cdot\|_*$.

Introduce also the normalized counting measure on $[0,1]$ with atoms at the points $\{i/n;\ i=1,\dots,n\}$ by the formula

$$\mu_n(A) := \frac{1}{n}\#\{i:\ i/n\in A\},\quad A\in\mathcal{A}.$$

It is clear that $\mu_n(A)$ is the distribution of the random variable $\omega_1^{(n)}$, and, as $n\to\infty$, the sequence $\{\mu_n\}$ converges weakly to the Lebesgue measure Λ.

Now we formulate the following condition for closeness of the kernel f_n to a function f:

Condition $(*)$. For every $\varepsilon>0$, there exist step functions $f_{N,M}$ on $[0,1]^m\times\mathcal{X}^m$, equal to $f_{i_1,\dots,i_m,j_1,\dots,j_m}$ on the partition element $T_{i_1}\times\dots\times T_{i_m}\times B_{j_1}\times\dots\times B_{j_m}$, $i_k\le M$, $j_k\le N$, $k=1,\dots m$, such that $\|f-f_{N,M}\|_*\le\varepsilon$ and $\|f_n-f_{N,M}\|_n\le\varepsilon$ for all $n\ge n_0$, where $T_i=\big((i-1)/M,\, i/M\big]$, $i\le M$, the naturals n_0, N, M depend only on ε and f, and the measurable subsets B_j, $j\le N$, form, generally speaking, an arbitrary partition of the sample space $\mathcal{X}$.

R e m a r k 1. Condition $(*)$ means that the functions under consideration can be approximated in the norms $\|\cdot\|_*$ and $\|\cdot\|_n$ by the same step functions which are constant on parallelepipeds of a special type. For instance, in the case $\mathcal{X}=[\prime,\infty]$ and $P=\Lambda$, this condition holds whenever the kernel functions are Riemann square integrable on the cube $[0,1]^{2m}$ and on all its intersections with the linear subspaces containing the main diagonal (of course, under the condition regarding closeness of the functions in the above-introduced norms). Here we can consider intervals of the type $(a,b]$ as the partition elements B_j. In the case $f_n(\bar t,\overline{X})=s_n(\bar t)g(\overline{X})$ (under the corresponding notations of the vector@-valued arguments) we can require Riemann integrability only of the component $s_n(\bar t)$ and its L_2-limit. Moreover, if $s_n(\bar t)=a\big([nt_1],\dots,[nt_m]\big)$ (for example, see [87]), then we do not need norms like $\|\cdot\|_n$ in Condition $(*)$. It is sufficient to postulate there only convergence of the sequence f_n to f $\big($or $s_n(\bar t)$ to the corresponding limit$\big)$ in the norm $\|\cdot\|_*$.

Theorem. *Let $M_n^{(m)}$ be defined by (1)–(3) and let Condition $(*)$ be fulfilled. Then, as $n\to\infty$,*

$$M_n^{(m)}\Rightarrow\int f(t_1,\dots,t_m,x_1,\dots,x_m)\mathbf{K}_P(dt_1,dx_1)\cdots\mathbf{K}_P(dt_m,dx_m)$$

$$\stackrel{d}{=}\int f(t_1,\dots,t_m,x_1,\dots,x_m)\mathbf{B}_P(dt_1,dx_1)\cdots\mathbf{B}_P(dt_m,dx_m),\tag{6}$$

where the $2m$-fold integrals with integrability domain $[0,1]^m\times\mathcal{X}^m$ are understood to be stochastic (L_2-limits of the corresponding integral sums), the symbol $\Rightarrow$ denotes weak convergence of distributions of random variables, and the symbol $\stackrel{d}{=}$ denotes equality of distributions.

R e m a r k 2. The theorem generalizes the main result in [87] for

$$f_n(t_1,\dots,t_m,x_1,\dots,x_m)=a\big([nt_1],\dots,[nt_m]\big)g(x_1,\dots,x_m)$$

under the condition that $a\big([nt_1],\ldots,[nt_m]\big) \to A(t_1,\ldots,t_m)$ in $L_2[0,1]^m$, where $A(t_1,\ldots,t_m)$ is an arbitrary continuous function, and X_i are uniformly distributed on $[0,1]$. Here the continuity condition for $A(t_1,\ldots,t_m)$ can be replaced by the weaker condition of its Riemann integrability (see Remark 1).

Formulate another useful consequence of the above theorem.

Corollary. *Let the canonical kernels f_n in (1), continuous on $[0,1]^{2m}$, converge uniformly to a function f. Then (6) holds for $P = \Lambda$.*

2. Auxiliary propositions

Introduce the following atomic stochastic measure (point process) defined on the semiring of all canonical rectangles in $[0,1]\times\mathcal{X}$:

$$\mathbf{S}_n(A,B) := \frac{1}{\sqrt{n}}\sum_{i:\, i/n\in A}\big(\mathbf{I}(X_i\in B)-P(B)\big),$$

where $A\in\mathcal{A}$, $B\in\mathcal{B}$, $\mathbf{I}(\cdot)$ is the indicator function of an event. The atoms of this measure have the form $(i/n,\ X_i)$, $i=1,\ldots,n$. By the Fubini theorem it is easy to obtain the following representation of von Mises statistics with canonical kernels which is a key for understanding the specific character of their limit behavior:

Proposition 1. *The statistic $M_n^{(m)}$ admits representation as the following $2m$-fold stochastic integral on $[0,1]^m\times\mathcal{X}^m$:*

$$M_n^{(m)} = \int f_n(t_1,\ldots,t_m,x_1,\ldots,x_m)\mathbf{S}_n(dt_1,dx_1)\cdots\mathbf{S}_n(dt_m,dx_m).$$

Now denote by $\mathcal{A}_0$ the subclass of subsets in $\mathcal{A}$ satisfying the following conditions: $\mu_n(A)\to\Lambda(A)$, $\mu_n(B)\to\Lambda(B)$ and, moreover, as $n\to\infty$, $\mu_n(A\cap B)\to\Lambda(A\cap B)$ for all $A,B\in\mathcal{A}_0$. Note that, if the subsets A and B are Jordan measurable, then these conditions hold. In general, fulfillment of the first and the second conditions does not imply that of the third. As an example, we can consider the Borel subsets $A=[0,1/2]$ and $B=R[0,1/2]\cup \mathrm{Ir}[1/2,1]$, where $R[\cdot]$ and $\mathrm{Ir}[\cdot]$ denote the subsets of all rational and irrational numbers of the intervals indicated. Here the values $\mu_n(A)$, $\mu_n(B)$, and $\mu_n(A\cap B)$ tend to $1/2$ (the Lebesgue measure of the sets A and B) as $n\to\infty$. However, $\Lambda(A\cap B)=0$.

Consider the random processes

$$\big\{\mathbf{S}_n(A,B);\ A\in\mathcal{A}_0,\ B\in\mathcal{B}\big\}\ \text{ and }\ \big\{\mathbf{K}_{\mathcal{P}}(\mathcal{A},\mathcal{B});\ \mathcal{A}\in\mathcal{A}_{\prime},\ \mathcal{B}\in\mathcal{B}\big\}.$$

Proposition 2. *As $n\to\infty$,*

$$\mathbf{S}_n(\cdot\,,\cdot)\Rightarrow\mathbf{K}_P(\cdot\,,\cdot),$$

where the symbol $\Rightarrow$ means weak convergence of finite-dimensional distributions of random processes;

$$\mathbf{E}\big(\mathbf{S}_n^{r_1}(A_1,B_1)\cdots\mathbf{S}_n^{r_d}(A_d,B_d)\big)\to\mathbf{E}\big(\mathbf{K}_P^{r_1}(A_1,B_1)\cdots\mathbf{K}_P^{r_d}(A_d,B_d)\big)$$

for all natural r_j and all $A_j \in \mathcal{A}_0$, $B_j \in \mathcal{B}$, $j = 1, \dots, d$.

Proof. The first statement follows from the multivariate central limit theorem. We must only emphasize that, for all $A_1, A_2 \in \mathcal{A}_0$ and $B_1, B_2 \in \mathcal{B}$, we have

$$\mathbf{E}\,\mathbf{S}_n(A_1, B_1)\mathbf{S}_n(A_2, B_2)$$

$$= \frac{1}{n}\mathbf{E}\Bigg\{ \sum_{i:\, i/n \in A_1} \big(\mathbf{I}(X_i \in B_1) - \Lambda(B_1)\big) \sum_{j:\, j/n \in A_2} \big(\mathbf{I}(X_j \in B_2) - \Lambda(B_2)\big) \Bigg\}$$

$$= \frac{1}{n}\mathbf{E}\Bigg(\sum_{i:\, i/n \in A_1 \cap A_2} \big(\mathbf{I}(X_i \in B_1) - \Lambda(B_1)\big) \Bigg)^2$$

$$= \mu_n(A_1 \cap A_2)\big(\Lambda(B_1 \cap B_2) - \Lambda(B_1)\Lambda(B_2)\big)$$

$$\to \Lambda(A_1 \cap A_2)\big(\Lambda(B_1 \cap B_2) - \Lambda(B_1)\Lambda(B_2)\big).$$

To prove the second statement we note that, because of additivity in each of the two arguments of the stochastic measures under consideration and independence of their values for the pairwise disjoint subsets A_j, the problem can be reduced to the case in which $A_1 = \dots = A_d$ and the subsets B_j are pairwise disjoint.

It is well known that, generally speaking, weak convergence does not imply moment convergence. But this assertion holds if the prelimit random variables are uniformly integrable. The last requirement is fulfilled if we prove that every absolute moment of the random variable $\mathbf{S}_n(A, B)$ is uniformly bounded in n. It immediately implies uniform boundedness of all mixed moments of the type $\mathbf{E}\left(\mathbf{S}_n^{r_1}(A, B_1) \cdots \mathbf{S}_n^{r_d}(A, B_d)\right)$, i.e; it provides the uniform integrability mentioned above. In order to justify this, we apply the so-called poissonization, i.e. replacement of the empirical measure $\mathbf{S}_n(\cdot\,,\cdot)$ by the corresponding Poisson point process, say, as in [16]. In this case, if the value $\max_j \Lambda(B_j)$ is sufficiently small, then, without loss of generality, we may assume that, for all $N \geq N_0(d)$ (see [16]),

$$\big|\mathbf{S}_n^{r_1}(A, B_1) \cdots \mathbf{S}_n^{r_d}(A, B_d)\big| \leq C\,\mathbf{E}\big|\mathbf{Q}_\lambda^{r_1}(B_1) \cdots \mathbf{Q}_\lambda^{r_d}(B_d)\big|$$

$$= C \prod_{j \leq d} \mathbf{E}\big|\mathbf{Q}_\lambda^{r_j}(B_j)\big| \leq C\Bigg(\prod_{j \leq d} \mathbf{E}\,\mathbf{Q}_\lambda^{2r_j}(B_j) \Bigg)^{1/2},$$

where the constant C depends only on $N_0(d)$ and $\max_{j \leq N} \Lambda(B_j)$; $\mathbf{Q}_\lambda(\cdot) = \frac{1}{\sqrt{n}}\big(\widetilde{\mathbf{Q}}_\lambda(\cdot) - \lambda(\cdot)\big)$, with $\lambda(\cdot) = \Lambda(\cdot)\mu_n(A)n$; by $\widetilde{\mathbf{Q}}_\lambda(\cdot)$ we denote the Poisson point process with mean measure $\lambda(\cdot)$, i.e; $\widetilde{\mathbf{Q}}_\lambda(B_1), \dots, \widetilde{\mathbf{Q}}_\lambda(B_d)$ are independent random variables for pairwise disjoint subsets $\{B_j\}$.

Finally, using the estimate $\mathbf{E}\big(\mathbf{Q}_\lambda(B_j)\big)^s \leq \mu_n(A)^{s/2} s!\Lambda(B_j)$ proven in [16] for an arbitrary natural s we obtain the required result. The proposition is proven.

For the centered Gaussian random processes indicated below, we can easily verify coincidence of the covariances which implies the following proposition well-known in the case $A, B \in \big\{[0, s];\ s \in [0, 1]\big\}$:

Proposition 3. *For all $A \in \mathcal{A}$ and $B \in \mathcal{B}$,*

$$\mathbf{K}_P(A,B) \overset{d}{=} \mathbf{B}_P(A,B) - P(B)\mathbf{B}_P(A,\mathcal{X}).$$

3. Proof of the Theorem

Consider the two-dimensional array of step functions satisfying Condition $(*)$:

$$\Bigg\{ f_{N,M}(t_1,\dots,t_m,x_1,\dots,x_m) := \sum_{i_1,\dots,i_m=1}^{M} \sum_{j_1,\dots,j_m=1}^{N} f_{i_1,\dots,i_m j_1,\dots,j_m}$$

$$\times \prod_{k=1}^{m} \mathbf{I}(t_k \in T_{i_k})\mathbf{I}(x_k \in B_{j_k});\ N,M \geq 1 \Bigg\},$$

where, as before, $T_i = \big((i-1)/M,\, i/M\big]$, $i \leq M$, and the measurable subsets $B_1,\dots,B_N$ form, generally speaking, an arbitrary partition of the sample space $\mathcal{X}$. Introduce the following notations:

$$J_{M,N}^{(n)} := \sum_{i_1,\dots,i_m=1}^{M} \sum_{j_1,\dots,j_m=1}^{N} f_{i_1,\dots,i_m,j_1,\dots,j_m} \mathbf{S}_n(T_{i_1},B_{j_1}) \cdots \mathbf{S}_n(T_{i_m},B_{j_m}), \tag{7}$$

$$J_{M,N} := \sum_{i_1,\dots,i_m=1}^{M} \sum_{j_1,\dots,j_m=1}^{N} f_{i_1,\dots,i_m,j_1,\dots,j_m} \mathbf{K}_P(T_{i_1},B_{j_1}) \cdots \mathbf{K}_P(T_{i_m},B_{j_m}). \tag{8}$$

As a consequence of Proposition 2, we obtain $J_{M,N}^{(n)} \Rightarrow J_{M,N}$ as $n \to \infty$. In what follows, as $N,M \to \infty$, we must prove, first, uniform in n convergence to zero of the second moment of the difference $J_{M,N}^{(n)} - M_n^{(m)}$, and, second, existence of an L_2-limit of the integral sums introduced in (8).

Evaluate the moment $\delta := \mathbf{E}\big(J_{M,N}^{(n)} - M_n^{(m)}\big)^2$. First, note that, without loss of generality, we may assume the step function $f_{N,M}$ to be canonical (i.e; to satisfy (3)). Otherwise, we can reduce it to the canonical form by a special centering as in [61]. For example, in the two-dimensional case, we can transform a function $\varphi(x,y)$ to the canonical by the transformation

$$\widetilde{\varphi}(x,y) = \varphi(x,y) - \mathbf{E}\,\varphi(X_1,y) - \mathbf{E}\,\varphi(x,X_1) + \mathbf{E}\,\varphi(X_1,X_2).$$

It is easy to see that this mapping transforms the step function $f_{N,M}$ to a new step function and accuracy of this replacement in the norm $\|\cdot\|_*$ has the same order ε (up to a factor depending only on m). Moreover, in this case, $J^{(n)}_{M,N}$ in (7) can be represented by (1), where the function f_n must be replaced by $f_{N,M}$. Then

$$\delta = \frac{1}{n^m}\sum_{\bar{i}}\sum_{\bar{j}}\mathbf{E}\left(f_n\left(\frac{\bar{i}}{n},\overline{X}_{\bar{i}}\right)-f_{M,N}\left(\frac{\bar{i}}{n},\overline{X}_{\bar{i}}\right)\right)$$

$$\times\left(f_n\left(\frac{\bar{j}}{n},\overline{X}_{\bar{j}}\right)-f_{M,N}\left(\frac{\bar{j}}{n},\overline{X}_{\bar{j}}\right)\right), \tag{9}$$

where the natural abbreviation is used for the vector@-valued sample and the multi-index.

Further, because of degeneracy of the kernels f_n and $f_{M,N}$, the moments in (9) are different from zero only in the case when each of the coordinates of the $2m$-variate vector $(\bar{i},\,\bar{j})$ coincides at least with another coordinate, i.e; the multiplicity of each coordinate is greater than 1. Hence the double sum in (9) has at most n^m nonzero summands. Using the Cauchy–Bunyakovskiĭ inequality twice (for the second time, for the n^m-points uniform distribution) we finally obtain

$$\delta \le \frac{1}{n^m}\sum_{\bar{i}}\mathbf{E}\left(f_n\left(\frac{\bar{i}}{n},\overline{X}_{\bar{i}}\right)-f_{M,N}\left(\frac{\bar{i}}{n},\overline{X}_{\bar{i}}\right)\right)^2. \tag{10}$$

Now we split the multiple sum in (10) into sums in which coincidence is fixed for certain groups of coordinates of the multi-index. For example, in the three@-dimensional case, the multiple sum over i_1, i_2, i_3, where each index i_k varies from 1 to n, splits into the following five sums: a triple sum over $i_1 \ne i_2 \ne i_3$, three double sums over $i_1 \ne i_2 = i_3$ and all permutations of the indices, and a sum over the main diagonal $i_1 = i_2 = i_3$. Moreover, in the m-dimensional case, the number of these sums can be estimated from above by m^m. Each of these sums can be easily evaluated by the corresponding moment on the right-hand side of (5) $\big($the definition of $\|\cdot\|_n\big)$, which was to be proven.

Now we establish existence of the L_2-limit for the sequence $J_{M,N}$ as $N, M \to \infty$. To verify the Cauchy criteria it is sufficient to prove the relation

$$\lim_{N,M\to\infty}\mathbf{E}\left(J_{M,N}\right)^2 = 0$$

as $N, M \to \infty$ whenever $\|f_{N,M}\|_* \to 0$. First, we evaluate the moment $\mathbf{E}\big(J^{(n)}_{M,N}\big)^2$, which is of interest in its own right. In order to do it, we use the above-indicated poissonization inequality that is true for all even moments of the random variable $J^{(n)}_{M,N}$ (for details, see [3", Section 2]):

$$\mathbf{E}\big(J^{(n)}_{M,N}\big)^2 \le \widetilde{C}\sum_{i_1,\dots,i_m}\sum_{j_1,\dots,j_m}\sum_{i_1,\dots,i'_m}\sum_{j_1,\dots,j'_m}\left|f_{i_1,\dots,i_m,j_1,\dots,j_m}f_{i'_1,\dots,i'_m,j'_1,\dots,j'_m}\right|$$

$$\times\mathbf{E}\big\{\mathbf{Q}_{\lambda_{i_1}}(B_{j_1})\cdots\mathbf{Q}_{\lambda_{i_m}}(B_{j_m})\mathbf{Q}_{\lambda_{i'_1}}(B_{j'_1})\cdots\mathbf{Q}_{\lambda_{i'_m}}(B_{j'_m})\big\}, \tag{11}$$

where $N \geq N_0$, the constant $\widetilde{C}$ depends only on m, N_0, and $\max_{j \leq N} P(B_j)$; $\mathbf{Q}_{\lambda_p}(\cdot) = \frac{1}{\sqrt{n}}\big(\widetilde{\mathbf{Q}}_{\lambda_p}(\cdot) - \lambda_p(\cdot)\big)$, with $\lambda_p(\cdot) = P(\cdot)\mu_n(T_p)n$; here we again denote by $\widetilde{\mathbf{Q}}_{\lambda_p}(\cdot)$ the Poisson point process introduced in the proof of Proposition 2 and having mean measure $\lambda_p(\cdot)$; thus, $\widetilde{\mathbf{Q}}_{\lambda_p}(B_1), \ldots, \widetilde{\mathbf{Q}}_{\lambda_p}(B_k)$ are independent random variables for pairwise disjoint subsets $\{B_j\}$. Moreover, we suppose that the stochastic signed measures $\mathbf{Q}_{\lambda_{p_1}}(\cdot), \ldots, \mathbf{Q}_{\lambda_{p_s}}(\cdot)$ are independent for pairwise different $p_1, \ldots, p_s$. From [3] and relation (12) below, we conclude that all mixed moments on the right-hand side of (11) are nonnegative.

Further, because of the above arguments, each mixed moment in (11) admits the following representation:

$$\mathbf{E}\big\{\mathbf{Q}_{\lambda_{i_1}}(B_{j_1}) \cdots \mathbf{Q}_{\lambda_{i_m}}(B_{j_m})\mathbf{Q}_{\lambda_{i'_1}}(B_{j'_1}) \cdots \mathbf{Q}_{\lambda_{i'_m}}(B_{j'_m})\big\}$$

$$= \prod_l \prod_d \mathbf{E}\big(\mathbf{Q}_{\lambda_l}(B_d)\big)^{r(l,d)}, \tag{12}$$

where the products in (12) are taken over $l \in \{i_1, \ldots, i_m, i'_1, \ldots, i'_m\}$ and $d \in \{j_1, \ldots, j_m, j'_1, \ldots, j'_m\}$ (without account taken of the multiplicity of the indices $\{i_k, i'_k\}$ and $\{j_k, j'_k\}$); $r(l,d)$ is the number of pairs (i,j) in the Cartesian product $\{i_1, \ldots, i_m, i'_1, \ldots, i'_m\} \times \{j_1, \ldots, j_m, j'_1, \ldots, j'_m\}$ satisfying the condition $i = l$ and $j = d$. It is important to emphasize that, by the definition of $r(l,d)$ and because of centering the Poisson point process $\mathbf{Q}_\lambda(\cdot)$, the following two-sided inequality holds: $2 \leq r(l,d) \leq (2m)^2$. Thus, using the above-mentioned estimate $\mathbf{E}\big(\mathbf{Q}_{\lambda_p}(B)\big)^s \leq \mu_n(T_l)^{s/2} s! P(B)$ (see [16]), we obtain the inequality

$$\mathbf{E}\big(\mathbf{Q}_{\lambda_l}(B_d)\big)^{r(l,d)} \leq C_0(f,m)\mu_n(T_l)P(B_d).$$

Substituting the last estimate in (12), we finally obtain from (11) the upper bound

$$\mathbf{E}\big(J^{(n)}_{M,N}\big)^2 \leq C_1 \|f_{N,M}\|^2_n, \tag{13}$$

where C_1 is a constant independent of n. Passing to the limit in (13) as $n \to \infty$, we obtain by Proposition 2 the required final upper bound:

$$\mathbf{E}\big(J_{M,N}\big)^2 = \lim_{n\to\infty} \mathbf{E}\big(J^{(n)}_{M,N}\big)^2 \leq C_1 \|f_{N,M}\|^2_*. \tag{14}$$

Note that convergence of the norm on the right-hand side of (13) to the corresponding limit on the right-hand side of (14) follows immediately from the structure of the step function $f_{N,M}$. In other words, we have proven that $\{J_{M,N};\ M,N = 1,2,\ldots\}$ is a Cauchy sequence and, hence, it has a limit as $M,N \to \infty$ in the Hilbert space of random variables with finite second moments. Therefore, in this space, there exists a limit point J which actually represents the first multiple stochastic integral in (6):

$$J := \int f(t_1, \ldots, t_m, x_1, \ldots, x_m)\mathbf{K}(dt_1, dx_1) \cdots \mathbf{K}_P(dt_m, dx_m).$$

R e m a r k 3. To prove inequality (14) we actually do not need results similar to inequality (13). We can directly estimate the mixed moments

$$\mathbf{E}\,\mathbf{K}_P^{r_1}(A,B_1)\cdots\mathbf{K}_P^{r_d}(A,B_d)$$

which appear in calculating the second moment of the random variable $J_{M,N}$ (see the proof of Proposition 2). In order to use the corresponding decoupling arguments for the mixed moments, we can use Proposition 3 instead of poissonization. This allows us to reduce the problem to analyzing the multilinear forms of moments of the type $\mathbf{E}\,\mathbf{B}_P^r(A,B) = (r-1)!!P(A)^{r/2}$, where $r \geq 2$. Whence, upper bound (14) will be immediately obtained.

R e m a r k 4. Construction of integrals of nonrandom functions with respect to stochastic measures with orthogonal (uncorrelated) values on disjoint subsets was independently proposed in 1940 by A. N. Kolmogorov [72–74] and H. Cramér [34]. This construction is based on the Hilbert-space technique (for example, see [54]). More particular schemes of constructing stochastic integrals with respect to increments of a Wiener process is contained in the classical paper by N. Wiener. Multiple stochastic integrals on orthogonal Gaussian random measures (the Itô–Wiener-type integrals) were studied in [67, 85]. The Hilbert-space technique of the above-cited papers can be transferred to constructing stochastic integrals (including multiple integrals) of nonrandom functions with respect to stochastic measures which are not necessarily orthogonal. Such an integral can be correctly defined if the second moment exists of the integrand with respect to the marginal projection of the total variation of the symmetric measure defined by the relation $m(A,B) := \mathbf{E}\mu(A)\mu(B)$. For example, for a Brownian bridge $W^0(\cdot)$ on the interval $[0,1]$, we have

$$\big|m(A,B)\big|^* = \Lambda(A\cap B) + \Lambda(A)\Lambda(B),$$

where $\big|m(A,B)\big|^*$ is the value of the total variation of $m(\cdot\,,\cdot)$ on the Cartesian product $A\times B$. The projection of this measure is calculated by the formula $\big|m\big(A,[0,1]\big)\big|^* = 2\Lambda(A)$. The last relation provides a correct definition of the stochastic integrals with respect to $W^0(\cdot)$ if the second moment $\big($in the Lebesgue measure on $[0,1]\big)$ of the kernel exists. Of course, we can use the well-known representation of the Brownian bridge $W^0(\cdot)$ via a Wiener process (for example, this follows from Proposition 3) and reduce constructing the corresponding stochastic integral to the Wiener construction (cf. [91]). A similar remark can be made regarding the construction of more general multiple stochastic integrals than those of [67, 85].

So, we have proven the following limit transitions:

$$\begin{array}{ccc} J_{M,N}^{(n)} & \underset{n\to\infty}{\Longrightarrow} & J_{M,N} \\ \downarrow & & \downarrow \\ M_n^{(m)} & & J \end{array}$$

where the vertical arrows denote L_2-convergence of the corresponding integral sums; the left vertical arrow denotes uniform convergence in n. Whence, the weak convergence $M_n^{(m)} \Rightarrow J$ follows, as $n\to\infty$.

By Proposition 3 and condition (3) on the kernel functions, we can easily obtain another representation of the random variable J. To this end, we can consider the integral sums $J_{M,N}$ and, next, pass to the corresponding limit:

$$J = \int f(t_1, \ldots, t_m, x_1, \ldots, x_m)\mathbf{K}(dt_1, dx_1) \cdots \mathbf{K}_P(dt_m, dx_m)$$

$$\stackrel{d}{=} \int f(t_1, \ldots, t_m, x_1, \ldots, x_m)\mathbf{B}_P(dt_1, dx_1) \cdots \mathbf{B}_P(dt_m, dx_m).$$

Therefore, the theorem is proven.

11 Limit theorems for canonical Von Mises statistics based on dependent observations

1. Statement of the main results

In the chapter we study limit behavior of canonical Von Mises statistics based on samples from a sequence of weakly dependent stationary observations satisfying ψ-mixing condition. The approach is based on representation of such statistics as multiple stochastic integrals with respect to the corresponding normalized empirical product-measure as well as on the results in [23]. These results allow us to interpret the corresponding limit random element as a multiple stochastic integral of the kernel of the statistic under consideration with respect to a Gaussian noise which plays a role of the weak limit for the above-mentioned normalized empirical measures.

Let $\{X_i;\, i \in \mathbf{Z}\}$ be a stationary sequence of random variables taking values in an arbitrary measurable space $\{X, \mathcal{A}\}$ and having a marginal distribution P. Consider a measurable function $f(t_1, \ldots, t_d) : X^d \to \mathbf{R}$. Define Von Mises statistics (or V-statistics) by the formula

$$V_n := n^{-d/2} \sum_{1 \le i_1, \ldots, i_d \le n} f(X_{i_1}, \ldots, X_{i_d}), \quad n = 1, 2, \ldots, \tag{1}$$

where $d \geq 2$ and the subscripts i_k independently take all the integers from 1 to n, and the function f satisfies the degeneracy condition

$$\mathbf{E} f(t_1, \ldots, t_{k-1}, X_k, t_{k+1}, \ldots, t_d) = 0 \tag{2}$$

for all $t_1, \ldots, t_d \in X$ and $k = 1, \ldots, d$. The function f is called a *kernel* of a Von Mises statistic, and the statistics with degenerate kernels are called *canonical*.

In the case of independent observations $\{X_i\}$ such statistics were studied in the second half of the last century. (see in [77] the references and examples of such statistics). For the first time some limit theorems for these statistics were obtained by Von Mises [89] and Hoeffding [59], and , moreover, in these papers, there were introduced the so-called U-statistics:

$$U_n := n^{-d/2} \sum_{1 \le i_1 \ne \cdots \ne i_d \le n} f(X_{i_1}, \ldots, X_{i_d}) \tag{3}$$

or

$$U_n^0 := n^{-d/2} \sum_{1 \le i_1 < \cdots < i_d \le n} f_0(X_{i_1}, \ldots, X_{i_d}), \tag{4}$$

where, as a rule, the kernel f_0 in (4) is symmetric with respect to all permutations of the arguments. Notice that to obtain the same limit behavior as that for V-statistics the factor $n^{-d/2}$ in (4) is replaced by $\left(C_n^d\right)^{-1/2}$.

The main distinction of U-statistics from V-statistics is absence of the so-called *diagonal subspaces* in the region of summation in multiple sums in (3) and (4), i.e., absence of subscripts of multiplicities greater than 1 in the definitions (3) and (4).

Under unrestricted conditions on the sample distribution and on the kernels this distinction is not essential since the number of various vector subscripts $(i_1, \dots, i_d)$ with the coordinates having multiplicity greater than 1 in the sums above has the order $O(n^l)$, $l < d$, where l is the number of free coordinates i_k (in other words, the *dimension* of the corresponding diagonal subspace). Therefore, under supplementary moment restrictions on the kernel f, we can easily prove equivalence in probability of representations (1) and (3) . Moreover, it is easy to see that the representations (3) and (4) of U-statistics, are equivalent: If we set in (4)

$$f_0(t_1, \dots, t_d) := \sum f(t_{i_1}, \dots, t_{i_d}),$$

where summation is taken over all permutations $i_1, \dots, i_d$ of the numbers $1, \dots, d$, then we reduce representation (3) to (4).

If the sample distribution P has no atoms then the corresponding U-statistic in (4) coincides in distribution up to the factor $(d!)^{-1}$ with the corresponding Von Mises statistic with the symmetric kernel vanishing on all the diagonal subspaces of X^d. This comment is the central in the limit theory of U-statistics. Notice also that each U-statistic can be represented as a finite linear combination of canonical U-statistics of all dimensions from 1 to d. It is the so-called *Hoeffding decomposition* (for detail, see [77, 61]). This fact allows us to reduce an asymptotic analysis of arbitrary U- and V-statistics to that for canonical ones. Essential preference of canonical V-statistics over U-statistics is the integral representation below.

The stochastic process

$$S_n(B) = n^{-\frac{1}{2}} \sum_{i=1}^{n} (I(X_i \in B) - P(B)), \quad B \in \mathcal{A},$$

is called *normalized empirical measure* (a signed random measure) based on the observations $X_1, \dots, X_n$. It is well known (for example, see [77, 21]) that the statistic V_n admits a representation as the d-fold stochastic integral which is path-wise determined as the classical Lebesgue integral with respect to a finite signed measure (since the stochastic part of $S_n(\cdot)$ is a pure atomic measure):

$$V_n = \int\limits_{X^d} f(x_1, \dots, x_d) S_n(dx_1) \dots S_n(dx_d). \tag{5}$$

It is known (see [59, 21, 52]) that, in the IID case, under the condition

$$\sum_{1 \le j_1, \dots, j_d \le d} \mathbf{E} f^2(X_{j_1}, \dots, X_{j_d}) < \infty,$$

the weak limit of the sequence V_n can be interpreted as a multiple stochastic integral which, under some additional restrictions, (say, if the distribution P has a bounded density) with the corresponding Itô — Wiener multiple stochastic integral (see [77, 67, 47]):

$$V_n \xrightarrow{d} \int\limits_{X^d} f(t_1, \dots, t_d) W_P(dt_1) \dots W_P(dt_d) \tag{6}$$

as $n \to \infty$, where, hereinafter, the symbol "$\xrightarrow{d}$" denotes weak convergence of the corresponding distributions, and $W_P(A)$ is a "White noise" with the structure function P, i. e., it is an elementary stochastic orthogonal Gaussian measure on $\mathcal{A}$ with mean zero and the covariance $\mathbf{E}W_P(A)W_P(B) = P(A \cap B)$. Notice that the integral in (6) cannot be path-wise defined almost surely without some supplementary restrictions on the kernel since the White noise has an unbounded total variation almost surely for any nonatomic sample distributions P, say, in $\mathbf{R}^k$.

Our goal is to prove limit theorems in a similar form as that in (6) in the case when the observations are weakly dependent. We know only few results in this direction. First of all, we note the paper [48] in which, in the case $d = 2$, the following representation of the limit random variable for statistics (3) was obtained:

$$\sum_{k=1}^{\infty} \lambda_k \big(\tau_k^2 - 1\big), \tag{7}$$

where $\{\lambda_k\}$ are eigenvalues of the integral operator with the kernel f which are additionally assumed to be summable, and $\{\tau_k\}$ is a Gaussian sequence with the covariance function depending on these eigenvalues as well as on the covariance function of the initial stationary sequence $\{X_i\}$ satisfying φ-mixing condition.

As a consequence of the main result of the paper we prove that the random variable in (7) may be interpret as a bivariate multiple stochastic integral with respect to a Gaussian process with nonorthogonal increments. Such integrals were introduced in [23].

Notice that, in the IID case, the random variables $\{\tau_k\}$ are independent as well, and, in this case, the representation (7) was obtained in [89]. Latter it was extended on the statistics of an arbitrary dimension (see [109]), and, moreover, there was proved another interpretation of the limit random variables as multiple stochastic integrals of type (6) (see) [47, 52, 67]. It seems that the second interpretation of the limit law is more constructive than the first one since, as a rule, we cannot explicitly study the sets of eigenvalues and eigenfunctions of the above-mentioned integral operator for the kernel from a sufficiently wide class.

In the special case when the observations are defined by a nonrandom transform of a Gaussian stationary sequence, under another dependency restriction, limit behavior of canonical U-statistics was investigated in [43]. So, in this paper, other phenomena are studied and the limit random variables are described as nonrandom transforms of the classical multiple Itô–Wiener stochastic integrals.

Finally, we note that a normalized inner product squared of sums of weakly dependent random variables taking values in a Hilbert space admits representation (1) in the case $d = 2$. This particular case has been studied in detail (for example, see [116]). In this case relation (6) follows from the corresponding central limit theorem.

We now introduce the main restrictions on parameters of the problem under consideration. Denote by $\mathcal{F}_j^k$ the *sigma*-field of the events generated by the random variables

$X_j, \dots, X_k,\ j \le k$. For every $m \ge 1$ we define the coefficient

$$\psi(m) := \sup\left\{\left|\frac{\mathbf{P}(AB)|}{\mathbf{P}(A)\mathbf{P}(B)} - 1\right|;\ A \in \mathcal{F}_{-\infty}^{k},\ B \in \mathcal{F}_{k+m}^{\infty},\ \mathbf{P}(A)\mathbf{P}(B) > 0\right\}. \tag{8}$$

It is clear that the sequence $\{\psi(m)\}$ is not increase. A stationary sequence of random variables is called a *sequence with ψ-mixing* if $\lim_{m\to\infty}\psi(m) = 0$. Such dependency conditions were studied in [13, 94, 111]. If a sequence of random variables satisfies ψ-mixing then it satisfies φ-mixing as well.

We also note that, under ψ-mixing condition, relation (8) implies absolutely continuity of any finite dimensional distribution with respect to the corresponding product-measure generated by the marginal distribution. For instance, if the random variable X_1 has the $[0,1]$-uniform distribution then the pair (X_0, X_m) has an absolutely continuous distribution on the square $[0,1]^2$ with a density bounded by the constant $1+\psi(m)$. An analogous statement for every finite family $\{X_{i_1}, \dots, X_{i_m}\}$ is easily deduced by induction.

Main assumptions and definitions.

I. $X = [0,1]$.

II. The stationary sequence of random variables $\{X_i;\ i \in \mathbf{Z}\}$ satisfies ψ-mixing condition.

III. The random variable X_0 has the $[0,1]$-uniform distribution.

Notice that, in the univariate case, without loss of generality we can study the $[0,1]$-uniform sample distribution since by the corresponding quantile transform we can reduce any sample distribution to the $[0,1]$-uniform one. Thus, after the corresponding redetermination of the kernel, we deal with a new V-statistic based on a stationary connected observations with $[0,1]$-uniform marginal distribution.

Denote by $F_k(t,s)$ the joint distribution function of the couple (X_0, X_k). Due to the comment above every distribution function $F_k(t,s)$ has a density which is denoted by $p_k(t,s)$. By (8) these densities are uniformly bounded on the square $[0,1]^2$.

Introduce a centered Gaussian process with the covariance function

$$\mathbf{E}Y(t)Y(s) = \min(t,s) - ts + \sum_{k\ge1}\left(F_k(t,s) + F_k(s,t) - 2ts\right). \tag{9}$$

where $t, s \in [0,1]$. Notice that the definition (8) and summability condition of the coefficients $\psi(k)$ (it follows from condition (15) below) provide absolute summability of the function series on the right-hand side of (9) as well as its uniform boundedness on $[0,1]^2$.

Gaussian processes with the covariance of the form (9) represent weak limits of the sequence of the classical empirical processes $S_n((-\infty,t))$ under some dependency conditions of the random variables $\{X_k\}$. In particular, such weak convergence is valid if the sequence $\{X_k\}$ satisfies φ-mixing condition (hence, and for ψ-mixing sequences as well) under the following restrictions on the corresponding coefficients: (see [12]):

$$\sum_{k\ge1}\psi^{1/2}(k) < \infty. \tag{10}$$

We need a result from [23] regarding definition of multiple stochastic integrals with product-noises generated by increments of Gaussian processes like $Y(t)$.

Introduce some notation we need. Let μ be a noise generated by increments of the above-mentioned Gaussian process $Y(t)$ on the interval $[0,1]$:

$$\mu((t,t+\delta]) = Y(t+\delta) - Y(t), \quad \mu([0,\delta]) = Y(\delta) - Y(0).$$

Consider the function space

$$S_0 := \Big\{ f : \sum_{1\le j_1,\ldots,j_d \le d} \mathbf{E} f^2(X^*_{j_1},\ldots,X^*_{j_d}) < \infty \Big\},$$

where $\{X^*_j\}$ are *independent* $[0,1]$-uniformly distributed random variables. In this space we define the combined L_2-norm

$$\|f\|^2 := \sum_{1\le j_1,\ldots,j_d \le d} \mathbf{E} f^2(X^*_{j_1},\ldots,X^*_{j_d}). \tag{11}$$

Notice that the normed function space S_0 is *embedded* to the normed space S introduced in [23]. This embedding follows from conditions I–III and from uniform boundedness (due to (8) and condition II) on the square $[0,1]^2$ of the function

$$b(t,s) := \sum_{k\ge 1} |p_k(t,s) + p_k(s,t) - 2|$$

which plays a key role in the construction of the corresponding stochastic integral in [23]. Therefore, we immediately deduce from [23] the following statement.

Theorem 1. *Let $f \in S_0$. Then there exists a sequence of step functions of the form*

$$f_N(x_1,\ldots x_d) := \sum_{j_1,\ldots,j_d=1}^{N} f_{j_1,\ldots,j_d} \prod_{k=1}^{d} I(x_k \in B_{j_k}) \tag{12}$$

such that, as $N \to \infty$, they converge to f in the norm (11) *of the function space S_0, where, for each $k \le d$, measurable* (*in the Lebesgue sense*) *subsets $\{B_{j_k}\}$ form a partition of the interval* $[0,1]$. *Moreover, as $N \to \infty$, the sequence*

$$\eta(f_N) := \sum_{j_1,\ldots,j_d=1}^{N} f_{j_1,\ldots,j_d} \prod_{k=1}^{d} \mu(B_{j_k}) \tag{13}$$

mean-square converges to some limit random variable $\eta(f)$ which does not depend on the sequence f_N.

The random variable $\eta(f)$ is called d-fold stochastic integral of a function f with respect to a noise generated by increments of a stochastic process $Y(t)$:

$$\eta(f) := \int_{X^d} f(t_1,\ldots,t_d)\, dY(t_1)\ldots dY(t_d). \tag{14}$$

We now formulate the main result of the paper.

Theorem 2. *Let the conditions* I–III *be fulfilled and*

$$\Psi(d) := \sum_{k\ge 1} \psi(k) k^{2d-2} < \infty. \tag{15}$$

Then, for any $f \in S_0$,

$$V_n \xrightarrow{d} \to \int\limits_{X^d} f(t_1, \ldots, t_d) dY(t_1) \ldots dY(t_d) \tag{16}$$

as $n \to \infty$.

2. Proof of Theorem 2

First of all, we prove a few auxiliary statements. We need the following elementary assertion ([77, 94]).

Lemma 1. *Let random variables* ξ *and* η *are measurable with respect to the* σ*-field* $\mathcal{F}_{-\infty}^{k}$ *and* $\mathcal{F}_{k+m}^{\infty}$ $(m \ge 1)$ *accordingly, and, moreover,* $\mathbf{E}|\xi| < \infty$ *and* $\mathbf{E}|\eta| < \infty$. *Then*

$$|\mathbf{E}\xi\eta - \mathbf{E}\xi\mathbf{E}\eta| \le \psi(m)\mathbf{E}|\xi|\mathbf{E}|\eta|. \tag{17}$$

Put $\widetilde{I}_k(A) := I(X_k \in A) - P(A)$, $A \in \mathcal{A}$; hereinafter the marginal distribution P is the Lebesgue measure on the interval $[0, 1]$.

Lemma 2. *For any natural numbers* q *and* $l_1, \ldots, l_q$ *as well as for any pair-wise disjoint measurable subsets* $A_1, \ldots, A_q \subseteq [0, 1]$ *the following inequality holds*

$$\mathbf{E}\big|\widetilde{I}_k^{\,l_1}(A_1) \ldots \widetilde{I}_k^{\,l_q}(A_q)\big| \le (q+1)P(A_1) \ldots P(A_q). \tag{18}$$

Proof. We will use induction on q.

1. The case $q = 1$. It is clear that

$$\mathbf{E}\big|\widetilde{I}_k^{\,l_1}(A)\big| \le \mathbf{E}|\widetilde{I}_k(A)| \le 2P(A).$$

2. Assume that inequality (18) is true for some $q \ge 1$. We then prove that it is true for $q + 1$. We have

$$\mathbf{E}\big|\widetilde{I}_k^{\,l_1}(A_1) \ldots \widetilde{I}_k^{\,l_{q+1}}(A_{q+1})\big| \le \mathbf{E}|\widetilde{I}_k(A_1) \ldots \widetilde{I}_k(A_{q+1})|$$

$$\le \mathbf{E}|\widetilde{I}_k(A_1) \ldots \widetilde{I}_k(A_q) I_k(A_{q+1})| + P(A_{q+1})\, \mathbf{E}|\widetilde{I}_k(A_1) \ldots \widetilde{I}_k(A_q)|$$

$$= \mathbf{E}|(-1)^q P(A_1) \ldots P(A_q) I_k(A_{q+1})| + P(A_{q+1})\, \mathbf{E}|\widetilde{I}_k(A_1) \ldots \widetilde{I}_k(A_q)|$$

$$\le P(A_1) \ldots P(A_{q+1}) + P(A_{q+1})(q+1)P(A_1) \ldots P(A_q).$$

The next to last equality above is valid due to the fact that the subsets $A_1, \dots, A_{q+1}$ are pair-vise disjoint. The last inequality is valid due to the induction condition. Thus,

$$\mathbf{E}\big|\widetilde{I}_k^{\,l_1}(A_1)\dots\widetilde{I}_k^{\,l_{q+1}}(A_{q+1})\big| \le (q+2)P(A_1)\dots P(A_{q+1}).$$

The Lemma is proved.

Lemma 3. *Let $q_1 < \cdots < q_s$ be arbitrary natural numbers. Consider arbitrary s collections of measurable subsets of the unit interval: $\{A_1, \dots A_{q_1}\}, \dots, \{A_{q_{s-1}+1}, \dots A_{q_s}\}$, where within every collection the subsets are pair-wise disjoint. Put*

$$\nu_{k_i} := \widetilde{I}_{k_i}^{\,l_{q_{i-1}+1}}(A_{q_{i-1}+1})\dots\widetilde{I}_{k_i}^{\,l_{q_i}}(A_{q_i}).$$

Then for any natural numbers $k_1 < \cdots < k_s$ and $l_1, \dots, l_{q_1}, \dots, l_{q_s}$ there is the following estimate:

$$\mathbf{E}|\nu_{k_1}\dots\nu_{k_s}| \le C(\psi(1), s, q_s)P(A_1)\dots P(A_{q_s}), \tag{19}$$

where the constant $C(\cdot)$ depends only on the arguments indicated.

Proof. By (18) we have

$$\mathbf{E}|\nu_{k_i}| \le (q_i - q_{i-1} + 1)P(A_{q_{i-1}+1})\dots\mathbf{P}(A_{q_i}).$$

It is clear that the random variables ν_{k_i} satisfy ψ-mixing condition. Using (17) we then obtain

$$\mathbf{E}|\nu_{k_1}\dots\nu_{k_s}| \le \prod_{j=1}^{s-1}(1+\psi(k_{j+1}-k_j))\mathbf{E}|\nu_{k_1}|\dots\mathbf{E}|\nu_{k_s}| \le C(\psi(1), s, q_s)P(A_1)\dots P(A_{q_s}).$$

The Lemma is proved.

A key auxiliary statement to prove the main result is estimating mixed moments of the form $\mathbf{E}S_n(A_1)\dots S_n(A_{2d})$.

Lemma 4. *Let d be a natural number and $l_1, \dots, l_q$ be a collection of natural numbers such that $l_1 + \cdots + l_q = 2d$, $d \ge 2$ and let $A_1, \dots, A_q$ be pair-wise disjoint measurable subsets of the interval $[0,1]$. Under condition (15) the following estimate is valid:*

$$\big|\mathbf{E}S_n^{l_1}(A_1)\dots S_n^{l_q}(A_q)\big| \le C(d, \Psi(d))P(A_1)\dots P(A_q), \tag{20}$$

where the constant $C(\cdot)$ depends only on the arguments indicated.

Proof. We start with the following simple estimate:

$$\big|\mathbf{E}S_n^{l_1}(A_1)\dots S_n^{l_q}(A_q)\big|$$

$$\le n^{-d}\sum_{k_1,\dots,k_{2d}\le n}|\mathbf{E}\widetilde{I}_{k_1}(A_1)\dots\widetilde{I}_{k_{l_1}}(A_1)\dots\widetilde{I}_{k_{2d-l_q+1}}(A_q)\dots\widetilde{I}_{k_{2d}}(A_q)|,$$

where the subscripts k_i independently take all integers from 1 to n.

The initial sum on the right-hand side of this inequality can be estimated by a finite sum of the following diagonal subsums

$$\sum_{k_1<\cdots<k_r\le n} |\mathbf{E}\nu_{k_1}\dots\nu_{k_r}|; \tag{21}$$

we denoted here $\nu_{k_i} := \widetilde{I}_{k_i}^{s_1(i)}(A_1)\dots\widetilde{I}_{k_i}^{s_q(i)}(A_q)$, where the integers $s_j(i)$ are defined by the corresponding "diagonal subspace" of the subscripts in the initial multiple sum and they satisfy the conditions $0 \le s_j(i) \le l_j$ for all $i \le r$ and $j \le q$, as well as $\sum\limits_{i\le r}\sum\limits_{j\le q} s_j(i) = 2d$.

Let $r \le d$. Estimating by (19) each summand in (21) and taking the normalized factor n^{-d} and the number of summands in (21) into account we obtain the upper bound we need.

In the sequel we do not indicate (say, as in Lemma 3) an obvious dependency of constants C or C_i on parameters of the problem under consideration. At the same time we use the subscripts only in the case when we need to distinguish some constants. Due to monotonicity of the function $\Psi(\cdot)$ we can conclude that the final constant depends only on the values d and $\Psi(d)$.

Let now $r > d$. We call the random variable ν_{k_i} *short product* if $\sum\limits_{j\le q} s_j(i) = 1$, i. e., $\nu_{k_i} = \widetilde{I}_{k_i}(A_{q_i})$ for some $q_i \le q$. Notice that if ν_{k_i} is a short product then

$$\mathbf{E}\nu_{k_i} = 0$$

.

To evaluate the sum in (21) we now consider an auxiliary multiple sum of the form

$$\sum_{k_{v_1}<\cdots<k_{v_2}\le n} |\mathbf{E}\nu_{k_{v_1}}\dots\nu_{k_{v_2}}|, \tag{22}$$

where $1 \le v_1 < v_2 \le r$, and the value $v := v_2 - v_1 + 1$ is the dimension of the corresponding multiple sum and ν_{k_i} are defined in (21). Introduce the following notation: $e_j(i) := \min\{1, s_j(i)\}$. We first prove the following assertion: If, in the summands in (22), there are at least m shorts products, where $0 \le m \le v$, then the following upper bound is valid:

$$\sum_{k_{v_1}<\cdots<k_{v_2}\le n} |\mathbf{E}\nu_{k_{v_1}}\dots\nu_{k_{v_2}}| \le Cn^{v-m/2}\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)}, \tag{23}$$

where $\alpha_j(v_1, v_2) := \sum\limits_{i=v_1}^{v_2} e_j(i)$. Notice that the set-function $\alpha_j(a, b)$ is additive on intervals $[a, b]$.

We prove this assertion by induction on m for all v_1 and v_2 such that $v \ge m$ and $v \le r$.

Let $m = 1$, i. e., the expectations in (22) contain at least one short product. Denote it by ν_{k_l}, where $k_{v_1} \le k_l \le k_{v_2}$. First we note that, in terms of the notation above, we can reformulate the statement of Lemma 3 for the absolute moment of each random product in (22) in such a way:

$$\mathbf{E}|\nu_{k_{v_1}}\dots\nu_{k_{v_2}}|\le C\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)}.$$

Taking this estimate into account we evaluate by (17) and (19) every summand in (22) setting $\xi:=\nu_{k_{v_1}}\dots\nu_{k_l}$ $\eta:=\nu_{k_{l+1}}\dots\nu_{k_{v_2}}$:

$$\sum_{k_{v_1}<\dots<k_{v_2}\le n}|\mathbf{E}\nu_{k_{v_1}}\dots\nu_{k_{v_2}}|$$

$$\le\sum_{k_{v_1}<\dots<k_{v_2}\le n}\psi(k_{l+1}-k_l)\mathbf{E}|\nu_{k_{v_1}}\dots\nu_{k_l}|\mathbf{E}|\nu_{k_{l+1}}\dots\nu_{k_{v_2}}|$$

$$+\sum_{k_{v_1}<\dots<k_l\le n}|\mathbf{E}\nu_{k_{v_1}}\dots\nu_{k_l}|\sum_{k_{l+1}<\dots<k_{v_2}\le n}\mathbf{E}|\nu_{k_{l+1}}\dots\nu_{k_{v_2}}|$$

$$\le C_1n^{v-1}\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)}\sum_{i\ge1}\psi(i)$$

$$+C_2n^{v_2-l}\prod_{j\le q}P(A_j)^{\alpha_j(l+1,v_2)}\sum_{k_{v_1}<\dots<k_l\le n}\psi(k_l-k_{l-1})\mathbf{E}|\nu_{k_1}\dots\nu_{k_{l-1}}|\mathbf{E}|\nu_{k_l}|$$

$$\le(C_3n^{v-1}+C_4n^{v_2-l}n^{l-v_1})\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)}\le C_5n^{v-1/2}\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)},$$

which required. In this chain of relations the second inequality is valid due to (17) and the equality $\mathbf{E}\nu_{k_l}=0$ as well.

We now assume that the upper bounds

$$\sum_{k_{v_1}<\dots<k_{v_2}\le n}|\mathbf{E}\nu_{k_{v_1}}\dots\nu_{k_{v_2}}|\le Cn^{v-z/2}\prod_{j\le q}P(A_j)^{\alpha_j(v_1,v_2)}$$

are true for all integers $z<m$, where z is the minimal possible number of short products in the expectations under consideration, and for all possible dimensions $v:\ z\le v\le r$ of multiple sums of the form (22), and, moreover, the moments in (22) contain at least m shorts products. Denote these products by $\nu_{k_{j_1}},\dots,\nu_{k_{j_m}}$. Consider $m-1$ pairs of neighboring products: $\nu_{k_{j_s}},\nu_{k_{j_{s+1}}},\ s=1,\dots,m-1$ Denote by $t_1,\dots,t_{m-1}$ differences between the subscripts in these pairs. We have

$$\sum_{k_{v_1}<\dots<k_{v_2}\le n}|\mathbf{E}\nu_{k_{v_1}}\dots\nu_{k_{v_2}}|\le R_1+\dots+R_{m-1},$$

where the sum R_s is taken over the set of subscripts

$$I_s:=\{(k_{v_1},\dots,k_{v_2}):k_{v_1}<\dots<k_{v_2}\le n,\ t_s=\max t_i\}.$$

We now estimate by (17) each summand in R_s setting

$$R_s \le \sum_{I_s} \psi(k_{j_s+1}-k_{j_s})\mathbf{E}|\nu_{k_{v_1}}\dots\nu_{k_{j_s}}|\,\mathbf{E}|\nu_{k_{j_s+1}}\dots\nu_{k_{v_2}}|$$

$$+\sum_{k_{v_1}<\dots<k_{j_s}} |\mathbf{E}\nu_{k_{v_1}}\dots\nu_{k_{j_s}}| \sum_{k_{j_s+1}<\dots<k_{v_2}} |\mathbf{E}\nu_{k_{j_s+1}}\dots\nu_{k_{v_2}}|. \qquad (24)$$

Consider the first sum on the right-hand side of (24). By (19) we have

$$\sum_{I_s} \psi(k_{j_s+1}-k_{j_s})\mathbf{E}|\nu_{k_{v_1}}\dots\nu_{k_{j_s}}|\,\mathbf{E}|\nu_{k_{j_s+1}}\dots\nu_{k_{v_2}}| \le C\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)} \sum_{I_s}\psi(t_s)$$

$$\le C\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)} n^{v-(m-1)} \sum_{t_i:\, t_i\le t_s} \psi(t_s)$$

$$\le C\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)} n^{v-m+1} \sum_{k}\psi(k)k^{m-2}$$

$$\le C\Psi(m/2)n^{v-m+1} \le C_1 n^{v-m/2}\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)}.$$

Notice that the last inequality is valid for $m \ge 2$.

Consider now the product of the sums on the right-hand side of (24). Let the summands of the first sum contain m_1 short products indicated above, and, in the summands of the second sum, there are $m-m_1$ short products indicated above. By the construction we have $1 \le m_1 \le m-1$. Hence, for these sums, we can use the induction condition. Finally, we have

$$\sum_{k_{v_1}<\dots<k_{j_s}} |\mathbf{E}\nu_{k_{v_1}}\dots\nu_{k_{j_s}}| \sum_{k_{j_s+1}<\dots<k_{v_2}} |\mathbf{E}\nu_{k_{j_s+1}}\dots\nu_{k_{v_2}}|$$

$$\le Cn^{j_s-m_1/2}n^{v-j_s-(m-m_1)/2}\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)} \le C_1 n^{v-m/2}\prod_{j\le q} P(A_j)^{\alpha_j(v_1,v_2)},$$

which required. Thus, for R_s, we obtained the upper bound we need. It implies the estimate in (23). The induction is over.

To finish the proof of Lemma 4 we should note that, first, by the definition, $\alpha_j(1,r) \ge 1$ for all $j \le q$, and, second, in the case $r > d$, the summands in (21) contain at least $2(r-d)$ short products. So, we should put in (23) $v_1 := 1$, $v_2 := r$, $m := 2(r-d)$ and $v := r$. It means that, for the sum in (21), the following upper bound is valid:

$$\sum_{k_1<\dots<k_r\le n} |\mathbf{E}\nu_{k_1}\dots\nu_{k_r}| \le Cn^d\prod_{j\le q} P(A_j)^{\alpha_j(1,r)} \le Cn^d P(A_1)\dots P(A_q).$$

The Lemma is proved.

R e m a r k. Under a stronger restriction on the coefficient $\psi(\cdot)$ the statement of Lemma 4 is also contained in [111] (in addition, see [77]). However, the proof in [111] was carried out only in the case $d = 2$, and the corresponding constant in the upper bound in [111] contains the factor $(1+\psi(0))^2$. It is easy to see that if the marginal distribution has a continuous component (for example, if it is the Lebesgue measure on $[0,1]$) or infinitely many atoms then $\psi(0) = \infty$. To verify this property we may put in (8) $A = B$, where A is an event from the σ-field $\mathcal{F}_0^0$. So, under the above-mentioned restrictions on the marginal P

$$\psi(0) \geq \sup_{A \in \mathcal{F}_0^0,\ P(A)>0} (1/P(A) - 1) = \infty.$$

We now study the integral sums for the integral representation of V-statistics in (5), where we substitute the step functions f_N for the initial kernel:

$$I_N^n := \sum_{j_1,\ldots,j_d \leq N} f_{j_1,\ldots,j_d} \prod_{k=1}^{d} S_n(B_{j_k}) = \int_{X_d} f_N(x_1,\ldots,x_d) \prod_{k=1}^{d} S_n(dx_k). \tag{25}$$

Lemma 5. *Let step functions of the form* (12) *converge to f in the norm of the function space S_0. Then the random variables* (25) *mean-square converge to the corresponding Von Mises statistic* (5) *uniformly on n.*

Proof It suffices to prove the Cauchy property of the sequence $\{I_N^n\}$ uniformly on n as $N \to \infty$.

Put $f_{M,N}(x_1,\ldots,x_d) := f_N(x_1,\ldots,x_d) - f_M(x_1,\ldots,x_d)$. Observe that the step functions $f_N(x_1,\ldots,x_d)$ and $f_M(x_1,\ldots,x_d)$ can be represented as finite linear combinations of indicator-type functions of the minimal collections of pair-wise disjoint subsets generated both families of subsets in the definition of f_N and f_M. Denote these subsets by the symbols A_j. We also denote by $f^{N,M}_{j_1,\ldots,j_d}$ the corresponding values of the step functions $f_{M,N}(x_1,\ldots,x_d)$ on elements of the above-mentioned universal partition. We then have

$$\begin{aligned}
\mathbf{E}(I_N^n - I_M^n)^2 &= \mathbf{E}\left(\int_{X_d} f_{M,N}(x_1,\ldots,x_d) \prod_{k=1}^{d} S_n(dx_k)\right)^2 \\
&= \sum_{j_1,\ldots,j_{2d}} f^{N,M}_{j_1,\ldots,j_d} f^{N,M}_{j_{d+1},\ldots,j_{2d}} \mathbf{E} S_n(A_{j_1}) \ldots S_n(A_{j_{2d}}) \\
&\leq C \sum_{i_1,\ldots,i_{2d} \leq d} \mathbf{E}|f_{M,N}(X^*_{i_1},\ldots,X^*_{i_d}) f_{M,N}(X^*_{i_{d+1}},\ldots,X^*_{i_{2d}})|.
\end{aligned}$$

Notice that the last inequality follows from (20) since the mixed moment $\mathbf{E}S_n(A_{j_1})\ldots S_n(A_{j_{2d}})$ can be represented in the form as in Lemma 4 (taking multiplicities of subscripts j_k into account). Using the Cauchy —Bunyakowsky inequality as well as the elementary evaluation of the sum on the right-hand side of the previous inequality we obtain the upper bound

$$\mathbf{E}\big(I_N^n - I_M^n\big)^2 \leq C\|f_{M,N}\|^2 = C\|f_M - f_N\|^2.$$

The right-hand side of the last inequality does not depend on n and vanishes as $M, N \to \infty$ since, by the condition of the Theorem, the sequence f_N converges in the norm of the function space S_0. Hence it is a Cauchy sequence. Thus we proved the Cauchy property (in the mean-square norm) of the sequence $\{I_N^n\}$ uniformly on n. So, we proved the uniform mean-square convergence of the integral sums (based on the empirical measure) to the corresponding stochastic integral, i.e., to Von Mises statistic of the form (5). The Lemma is proved.

Lemma 6. *The sequence of random variables I_N^n introduced in* (25) *converges in distribution to the random variable $\eta(f_N)$ in* (13) *as $n \to \infty$.*

Proof. This assertion is a direct consequence of the multivariate central limit theorem for finite-dimensional distributions of the standard empirical processes based on stationary connected observations with φ- or ψ-mixing. Notice that this multivariate theorem follows from the univariate version for stationary sequences of weakly dependent random variables if we take the classical Cramér—Wold method into account. This method allows us to reduce the multivariate case to the univariate one (for example, see [12, 94]). Notice also that to apply the multivariate central limit theorem for empirical processes we need the condition

$$\sum_{k\geq 1} \psi^{1/2}(k) < \infty$$

which follows from condition (15) in the case $d = 2$, due to the Cauchy—Bunyakowsky inequality:

$$\sum_{k\geq 1} \psi^{1/2}(k) \leq \sum_{k\geq 1} \psi(k)k^2 \sum_{k\geq 1} k^{-2}.$$

On the other hand, this condition and Lemma 1 provide fulfillment of the conditions of Theorem 6 in [94] (the central limit theorem for stationary sequences of numerical random variables)

The next elementary assertion is contained, in fact, in [12]. We formulate it in a more convenient form.

Lemma 7*Let $\xi, \xi_1, \xi_2, \ldots$ be a sequence of random variables taking values in an arbitrary measurable space $(X, \mathcal{B})$ and defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. Let $\mathcal{F} := \{F\}$ be a family of $\mathcal{B}$-measurable functionals in X such that $F(\xi_n) \stackrel{d}{\to} \to F(\xi)$ as $n \to \infty$. Assume that there exist numerical random variables $\eta, \eta_1, \eta_2, \ldots$ defined on $(\Omega, \mathcal{A}, \mathbf{P})$ such that for some sequence $F_k \in \mathcal{F}$, as $k \to \infty$, the following relations are valid:*

(a) $F_k(\xi) \to \eta$ *in the mean-square norm,*

(b) $F_k(\xi_n) \to \eta_n$ *in the mean-square norm uniformly on n.*

Then $\eta_n \stackrel{d}{\to} \to \eta$.

Proof of Theorem 2. We set in the notation of Lemma 7

$$\eta := \eta(f) = \int\limits_{X^d} f(t_1, \ldots, t_d)\, dY(t_1) \ldots dY(t_d), \quad \eta_n := V_n,$$

$$\xi := \mu(\cdot), \quad F_k(\xi) := \eta(f_k), \quad \xi_n := S_n(\cdot), \quad F_k(\xi_n) := I_k^n.$$

We consider $\mathcal{F}$ as the class of all multilinear transforms of the form (25) which are defined of all elementary stochastic measures. Without loss of generality we assume the all the random variables under consideration are defined on a common probability space (for instance, we may assume that the random process $Y(t)$ and the sequence $\{X_i\}$ are independent). Convergence (a) follows from Theorem 1 and convergence (b) follows from Lemma 5. Finally, the weak convergence $F(\xi_n) \xrightarrow{d} \to F(\xi)$ was proved in Lemma 6.

Thus, Theorem 2 is proved.

12 The functional limit theorem for the canonical U-processes based on dependent trials

1. Main Definitions and Notions

From the mid-1950s the functional limit theorem is available, i.e., the invariance principle for the partial sum processes based on the sequences of independent and weakly dependent random variables (for example, see [12, 31, 53]). In the 1980s the limit theory (including the invariance principle) was created for a more general object, the so-called U-statistics and Von Mises' statistics (V-statistics) of arbitrary order which are based on independent trials (for example, see [42, 44, 104, 109]). To study the limit behavior of noncanonical U- and V-statistics, as a rule we reduce the problems to the asymptotic analysis of the sums of random variables. But in the case of canonical statistics, the study is much complicated. For independent observations, the limit distribution of such statistics can be represented as the distribution of a polynomial of infinite sequence of independent Gaussian random variables (see [109]) or as the distribution of a multiple stochastic integral with the integrating Wiener product-measure (see [44]). The respective weak limits in the functional limit theorem can be represented as either the analogous polynomials of independent Wiener processes (see [104]) or a one-parameter family of multiple stochastic integrals with the integrating Gaussian product-measure generated by the so-called Kiefer two-parametric process (see [42]).

In the case of weakly dependent observations, the asymptotic behavior of canonical U- and V-statistics is essentially complicated in comparison with the case of independent observations. First of all, this remark relates to description of the limit distribution as a multiple stochastic integral (see [24]). In the case of dependent trials, the approach is worth noting that is based on the orthogonal series technique (see [25, 109]) that was firstly applied to independent trials and canonical statistics of second order in 1947 in the classical paper by Von Mises [89]. Later this result was extended to the canonical statistics of arbitrary order (see [109]).

Recall some principal points of the above-mentioned approach in the case of weakly dependent observations under ϕ-mixing condition (for detail, see [25]).

Let $X_1, X_2, \dots$ be a stationary sequence of random variables defined on some probability space $(\Omega, F, \mathbf{P})$. Denote by F the distribution of X_1. Consider $f \in L_2(\mathbf{R}^m, F^m)$, where F^m is the product-measure with marginal F. Then the following series expansion is valid (see [11]):

$$f(t_1, \dots, t_m) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} f_{k_1 \dots k_m} e_{k_1}(t_1) \dots e_{k_m}(t_m), \tag{1}$$

and this series converges in $L_2(\mathbf{R}^m, F^m)$; here $\{e_{k_j}\}$ is an orthonormal basis for $L_2(\mathbf{R}, F)$ and, without loss of generality, we can assume that $e_0 \equiv 1$. Then $\mathbf{E}e_j(X_1) = 0$, $j \geq 1$, by orthogonality with e_0, and $\mathbf{E}e_i(X_1)e_j(X_1) = \delta_{i,j}$ for all $i \neq j$.

Denote by $\{X_i^*\}$ the sequence of independent copies of X_1. If the expansion coefficients $\{f_{k_1 \dots k_m}\}$ are absolutely summable then, by B. Levi's theorem and the simple estimate $\mathbf{E}|e_{k_1}(X_1^*)\dots e_{k_m}(X_m^*)| \le 1$, the series in (1) almost surely converges after replacement of the nonrandom arguments $t_1,\dots,t_m$ with the random variables $X_1^*,\dots,X_m^*$.

Definition 1. A function $f(t_1,\dots,t_m)\in L_2(\mathbf{R}^m, F^m)$ is called *canonical* (or *degenerate*) if

$$\mathbf{E}f(y_1,\dots,y_{i-1},X_1,y_{i+1},\dots,y_m)=0$$

for all $y_j\in\mathbf{R}$ and $i\in\{1,\dots,m\}$, where the cases $i=1$ and $i=m$ correspond to the extreme positions of the coordinate X_1 of the vector argument of f.

Note the important property of canonical functions (see [25]):

Proposition 1 *If $f(t_1,\dots,t_m)$ is canonical then e_0 is absent in (1), i.e.,*

$$f(t_1,\dots,t_m)=\sum_{k_1=1}^{\infty}\cdots\sum_{k_m=1}^{\infty} f_{k_1\dots k_m} e_{k_1}(t_1)\dots e_{k_m}(t_m). \tag{2}$$

Define the *canonical U-statistic* of a sample of size n from a stationary sequence of observations:

$$U_n := n^{-m/2}\sum_{1\le i_1\ne}\cdots\sum_{\ne i_m\le n} f(X_{i_1},\dots,X_{i_m}),$$

where f is canonical.

In the chapter we study the sequence of U-statistics

$$U_n(t) := n^{-m/2}\sum_{1\le i_1\ne}\cdots\sum_{\ne i_m\le [nt]} f(X_{i_1},\dots,X_{i_m}),\quad t\in[0,1],$$

as a stochastic process in $D[0,1]$ which is called a *U-process.*

The asymptotic behavior of canonical U-statistics has been thoroughly studied. For instance, in the case of independent observations it was proved in [109] that

$$U_n \xrightarrow{d} \sum_{k_1=0}^{\infty}\cdots\sum_{k_m=0}^{\infty} f_{k_1\dots k_m}\prod_{j=1}^{\infty} H_{v_j(k_1,\dots,k_m)}(\tau_j), \tag{3}$$

where $\{\tau_j\}$ is a sequence of independent random variables having the standard normal distribution, $v_j(i_1,\dots,i_m)$ is the number of the subscripts among $i_1,\dots,i_m$ that equal j, and $H_k(x)$ are the classical Hermite polynomials defined by the formula

$$H_k(x)=(-1)^k\exp(x^2/2)\frac{d^k}{dx^k}\exp(-x^2/2),\quad k\ge 0.$$

In [25], an analog of (3) was obtained for observations under α- and φ-mixing conditions.

In the chapter we study only stationary sequences $\{X_i\}$ satisfying φ-mixing. Recall the corresponding definition. Denote by M_j^k, with $j\le k$, the σ-field of the events generated by $X_j,\dots,X_k$.

Definition 2. A sequence of random variables $X_1,X_2,\dots$ satisfies *φ-mixing* if

$$\varphi(i) := \sup_{k\ge 1}\ \sup_{A\in M_1^k B\in M_{k+1}^{\infty}\mathbf{P}(\mathbf{A})>\mathbf{0}} \frac{|\mathbf{P}(AB)-\mathbf{P}(A)\mathbf{P}(B)|}{\mathbf{P}(A)} \to 0 \quad \text{as } i\to\infty.$$

In the sequel we will assume that

$$\sum_{k=1}^{\infty} \varphi(k)^{1/2} < \infty.$$

Notice that this well-known condition provides the central limit theorem for the corresponding stationary sequence of random variables (for example, see [12]).

To study dependent observations, the principle difficulty occurs: After replacement of the nonrandom arguments $(t_1, \ldots, t_m)$ with the dependent random variables $(X_1, \ldots, X_m)$, equality in (2) may be false with positive probability (see the corresponding counterexample in [25]). Introduce some restriction on the joint distributions of the dependent random variables $\{X_i\}$, which provides possibility of the above-mentioned replacement in (2) of nonrandom arguments with random arguments (see [25]):

(AC) *For every collection of pairwise distinct subscripts $(j_1, \ldots, j_m)$ the distribution of the random vector $(X_{j_1}, \ldots, X_{j_m})$ is absolutely continuous with respect to the distribution of $(X_1^*, \ldots, X_m^*)$.*

For instance, Condition (AC) is fulfilled for any stationary sequence under the so-called ψ-mixing (see [24, 25]). Moreover, it is easy to define some sequence of moving averages based on a sequence of $[0, 1]$-uniformly distributed independent random variables such that m elements from the sequence of moving averages have bounded density of the joint distribution, which means the fulfilment of Condition (AC).

As noted above, the condition

$$\sum_{k_1=1}^{\infty} \cdots \sum_{k_m=1}^{\infty} |f_{k_1 \ldots k_m}| < \infty \tag{4}$$

implies convergence of (2) almost surely with respect to the distribution of $(X_1^*, \ldots, X_m^*)$. Hence, under Condition (AC), the convergence is valid almost surely with respect to the distribution of the random vector $(X_{j_1}, \ldots, X_{j_m})$. In other words, under Condition (AC), we can substitute in (2) the random variables $X_{j_1}, \ldots, X_{j_m}$ for nonrandom arguments $t_1, \ldots, t_m$ for all pairwise distinct subscripts $j_1, \ldots, j_m$.

Thus, under the fulfilment of Condition (AC) and (4), every U-statistic can be represented as the following series converging with probability 1:

$$U_n(t) = n^{-m/2} \sum_{1 \le i_1 \ne} \cdots \sum_{\ne i_m \le [nt]} \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} f_{k_1 \ldots k_m} e_{k_1}(X_{i_1}) \ldots e_{k_m}(X_{i_m})$$

$$= n^{-m/2} \sum_{k_1=0}^{\infty} \cdots \sum_{k_m=0}^{\infty} f_{k_1 \ldots k_m} \sum_{1 \le i_1 \ne} \cdots \sum_{\ne i_m \le [nt]} e_{k_1}(X_{i_1}) \ldots e_{k_m}(X_{i_m}).$$

Further arguments are quite similar to those in the case of independent observations. They are reduced to sequential extraction of statistics with splitting kernels from the multiple sum on the right-hand side of the identity above. Indeed, the expression

$$n^{-m/2} \sum_{1 \le i_1 \ne} \cdots \sum_{\ne i_m \le [nt]} e_{k_1}(X_{i_1}) \ldots e_{k_m}(X_{i_m})$$

is represented as a linear combination of products of the following values:

$$n^{-1/2}\sum_{i=1}^{[nt]} e_k(X_i), \quad n^{-1}\sum_{i=1}^{[nt]} e_{k_1}(X_i)e_{k_2}(X_i), \ldots, n^{-l/2}\sum_{i=1}^{[nt]} e_{k_1}(X_i)\cdots e_{k_l}(X_i).$$

The proof is pure combinatoric and does not depend on the joint distribution of $\{X_i\}$ (see [109]).

Below we presume that the basis $\{e_j(t)\}$ in (2) satisfies the following additional restriction to be of use in proving various limit theorems for dependent random variables (see [25]):

$$\sup_i \mathbf{E}|e_i(X_1)|^m < \infty. \tag{5}$$

Introduce a sequence of dependent Wiener processes $\{w_i(t)\}$ with joint covariance

$$\mathbf{E}w_k(t_1)w_k(t_2) = \min(t_1,t_2)\left(1 + 2\sum_{j=1}^{\infty} \mathbf{E}e_k(X_1)e_k(X_{j+1})\right); \tag{6}$$

$$\mathbf{E}w_k(t_1)w_l(t_2) = \min(t_1,t_2)\left(\sum_{j=1}^{\infty} \mathbf{E}e_k(X_1)e_l(X_{j+1}) + \sum_{j=1}^{\infty} \mathbf{E}e_l(X_1)e_k(X_{j+1})\right), \quad l \neq k.$$

Finiteness of the series in (6) follows from the above-mentioned condition on the mixing coefficient (see [12]). The sequence of dependent Wiener processes $\{w_i(t)\}$ with joint covariance (6) exists due to the Kolmogorov extension theorem (for example, see [27, 31]) and it will play the role of the weak limit as $n \to \infty$ for the sequence of stochastic processes $\left\{n^{-1/2}\sum_{j=1}^{[nt]} e_i(X_j);\ i \geq 1\right\}$. Notice that, for every fixed N, the multidimensional stochastic process $\left\{n^{-1/2}\sum_{j=1}^{[nt]} e_i(X_j);\ 1 \leq i \leq N\right\}$ C-converges to $\{w_i(t);\ 1 \leq i \leq N\}$ as $n \to \infty$ by the corresponding invariance principle. Note that we consider the case when the coefficient of $\min(t_1,t_2)$ in (6) vanishes. In other words, it is convenient for us to interpret the zero function on $[0,1]$ as a Wiener process with zero variance. It is worth noting that the class of all degenerate distributions is a set of limit points for the class of Gaussian distributions in the weak convergence topology.

Recall that we say about C-convergence of k-dimensional stochastic processes $\{\xi_n(t)\}$ having the paths in $D^k[0,1]$ with the Skorokhod product-topology, to an a.s. continuous process $\xi(t)$ if, for every measurable functional $g(\cdot)$ continuous at the points of $C^k[0,1]$ in the sup-norm, the sequence $g(\xi_n)$ converges in distribution to the random variable $g(\xi)$ (see [28]).

Closing this section, we formulate some important statement that is an analog of the classical Rosenthal's moment inequality for sums of independent random variables.

Theorem 1 [119] *Let $\{\xi_i\}$ be a sequence of centered random variables with finite moments of order $t \geq 2$ satisfying φ-mixing, and moreover $\varphi := \sum_{k=1}^{\infty} \varphi^{1/2}(2^k) < \infty$. Then, for $t \geq 2$,*

$$\mathbf{E}\max_{1\leq k\leq n} |S_k|^t \leq (tc(\varphi))^t \left(\sum_{i=1}^{n} \mathbf{E}|\xi_i|^t + \left(\sum_{i=1}^{n} \mathbf{E}|\xi_i|^2\right)^{t/2}\right),$$

where the constant $c(\varphi)$ depends only on φ.

2. The Functional Limit Theorem for U-Processes

Introduce the random process

$$U(t) := \sum_{k_1=1}^{\infty} \cdots \sum_{k_m=1}^{\infty} f_{k_1\dots k_m} t^{m/2} \prod_{j=1}^{\infty} H_{v_j(k_1,\dots,k_m)}(t^{-1/2} w_j(t)). \tag{7}$$

Theorem 2*Let a stationary sequence of random variables $\{X_i\}$ satisfy φ-mixing, with the restrictions* (5) *and $\sum_{k=1}^{\infty} \varphi(k)^{1/2} < \infty$. For the canonical kernel $f \in L_2(\mathbf{R}^m, F^m)$, let* (4) *and* (AC) *be fulfilled.*

Then, as $n \to \infty$, the sequence of stochastic processes $U_n(t)$ C-converges to the stochastic process $U(t)$ defined in (7)*, and the corresponding multiple series a.s. converges for every $t \in [0,1]$, and moreover it is continuous in t.*

Proof. To prove the C-convergence we need to verify convergence of the finite-dimensional distributions and the property of density of the family of prelimit distributions with respect to the uniform topology (for example, see [12]).

I. Convergence of the finite-dimensional distributions. We prove that, as $n \to \infty$,

$$(U_n(t_1), \dots, U_n(t_q)) \xrightarrow{d} (U(t_1), \dots, U(t_q)),$$

where $t_1, \dots, t_q$ is an arbitrary finite collection of points from $[0,1]$.

Introduce the corresponding partial sum in the definition of $U_n(t)$:

$$U_n^N(t) := n^{-m/2} \sum_{k_1=1}^{N} \cdots \sum_{k_m=1}^{N} f_{k_1\dots k_m} \sum_{1 \le i_1 \ne} \cdots \sum_{\ne i_m \le [nt]} e_{k_1}(X_{i_1}) \dots e_{k_m}(X_{i_m}),$$

and the analogous partial sum for $U(t)$:

$$U^N(t) := \sum_{k_1=1}^{N} \cdots \sum_{k_m=1}^{N} f_{k_1\dots k_m} t^{m/2} \prod_{j=1}^{\infty} H_{v_j(k_1,\dots,k_m)}(t^{-1/2} w_j(t)).$$

Firstly, given a natural N, we establish the convergence

$$\big(U_n^N(t_1), \dots, U_n^N(t_q)\big) \xrightarrow{d} (U^N(t_1), \dots, U^N(t_q)).$$

Consider the statistic

$$U_n^N(t) := n^{-m/2} \sum_{k_1=1}^{N} \cdots \sum_{k_m=1}^{N} f_{k_1\dots k_m} \sum_{1 \le i_1 \ne} \cdots \sum_{\ne i_m \le [nt]} e_{k_1}(X_{i_1}) \dots e_{k_m}(X_{i_m})$$

that is represented as a finite linear combination of U-statistics of the form

$$U_n^N(e_{k_1}, \dots, e_{k_m})(t) := n^{-m/2} \sum_{1 \le i_1 \ne} \cdots \sum_{\ne i_m \le [nt]} e_{k_1}(X_{i_1}) \dots e_{k_m}(X_{i_m}).$$

Further, by the arguments similar to those in the case of independent observations, we represent the last U-statistic as a sum of Von Mises' statistics where summation is taken over all different collections (not necessarily pairwise distinct) of the subscripts $j_1, \ldots, j_m$. Next, changing the order of summation, we reduce the problem to studying the polynomials of the following stochastic processes:

$$n^{-1/2} \sum_{i=1}^{[nt]} e_k(X_i),\ n^{-1} \sum_{i=1}^{[nt]} e_{k_1}(X_i) e_{k_2}(X_i), \ldots, n^{-l/2} \sum_{i=1}^{[nt]} e_{k_1}(X_i) \cdots e_{k_l}(X_i).$$

For all $2 < l \leq m$, as $n \to \infty$, these sums converge in probability to 0 since, by condition (5) and Hölder's inequality, the value $\mathbf{E}|e_{k_1}(X_i) \cdots e_{k_l}(X_i)|$ is finite. Therefore, by the law of large numbers for weakly dependent random variables, we have

$$n^{-1} \sum_{i=1}^{[nt]} e_{k_1}(X_i) \cdots e_{k_l}(X_i) \xrightarrow{p} t\mathbf{E} e_{k_1}(X_i) \cdots e_{k_l}(X_i).$$

Hence, for $l > 2$, we obtain

$$n^{-l/2} \sum_{i=1}^{[nt]} e_{k_1}(X_i) \cdots e_{k_l}(X_i) \xrightarrow{p} 0$$

for all $t \in [0, 1]$.

Thus, the summands containing such sums as factors, also converge in probability to zero. If $l = 2$ then we can also apply the law of large numbers to the sums under consideration. By the orthonormality of the basis, the limits of these values equal $t\delta_{k_1,k_2}$, where δ_{k_1,k_2} is the Kronecker delta. Thus, the limiting result for the partial sums under consideration is similar to that in the case of independent observations (see [4]):

$$U_n^N(t) \xrightarrow{d} U^N(t) = \sum_{k_1=1}^{N} \cdots \sum_{k_m=1}^{N} f_{k_1 \ldots k_m} t^{m/2} \prod_{j=1}^{\infty} H_{v_j(k_1,\ldots,k_m)}(t^{-1/2} w_j(t)).$$

From here we see that, as $n \to \infty$,

$$\bigl(U_n^N(t_1), \ldots, U_n^N(t_q)\bigr) \xrightarrow{d} (U^N(t_1), \ldots, U^N(t_q)),$$

where the dependent Wiener processes $\{w_j(t)\}$ have joint Gaussian distributions with covariances defined in (6). This assertion is a direct consequence of the multivariate central limit theorem for stationary sequences of random variables with mixing since, in this case, we can use the classical Cramér–Wold method (see [12]).

II. Density of the distributions of U-processes. We need to prove that

$$\lim_{\Delta \to 0} \limsup_{n \to \infty} P(\sup_t |U_n(t + \Delta) - U_n(t)| > c) = 0.$$

First we prove the analogous assertion for U_n^N:

$$\lim_{\Delta\to 0}\limsup_{n\to\infty} P\Big(\sup_t\big|U_n^N(t+\Delta)-U_n^N(t)\big|>c\Big)=0.$$

If the density property is valid for every stochastic process (distribution) from a finite collection then it holds for the sum of these processes. Therefore, we study this property separately for every stochastic process which forms the truncated statistic U_n^N. Note that the statistic

$$U_n^N(t)=n^{-m/2}\sum_{k_1=0}^{N}\cdots\sum_{k_m=0}^{N}\sum_{1\le i_1\ne}\cdots\sum_{\ne i_m\le[nt]} f_{k_1\dots k_m}e_{k_1}(X_{i_1})\dots e_{k_m}(X_{i_m})$$

consists of finitely many sums of the form

$$S_m(t)\equiv S_n(m,k_1,\dots,k_m)(t):=n^{-m/2}\sum_{1\le i_1\ne}\cdots\sum_{\ne i_m\le[nt]} e_{k_1}(X_{i_1})\dots e_{k_m}(X_{i_m}),$$

where the indices $1\le k_1,\dots,k_m\le N$ are arbitrary.

Notice that $\mathbf{E}\sup_{\mathbf{t}}|\mathbf{S_m(t)}|<\mathbf{C(m)}<\infty$. Indeed, adding and subtracting the diagonal elements of the sum $S_m(t)$, we obtain a finite linear combination of the following products:

$$S_1^{(1)}(t)\dots S_1^{(j)}(t),\quad 1\le j\le m,$$

where $S_1^{(j)}(t)=n^{-1/2}\sum_{i\le[nt]}e_j(X_i)$. Using Theorem 1, we have

$$\mathbf{E}\sup_t\big|S_1^{(1)}(t)\dots S_1^{(j)}(t)\big|\le\Big(\mathbf{E}\sup_t\big|S_1^{(1)}(t)\big|^j\dots\mathbf{E}\sup_t\big|S_1^{(j)}(t)\big|^j\Big)^{1/j}<\infty.$$

Prove the density for the family of the distributions of $S_n(m,\dots)(t)$ by induction on m. For $m=1$, the statement is contained in [12].

Assume that the statement was proved for all $m\le l$. Now, prove this assertion for $m=l+1$. Use the identity

$$n^{-(l+1)/2}\sum_{1\le i_1\ne}\cdots\sum_{\ne i_{l+1}\le[nt]} e_{k_1}(X_{i_1})\dots e_{k_l}(X_{i_l})e_{k_{l+1}}(X_{i_{l+1}})$$

$$=n^{-l/2}\Big(\sum_{1\le i_1\ne}\cdots\sum_{\ne i_l\le[nt]} e_{k_1}(X_{i_1})\dots e_{k_l}(X_{i_l})\Big)n^{-1/2}\sum_{i_{l+1}=1}^{[nt]}e_{k_{l+1}}(X_{i_{l+1}})$$

$$-n^{-1}\sum_{i=1}^{[nt]}e_{k_1}(X_i)e_{k_{l+1}}(X_i)n^{-(l-1)/2}\sum_{1\le i_2\ne}\cdots\sum_{\ne i_l\le[nt]}e_{k_2}(X_{i_2})\dots e_{k_l}(X_{i_l})-\cdots$$

$$-n^{-1}\sum_{i=1}^{[nt]}e_{k_l}(X_i)e_{k_{l+1}}(X_i)n^{-(l-1)/2}\sum_{1\le i_1\ne}\cdots\sum_{\ne i_{l-1}\le[nt]}e_{k_1}(X_{i_1})\dots e_{k_{l-1}}(X_{i_{l-1}}).$$

It is more convenient to rewrite this relation as follows:

$$S_n(l+1,k_1,\dots,k_{l+1})(t) = S_n(l,k_1,\dots,k_l)(t)S_n(1,k_{l+1})(t)$$

$$-S_n(l-1,k_2,\dots,k_l)\theta_n^1(t) - \dots - S_n(l-1,k_1,\dots,k_{l-1})\theta_n^l(t), \tag{8}$$

where $\theta_n^j(t) = n^{-1}\sum_{i=1}^{[nt]} e_{k_j}(X_i)e_{k_{l+1}}(X_i)$ and by definition $S_0(t) \equiv 1$.

First we note that, using (8), one can establish by induction on l the uniform stochastic boundedness of the stochastic processes $S_n(l,k_1,\dots,k_l)(t)$, i.e., we mean the relation

$$\lim_{K\to\infty} \mathbf{P}(\sup_{0\le t\le 1} |S_n(l,k_1,\dots,k_l)(t)| > K) = 0.$$

The induction base ($l = 1$) is straightforward from Theorem 1 and the simple relation (the law of large numbers for weakly dependent identically distributed random variables)

$$\sup_{0\le t\le 1}\left|\theta_n^j(t)\right| \le n^{-1}\sum_{i=1}^{n} |e_{k_j}(X_i)e_{k_{l+1}}(X_i)| \xrightarrow{p} \mathbf{E}|e_{k_j}(X_1)e_{k_{l+1}}(X_1)| < \infty \quad \text{as } n\to\infty. \tag{9}$$

Now, prove the density of the distributions (for all n) of all pairwise products of the stochastic processes in (8) by induction on l once again. For example, consider the product $S_n(l,\cdot)(t)S_n(1,\cdot)(t)$ (the other pairwise products in (8) are studied similarly). In view of the elementary representation

$$S_l(t+\Delta)S_1(t+\Delta) - S_l(t)S_1(t) = (S_l(t+\Delta) - S_l(t))S_1(t+\Delta) + S_l(t)(S_1(t+\Delta) - S_1(t)),$$

we can conclude that the density property for the product of two processes is valid if this property is fulfilled for each of these two processes, and moreover, each of these processes is stochastically bounded uniformly in $t \in [0,1]$ which was proved. The induction base follows from Theorem 1. The induction step from l to $l+1$ is immediate from (8) and the fact that the stochastic processes $n^{1/2}\theta_n^j(t)$ are the classical partial sum processes (the so-called random broken lines) for which, under the above conditions, the density of their distributions has been proved (see [12]).

Estimate the first moment of the uniform norm of the tail statistic (multiple series):

$$\mathbf{E}\sup_t\left|U_n(t) - U_n^N(t)\right| = \mathbf{E}\sup_t\Big|\sum_{\max(k_j)\ge N+1}\dots\sum f_{k_1\dots k_m}S_n(m,\dots)(t)\Big|$$

$$\le \sum_{\max(k_j)\ge N+1}\dots\sum |f_{k_1\dots k_m}|\mathbf{E}\sup_t |S_n(m,\dots)(t)| \le C(m)\sum_{\max(k_j)=N+1}^{\infty} |f_{k_1\dots k_m}|.$$

Thus, for every $\varepsilon > 0$, by the relation $\sum_{k_1=1}^{\infty}\dots\sum_{k_m=1}^{\infty}|f_{k_1\dots k_m}| < \infty$, there exists a natural N such that

$$\sum_{\max(k_j)\ge N+1}\dots\sum |f_{k_1\dots k_m}| < \varepsilon.$$

Then the following is valid:

$$P(\sup_t |U_n(t+\Delta) - U_n(t)| > c)$$
$$\le P\big(\sup_t \big|U_n^N(t+\Delta) - U_n^N(t)\big| + 2\sup_t \big|U_n(t) - U_n^N(t)\big| > c\big)$$
$$\le P\big(\sup_t \big|U_n^N(t+\Delta) - U_n^N(t)\big| > c/3\big) + P\big(\sup_t \big|U_n(t) - U_n^N(t)\big| > c/3\big)$$
$$\le P\big(\sup_t \big|U_n^N(t+\Delta) - U_n^N(t)\big| > c/3\big) + 3c^{-1}\mathbf{E}\sup_t \big|U_n(t) - U_n^N(t)\big|$$
$$\le P\big(\sup_t \big|U_n^N(t+\Delta) - U_n^N(t)\big| > c/3\big) + 3c^{-1}\varepsilon C(m).$$

Hence,

$$\lim_{\Delta\to 0}\limsup_{n\to\infty} P(\sup_t |U_n(t+\Delta) - U_n(t)| > c) \le 3c^{-1}\varepsilon C(m).$$

Since ε is arbitrary, we conclude that

$$\lim_{\Delta\to 0}\limsup_{n\to\infty} P(\sup_t |U_n(t+\Delta) - U_n(t)| > c) = 0.$$

We should only prove a.s. continuity of the limit stochastic process $U(t)$. Since, for all j, the relation

$$\mathbf{E}|w_j(t+\delta) - w_j(t)|^4 = \mathbf{E}|w_j(\delta)|^4 = C\delta^2$$

is valid then, for all $l \le m$, we obtain

$$\mathbf{E}\big|w_j^l(t+\delta) - w_j^l(t)\big|^4 = \mathbf{E}|w_j(t+\delta) - w_j(t)|^4\big|w_j^{l-1}(t+\delta) + \cdots + w_j^{l-1}(t)\big|^4$$
$$\le \big(\mathbf{E}|w_j(t+\delta) - w_j(t)|^8\mathbf{E}\big|w_j^{l-1}(t+\delta) + \cdots + w_j^{l-1}(t)\big|^8\big)^{1/2} \le C_1(\delta^4)^{1/2} = C_1\delta^2. \quad (10)$$

It is clear that multiplication of the stochastic process $w_j^l(t)$ by any nonrandom Lipschitz function cannot essentially change (10) which is true up to a constant factor on the right-hand side of this inequality. So, multiplying these stochastic processes by t^k, $t \in [0,1]$ (in the representation of U-processes under consideration, there are only the factors t^k with integer $k \ge 0$), we obtain some new stochastic processes that satisfy the above-mentioned estimate in (10). Summing the values $t^k w_j^l(t)$ with the corresponding scalar coefficients, we obtain the expression $t^{v_j(k_1,\dots,k_m)/2}H_{v_j(k_1,\dots,k_m)}(t^{-1/2}w_j(t))$. Since, for all $j, k_1, \dots, k_m$, the order of polynomial $v_j(k_1,\dots,k_m)$ does not exceed m, there is a constant C such that

$$\mathbf{E}|(t+\delta)^{v_j(k_1,\dots,k_m)/2}H_{v_j(k_1,\dots,k_m)}((t+\delta)^{-1/2}w_j(t+\delta))$$
$$-t^{v_j(k_1,\dots,k_m)/2}H_{v_j(k_1,\dots,k_m)}(t^{-1/2}w_j(t))|^4 \le C_2\delta^2.$$

For the product of finitely many these processes, we then have

$$Y_{k_1,\dots,k_m}(t) = t^{m/2}\prod_{j=1}^{\infty} H_{v_j(k_1,\dots,k_m)}(t^{-1/2}w_j(t)). \quad (10)$$

By finiteness of the moments

$$\mathbf{E}|t^{v_j(k_1,\dots,k_m)/2}H_{v_j(k_1,\dots,k_m)}(t^{-1/2}w_j(t))|^l, \quad l \le m,$$

and Hölder's inequality, we have

$$\mathbf{E}|Y_{k_1,\dots,k_m}(t+\delta) - Y_{k_1,\dots,k_m}(t)|^4 \le C_3\delta^2.$$

Notice that, in the infinite product in (10), there is only a finite collection of the factors that do not equal to 1 since, for $j > \max k_i$, all factors of this product are equal to 1 (i.e., to the first Hermit polynomial). Denote

$$\Delta_0 = \mathbf{E}|U(t+\delta) - U(t)|^4 = \mathbf{E}\left|\sum_{k_1=1}^{\infty} \cdots \sum_{k_m=1}^{\infty} f_{k_1\dots k_m}(Y_{k_1,\dots,k_m}(t+\delta) - Y_{k_1,\dots,k_m}(t))\right|^4.$$

For convenience, we replace the multi-index $(k_1, \dots, k_m)$ with the symbol $\tilde{k}$ and denote

$$\triangle Y_{\tilde{k}} := Y_{\tilde{k}}(t+\delta) - Y_{\tilde{k}}(t).$$

Applying Hölder's inequality, we finally obtain

$$\Delta_0 \le \sum_{\tilde{k_1}}\sum_{\tilde{k_2}}\sum_{\tilde{k_3}}\sum_{\tilde{k_4}} |f_{\tilde{k_1}}f_{\tilde{k_2}}f_{\tilde{k_3}}f_{\tilde{k_4}}|\mathbf{E}|\triangle Y_{\tilde{k_1}}\triangle Y_{\tilde{k_2}}\triangle Y_{\tilde{k_3}}\triangle Y_{\tilde{k_4}}|$$

$$\le \sum_{\tilde{k_1}}\sum_{\tilde{k_2}}\sum_{\tilde{k_3}}\sum_{\tilde{k_4}} |f_{\tilde{k_1}}f_{\tilde{k_2}}f_{\tilde{k_3}}f_{\tilde{k_4}}|(\mathbf{E}|\triangle Y_{\tilde{k_1}}|^4\mathbf{E}|\triangle Y_{\tilde{k_2}}|^4\mathbf{E}|\triangle Y_{\tilde{k_3}}|^4\mathbf{E}|\triangle Y_{\tilde{k_4}}|^4)^{1/4}$$

$$\le \sum_{\tilde{k_1}}\sum_{\tilde{k_2}}\sum_{\tilde{k_3}}\sum_{\tilde{k_4}} |f_{\tilde{k_1}}f_{\tilde{k_2}}f_{\tilde{k_3}}f_{\tilde{k_4}}|K\delta^2 = K\delta^2\left(\sum_{k_1=1}^{\infty} \cdots \sum_{k_m=1}^{\infty} |f_{k_1\dots k_m}|\right)^4.$$

So, by the classical Kolmogorov's criterion (see [1]), the a.s. continuity of $U(t)$ is immediate from here.

The theorem is proven.

13 Gaussian approximation to the partial sum processes of moving averages

1. Introduction

We study approximation to the partial sum processes of moving averages of independent identically distributed (i.i.d.) observations to some Gaussian processes. The approximation of this kind is studied for a long time (for example, see [35, 64]) and is justified by the explicit applied character of the models under consideration. Note that the moving averages mentioned below and based on a sequence of i.i.d. random variables comprise a constructively defined sequence of stationary connected random variables. In general, the inter dependence of these random variables may be strong enough. In particular, for moving averages, the classical strong (or uniformly strong) mixing condition may fail here (see [64, 114]).

In [76] some approximation was studied to the above-mentioned partial sum processes by a fractional Brownian motions with Hurst parameters $H > 1/2$, where some rates of convergence were obtained in Donsker's and Strassen's invariance principles. In this paper we obtain analogous convergence rates for all $0 < H < 1$ in the above-mentioned invariance principles. Moreover, in Strassen's invariance principle we slightly extend the class of limit Gaussian processes. Also, for $H < 1/2$, in Donsker's invariance principle we weaken the moment restrictions on the initial sequence of i.i.d. random variables in [35] for convergence in distribution of the normalized partial sum processes to a fractional Brownian motion.

Let $\{\xi_k; k \in \mathbb{Z}\}$ be i.i.d. random variables with mean zero and variance one, where $\mathbb{Z}$ is the set of all integer numbers. Consider the sequence of random variables $\{X_j; j \in \mathbb{Z}\}$ defined by the formula

$$X_j = \sum_{k=-\infty}^{\infty} a_{j-k}\xi_k; \tag{1}$$

these random variables are called *moving averages* of the initial sequence $\{\xi_k; k \in \mathbb{Z}\}$ (see [113]). The following well-known condition guarantees convergence with probability 1 of the series on the right-hand side of (1):

$$0 < \sum_{k\in\mathbb{Z}} a_k^2 < \infty. \tag{2}$$

In the sequel we assume (2) to be fulfilled. Define the partial sum process of moving averages (1):

$$S_0 = 0, \quad S_n = \sum_{i=1}^{n} X_i, \ n = 1, 2, \dots.$$

This random process will be approximated by the following Gaussian process on the positive half-line $[0,\infty)$:

$$B_H(t)=\sigma L_H^{-1/2}\Big(\int_0^{\infty}((t+s)^{H-1/2}g(t+s)-s^{H-1/2}g(s))\,d\widetilde{W}(s)+\int_0^{t}(t-s)^{H-1/2}g(t-s)\,dW(s)\Big), \tag{3}$$

where

$$\sigma^2=\mathbf{D}B_H(1),\quad L_H=\frac{1}{2H}+\int_0^{\infty}((1+s)^{H-1/2}-s^{H-1/2})^2\,ds,$$

$\widetilde{W}(s)$ and $W(s)$ are two independent versions of the standard Wiener process, $0<H<1$, and g is a slowly varying function with the following properties:

(a) $g(x)$ is differentiable on the half-line $[0,\infty)$; moreover,

$$g'(x)=o(g(x)/x)\quad\text{as}\quad x\to\infty;$$

(b) $g(x)$ and $g'(x)$ are monotone and $g(x)$ is sign-preserving on the half-line $[0,\infty)$.

For example, it is easy to verify that every (positive or negative) power of the m-iterated logarithm for each natural m satisfies these conditions.

Note that if $g\equiv 1$ in (3) then the random process $B_H(t)$ is a fractional Brownian motion (see [88]), i.e., a centered Gaussian process with covariance function

$$R(t,s)=\frac{\sigma^2}{2}(t^{2H}+s^{2H}-|t-s|^{2H}).$$

In the sequel we will denote this process by B_H^0. It is easy to see that, in the case $H=1/2$, we deal with a Wiener process. Recall the well-known property (H-homogeneity) of a fractional Brownian motion (see [88]): For every $\lambda>0$, the finite-dimensional distributions of the random processes $\{B_H^0(\lambda t)\}$ and $\{\lambda^H B_H^0(t)\}$ coincide. Moreover, the random process B_H^0 has stationary increments. These properties of B_H^0 allow us to relate it to a class of objects called *fractals* whose every "distinguished" fragment (say, part of a trajectory) is similar in a sense to the whole object.

Denote

$$A_m=\quad a_0+\dots+a_m \text{ if } m\ge 0 \text{ and } A_{-1}=0, -(a_{m+1}+\dots+a_{-1}) \text{ if } m<-1. \tag{4}$$

For each fixed $n\ge 0$ and some $\alpha>2$ independent of n, we assume that, as $|l|\to\infty$, the following condition holds:

$$|A_{n+l}-A_l|=O(|l|^{-1/\alpha}), \tag{5}$$

implying in view of (2) that

$$\sum_{l\in\mathbb{Z}}(A_{n+l}-A_l)^2<\infty. \tag{6}$$

We note that the series in (6) coincides with $\mathbf{D}S_n$ (for details, see Subsection 2.2). If, as $l\to\infty$, the quantity $|A_{n+l}-A_l|$ tends monotonically to zero then (5) follows from (6) because, in this case, $|A_{n+l}-A_l|=o(|l|^{-1/2})$.

R e m a r k 1. In the sequel we will speak about the *nontrivial* (or *substantial*) estimates for closeness of the processes S_n and $B_H(n)$ if, as $n \to \infty$, the following holds:

$$S_n - B_H(n) = o(\sqrt{\mathbf{D}B_H(n)}) \quad \text{a. s.}$$

In other words, the substantial estimates for closeness should be essentially less than the random variables under consideration. But in this case (see the proof of Theorem 1 below), absolutely the same statement is valid for the Gaussian analog of the process S_n, when the random variables ξ_k in (1) are the standard Gaussian (moreover, the numerical value $\mathbf{D}S_n$ will be the same), i.e., in the above-mentioned asymptotic relation, we may consider S_n as a Gaussian sequence. It follows that, first, as $n \to \infty$, the distributions of the random variables $S_n/\sqrt{\mathbf{D}B_H(n)}$ converge weakly to the standard normal law; and second, for each n, the random variable $S_n/\sqrt{\mathbf{D}S_n}$ has the standard normal distribution. Whence we obtain the following *necessary* condition for the above-mentioned approximation to be valid: $\mathbf{D}S_n \sim \mathbf{D}B_H(n)$ as $n \to \infty$.

Proposition 1 *The following is valid:*

$$\mathbf{D}B_H(n) \sim \sigma^2 g^2(n) n^{2H} \quad \textit{as} \;\; n \to \infty.$$

By Proposition 1 we can rewrite the above-mentioned necessary condition as follows:

$$\mathbf{D}S_n = \sum_{k \in \mathbb{Z}} (A_{n-k} - A_{-k})^2 \sim \sigma^2 g^2(n) n^{2H} \quad \text{as} \;\; n \to \infty. \tag{H}$$

In addition to the sequence $\{A_m\}$ in (4), we need also the following notation connected with the initial coefficients $\{a_i\}$:

$$\Delta_n^{(1)} = \sum_{m=0}^{\infty} \big(A_{n+m} - A_m - \sigma L_H^{-1/2}(n+m+1)^{H-1/2} g(n+m+1)$$
$$+\sigma L_H^{-1/2}(m+1)^{H-1/2} g(m+1)\big)^2;$$
$$\Delta_n^{(2)} = \sum_{m=1}^{n} \big(A_{n-m} - A_{-m} - \sigma L_H^{-1/2}(n-m+1)^{H-1/2} g(n-m+1)\big)^2;$$
$$\Delta_n^{(3)} = \sum_{m>n} (A_{n-m} - A_{-m})^2; \quad \Delta_n = \Delta_n^{(1)} + \Delta_n^{(2)} + \Delta_n^{(3)};$$
$$\Delta_{\alpha,n}^{+} = \sum_{m=1}^{\infty} \max\{m^{1/\alpha}, n^{1/\alpha}\} |a_{n+m} - a_m|;$$
$$\Delta_{\alpha,n}^{-} = \sum_{m=1}^{\infty} \max\{m^{1/\alpha}, n^{1/\alpha}\} |a_{n-m} - a_{-m}|; \quad \Delta_{\alpha,n} = \Delta_{\alpha,n}^{+} + \Delta_{\alpha,n}^{-}.$$

In the sequel we will interpret a relation of the type $\psi(t) = O(f(t))$ a. s., where $\psi(t)$ is a random process and $f(t)$ is a certain nonrandom nonnegative function, as fulfillment of the inequality $|\psi(t)| \le Cf(t)$ for all $t \ge t_0$ (where, in general, t_0 depends on an elementary event) and some nonrandom positive constant C that is called the *O-symbol constant*; moreover, this constant may be absolute as well as dependant on some parameters of the problem under consideration. We will indicate the type of dependence in stating the claims to follow.

Theorem 1. *Let* (5) *be valid and* $\mathbf{E}|\xi_0|^\alpha < \infty$ *for some* $\alpha > 2$. *Then, as* $t \to \infty$, *the following holds:*

$$S_{[t]} - B_H(t) = \varepsilon(t)\Delta_{\alpha,[t]} + O\Big(\sqrt{\Delta_{[t]}\log t} + \sigma L_H^{-1/2}\big(\Upsilon_H^{1/2} + \Phi_H^{1/2} C_H\big)\sqrt{\log t}\Big) \quad a.\ s.,$$

where $\lim_{t\to\infty}\varepsilon(t) = 0$ *a. s.; moreover, if the distribution of the random variable* ξ_0 *is fixed then this convergence is uniform over all other parameters of the problem;*

$$\Upsilon_H = \int_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s+1))^2\,ds + \int_0^\infty s^{2H-1}(g(s+1)-g(s))^2\,ds,$$

$$\Phi_H = \frac{1}{4H}(g(1))^2 + \int_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s+1))^2\,ds + \int_0^\infty s^{2H-1}(g'(s))^2\,ds,$$

if g increases;

$$\Upsilon_H = \int_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s))^2\,ds + \int_0^\infty s^{2H-1}(g(s+1)-g(s))^2\,ds,$$

$$\Phi_H = \frac{1}{4H}(g(0))^2 + (g(0))^2\int_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2\,ds + \int_0^\infty s^{2H-1}(g'(s))^2\,ds,$$

if g decreases; $C_H = \sum_{k=1}^\infty 2^{-kH}k^{1/2}$; *moreover, the O-symbol constant is absolute.*

R e m a r k 2. In the case $H > 1/2$ and $g \equiv 1$ an analog of Theorem 1 is proven in [76], where there is no specification of the O-symbol constant.

In the statements below the O-symbol constant, in general, depends on the sequence $\{a_k; k \in \mathbb{Z}\}$.

Proposition 2. *Let the sequence* $\{a_k; k \in \mathbb{Z}\}$ *satisfy the following conditions:*

$$a_k = 0, \quad k < 0;$$

$$a_k - \sigma L_H^{-1/2}(k+1)^{H-1/2}g(k+1) + \sigma L_H^{-1/2}k^{H-1/2}g(k) = O(k^{\gamma-3/2}g(k)), \quad \gamma < H; \tag{7}$$

$$A_k - \sigma L_H^{-1/2}(k+1)^{H-1/2}g(k+1) = O(k^{\beta-1/2}g(k)), \quad 0 < \beta < H,$$

as $k \to \infty$. *Then, as* $n \to \infty$,

(i) *If* $H > 1/2$ *then* $\Delta_{\alpha,n} = O(g(n)n^{H-1/2+1/\alpha})$; $\Delta_n = O(g^2(n)n^{\max\{2\beta,2\gamma\}})$;

(ii) *If* $H < 1/2$ *then* $\Delta_{\alpha,n} = O(n^{1/\alpha})$; $\Delta_n = O(g^2(n)n^{\max\{2\beta,2\gamma\}})$;

(iii) *if* $H = 1/2$ *and* $\sum_{k=0}^{n} |a_k| = O(g(n))$ *for increasing* g *or* $\sum_{k=0}^{\infty} |a_k| < \infty$ *for decreasing* g *then* $\Delta_{\alpha,n} = O(g(n)n^{1/\alpha})$ *or* $\Delta_{\alpha,n} = O(n^{1/\alpha})$ *for increasing or decreasing* g*;* $\Delta_n = O(g^2(n)n^{\max\{2\beta,2\gamma\}})$.

Corollary 1. *Let the conditions of Theorem* 1 *and Proposition* 2 *be fulfilled. Then, as* $t \to \infty$*,*

(i) *In the case* $H > 1/2$*,*

$$S_{[t]} - B_H(t) = o(g(t)t^{H-1/2+1/\alpha}) + O(t^{\max\{\beta,\gamma\}} g(t)\sqrt{\log t}) \quad a.\ s.;$$

(ii) *In the case* $H < 1/2$*,*

$$S_{[t]} - B_H(t) = o(t^{1/\alpha}) + O(t^{\max\{\beta,\gamma\}} g(t)\sqrt{\log t}) \quad a.\ s.;$$

(iii) *In the case* $H = 1/2$*, if* g *increases then*

$$S_{[t]} - B_H(t) = o(g(t)t^{1/\alpha}) + O(t^{\max\{\beta,\gamma\}} g(t)\sqrt{\log t}) \quad a.\ s.;$$

(iv) *In the case* $H = 1/2$*, if* g *decreases then*

$$S_{[t]} - B_H(t) = o(t^{1/\alpha}) + O(t^{\max\{\beta,\gamma\}} g(t)\sqrt{\log t}) \quad a.\ s.$$

Remark 3 Condition (7) implies Condition (5). In items (i), (iii), and (iv) of Corollary 1 we have nontrivial estimates and Condition (H) is valid. In item (ii) we have substantial estimates in the case $\alpha \geq 1/H$ for $g(t) \asymp 1$ as $t \to \infty$. In the case when $g(t) \to 0$ as $t \to \infty$, these estimates are substantial for $\alpha > 1/H$.

In the statements below we prove Donsker's invariance principle for the processes under consideration. We also estimate the corresponding convergence rate (regarding other limit theorems for the sums of moving averages, see [63] for example).

Define the normalized partial sum process on the interval $[0,1]$:

$$Z_{n,H}(t) = \frac{S_{[nt]}}{n^H}.$$

Theorem 2. *Let* $\mathbf{D}S_n \sim \sigma^2 n^{2H}$ *as* $n \to \infty$*. Moreover, let* $\mathbf{E}|\xi_0|^\alpha < \infty$*, where* $\alpha \geq 2$ *and* $\alpha H > 1$*. Then, as* $n \to \infty$*, the distributions of the random processes* $Z_{n,H}(t)$ C*-converge in* $D[0,1]$ *to the distribution of a fractional Brownian motion* $B_H^0(t)$*.*

Recall that C-convergence in $D[0,1]$ is weak convergence of distributions of measurable functionals on $D[0,1]$ (in the Skorokhod topology) which are continuous in the uniform topology only at the points of $C[0,1]$ (see [28]).

Remark 4. In the case $H > 1/2$ the claim is proved in [76]. Necessity of the condition $\mathbf{D}S_n \sim \sigma^2 n^{2H}$ in Theorem 2 is proved in [35], where, in addition, C-convergence is proved in the case $H < 1/2$ under more restrictive moment conditions than those in Theorem 2.

Remark 5 Under the conditions of Corollary 1 in the case $H < 1/2$ and $\alpha = 1/H$, it is possible to define the processes $Z_{n,H}(t)$ on a common probability space together with a fractional Brownian motion so that, as $n \to \infty$,

$$\sup_{t\in[0,1]} \left| Z_{n,H}(t) - B_H^0(nt)/n^H \right| = o(1) \quad \text{a.s.}$$

Since, for each n, the random processes $B_H^0(nt)/n^H$ and $B_H^0(t)$ coincide in distribution, we deduce from this relation the claim of Theorem 2 for the particular case mentioned above for $H < 1/2$ and $\alpha = 1/H$.

Put $\delta_n = \sup_{t\in[0,1]} \Delta_{[nt]}$.

Theorem 3 *Assume the following:*

(i) *$H \geq 1/2$, $\mathbf{E}|\xi_0|^\alpha < \infty$ for some $\alpha > 2$, and $a_i = 0$ for all $i < 0$. Moreover, let, for all $i \geq N \geq 1$, the inequalities $a_i \geq a_{i+1} \geq 0$, $a_i \leq Ci^{H-3/2}$, and $\sum_{i=0}^n |a_i| \leq C_1(n+1)^{H-1/2}$ be valid for $n \geq 0$. Then there exists a probability space such that*

$$\mathbf{P}\big(\sup_{t\in[0,1]} \big|Z_{n,H}(t) - B_H^0(t)\big| \geq n^{-H}\sqrt{(H+1)\log n}(8\sqrt{2\Upsilon_H^0} + \sqrt{2\delta_n} + 2\sqrt{2}\sigma C_H)$$

$$+n^{-\frac{\alpha-2}{2(\alpha+1)}}(C_1(2^{H+3/2}+1) + C(1-2^{H-3/2+1/\alpha})^{-2})\big)$$

$$\leq 6n^{-H} + C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}}(2+\alpha/(\alpha-1))$$

for all $n \geq N$.

(ii) *Let $H \leq 1/2$ and $\mathbf{E}|\xi_0|^\alpha < \infty$ for some $\alpha > 1/H$. Moreover, let $a_i = 0$ for all $i < 0$, $a_i \leq a_{i+1} \leq 0$ and $|a_i| \leq Ci^{H-3/2}$ for all $i \geq N \geq 1$, and also $\sum_{i=0}^n |a_i| \leq C_1$ for $n \geq 0$. Then there exists a probability space such that*

$$\mathbf{P}\big(\sup_{t\in[0,1]} \big|Z_{n,H}(t) - B_H^0(t)\big| \geq n^{-H}\sqrt{(H+1)\log n}(8\sqrt{2\Upsilon_H^0} + \sqrt{2\delta_n} + 2\sqrt{2}\sigma C_H)$$

$$+5C_1 n^{-\frac{\alpha H-1}{\alpha+1}} + Cn^{-\frac{\alpha-2}{2(\alpha+1)}}(1-2^{H-3/2+1/\alpha})^{-2}\big)$$

$$\leq 6n^{-H} + 2C_\xi n^{-\frac{\alpha H-1}{\alpha+1}} + C_\xi \alpha(\alpha-1)^{-1} n^{-\frac{\alpha-2}{2(\alpha+1)}}$$

for all $n \geq N$, where C_ξ is a constant depending only on the distribution of ξ_0,

$$C_H = \sum_{k=1}^\infty 2^{-kH}k^{1/2}, \quad \Upsilon_H^0 = \int_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2\, ds.$$

Note that if $a_i = 0$ for all $i < 0$ and, in addition, $a_0 = \sigma L_H^{-1/2}$ and $a_i = \sigma L_H^{-1/2}((i+1)^{H-1/2} - i^{H-1/2})$ for $i > 0$ then $\Delta_n = \delta_n = 0$. We now obtain some estimates for δ_n under more general restrictions on the coefficients $\{a_i, i \in \mathbb{Z}\}$. These estimates are valid for all $H \in (0,1)$.

Proposition 3 *Let $a_i = 0$ for all $i < 0$. Then, as $n \to \infty$,*

(i) *If*

$$\sum_{k=0}^\infty \big(A_k - \sigma L_H^{-1/2}(k+1)^{H-1/2}\big)^2 < \infty$$

then

$$\delta_n = O(1);$$

(ii) *If*
$$\left|A_n - \sigma L_H^{-1/2}(n+1)^{H-1/2}\right| = O(n^{\beta-1/2}), \quad 0<\beta<H,$$
and
$$\left|a_n - \sigma L_H^{-1/2}((n+1)^{H-1/2} - n^{H-1/2})\right| = O(n^{\gamma-3/2}), \quad \gamma<H,$$
then
$$\delta_n = O(n^{\max\{2\gamma,2\beta\}}).$$

R e m a r k 6. Under the conditions of items (i) or (ii) of Proposition 3 the relation $n^{-H}\sqrt{\delta_n \log n} \to 0$ is valid as $n\to\infty$. In this case we obtain a nontrivial convergence rate in Theorem 3.

R e m a r k 7. Let $\sum_{k\in\mathbb{Z}}|a_k|<\infty$. Then it is easy to deduce the following identity:
$$n^{-1}\mathbf{D}S_n = n^{-1}\sum_{k\in\mathbb{Z}}(A_{n-k}-A_{-k})^2 = n^{-1}\sum_{k\in\mathbb{Z}}(a_{k+1}+\cdots+a_{k+n})^2$$
$$= \sum_{k\in\mathbb{Z}}a_k^2 + 2\sum_{j=1}^{n}\sum_{k\in\mathbb{Z}}a_{k+j}a_k - 2n^{-1}\sum_{j=1}^{n}j\sum_{k\in\mathbb{Z}}a_{k+j}a_k.$$
Moreover, by Kronecker's lemma (see [93, p. 328]) the rightmost double sum of this identity vanishes as $n\to\infty$. Thus, assuming the absolute summability of $\{a_k\}$, we have
$$\lim_{n\to\infty} n^{-1}\mathbf{D}S_n = \sigma^2 = \Big(\sum_{k\in\mathbb{Z}}a_k\Big)^2.$$
Moreover, if $\sigma^2>0$ then condition (H) is valid for $H=1/2$. Hence, under the conditions of Theorem 2 (i.e., if $\mathbf{E}|\xi_0|^\alpha<\infty$ for some $\alpha>2$) for *absolutely summable* coefficients $\{a_i\}$ the above-mentioned C-convergence to a Wiener process is valid if only $\sum_{k\in\mathbb{Z}}a_k\neq 0$. Note that, under the finiteness of the fourth moment of the random variable ξ_0, an analogous proposition follows from [35].

For example, we can apply these arguments to the simplest particular case in which $a_i=0$ for all $i<0$ and $i>m$, where m is a fixed natural number (in this case, $\{X_j; j\ge 1\}$ are m-dependent random variables), and $\sigma=\sum_{i=0}^m a_i\neq 0$.

R e m a r k 8. If the second moment of ξ_0 exists then in [12, pp. 255 and 264] there are given some sufficient conditions guaranteeing the C-convergence of $Z_{n,1/2}$ to a Wiener process. In particular, the article [12] contains the following additional restriction on the coefficients $\{a_k\}$:
$$\sum_{l=1}^{\infty}\Big(\sum_{|i|>l}a_i^2\Big)^{1/2}<\infty. \tag{8}$$
It is clear that this implies the absolute summability of $\{a_i\}$ of Remark 7. Further, (8) was slightly weaken as follows:
$$\sum_{l=1}^{\infty}\Big(\Big(\sum_{i>l}a_i\Big)^2+\Big(\sum_{i>l}a_{-i}\Big)^2\Big)<\infty.$$

In particular, the last condition implies the *conditional* convergence of $\sum a_i$. Note that, in this case, the following holds:

$$\lim_{n\to\infty} n^{-1}\mathbf{D}S_n = \Big(\sum_{k\in\mathbb{Z}} a_k\Big)^2,$$

which is actually assumed in Theorem 2. As an example we may consider the coefficients $a_i = 0$ for $i \le 0$ and $a_i = (-1)^i/i$ for $i \ge 1$.

However, if the series tail $\sum_{|i|\ge k} a_i$ decreases slowly enough then the above-mentioned two conditions of [12] may not be fulfilled. Say, the sequence $a_i = \sigma_0(i^{-1/2} - (i+1)^{-1/2})$ for $i > 0$, $a_0 = \sigma_0$, and $a_i = 0$ for $i < 0$ does not satisfy these conditions, but it enjoys the conditions of Remark 7, i.e., for these coefficients $\{a_i\}$ the above-mentioned C-convergence to a Wiener process is valid (but under somewhat stronger moment restrictions than those in [12]).

2. Proof of the Main Results.

2.1. Proof of Proposition 1. We consider only the case of g increasing. The case of g decreasing is settled by analogy. We have

$$\mathbf{D}B_H(n) = \sigma^2 L_H^{-1} n^{2H} g^2(n)\bigg(\int_0^1 s^{2H-1} g^2(ns)/g^2(n)\,ds$$

$$+\int_0^\infty ((1+s)^{H-1/2} g(n(1+s))/g(n) - s^{H-1/2} g(ns)/g(n))^2\,ds\bigg). \tag{9}$$

Consider the first summand on the right-hand side of (9). By the Lebesgue dominated convergence theorem,

$$\int_0^1 s^{2H-1} g^2(ns)/g^2(n)\,ds \to 1/(2H) \quad \text{as} \quad n\to\infty.$$

We now split the second integral on the right-hand side of (9) into the following two: $\int_0^\infty = \int_0^1 + \int_1^\infty$. Estimate the first integral:

$$\int_0^1 ((1+s)^{H-1/2} g(n(1+s))/g(n) - s^{H-1/2} g(ns)/g(n))^2\,ds$$

$$\le 2\int_0^1 ((1+s)^{2H-1} g^2(2n)/g^2(n) + s^{2H-1})\,ds;$$

moreover, the sequence $\{g^2(2n)/g^2(n)\}$ is bounded. Hence we construct an integrable majorant for the integrand. Thus, by the Lebesgue dominated convergence theorem,

$$\lim_{n\to\infty}\int_0^1 ((1+s)^{H-1/2}g(n(1+s))/g(n) - s^{H-1/2}g(ns)/g(n))^2\,ds$$

$$= \int_0^1 ((1+s)^{H-1/2} - s^{H-1/2})^2\,ds. \tag{10}$$

We now estimate the second integral:

$$\int_1^\infty ((1+s)^{H-1/2}g(n(1+s))/g(n) - s^{H-1/2}g(ns)/g(n))^2\,ds$$

$$\le 2\int_1^\infty ((1+s)^{H-1/2} - s^{H-1/2})^2 g^2(n(1+s))/g^2(n)\,ds$$

$$+2\int_1^\infty s^{2H-1}(g(n(1+s)) - g(ns))^2/g^2(n)\,ds. \tag{11}$$

Consider the first summand on the right-hand side of (11). It is clear that

$$\int_1^\infty ((1+s)^{H-1/2} - s^{H-1/2})^2 g^2(n(1+s))/g^2(n)\,ds \le 2\int_1^\infty s^{2H-3}g^2(ns)/g^2(n)\,ds.$$

For the second summand in (11) we obtain the following simple estimate:

$$\int_1^\infty s^{2H-1}(g(n(1+s)) - g(ns))^2/g^2(n)\,ds \le C\int_1^\infty s^{2H-3}g^2(ns)/g^2(n)\,ds, \quad n\ge 1,$$

where C is a positive constant. So, we have

$$\int_1^\infty ((1+s)^{H-1/2}g(n(1+s))/g(n) - s^{H-1/2}g(ns)/g(n))^2\,ds$$

$$\le (4+2C)\int_1^\infty s^{2H-3}g^2(ns)/g^2(n)\,ds. \tag{12}$$

We now study the right-hand side of (12). Using the L'Hospital principle we have

$$\lim_{n\to\infty}\int_1^\infty s^{2H-3}g^2(ns)/g^2(n)\,ds = \lim_{n\to\infty}((g^2(n)n^{2H-2})^{-1}\int_n^\infty g^2(s)s^{2H-3}\,ds)$$

$$= \lim_{n\to\infty}(2-2H-2ng'(n)/g(n))^{-1} = (2-2H)^{-1} = \int_1^\infty s^{2H-3}\lim_{n\to\infty}g^2(ns)/g^2(n)\,ds.$$

Thus, the functions $\{s^{2H-3}g^2(ns)/g^2(n)\}_{n\geq 1}$ are uniformly integrable. Therefore, we construct a uniformly integrable majorant for the sequence of functions

$$\{((1+s)^{H-1/2}g(n(1+s))/g(n) - s^{H-1/2}g(ns)/g(n))^2\}_{n\geq 1}.$$

Hence,

$$\lim_{n\to\infty}\int_1^\infty((1+s)^{H-1/2}g(n(1+s))/g(n) - s^{H-1/2}g(ns)/g(n))^2\,ds$$

$$= \int_1^\infty((1+s)^{H-1/2} - s^{H-1/2})^2\,ds. \tag{13}$$

¿From (10) and (13) we have

$$\lim_{n\to\infty}\int_0^\infty((1+s)^{H-1/2}g(n(1+s))/g(n) - s^{H-1/2}g(ns)/g(n))^2\,ds$$

$$= \int_0^\infty((1+s)^{H-1/2} - s^{H-1/2})^2\,ds.$$

The proof of the lemma is complete.

2.2. Proof of Theorem 1. We will follow the scheme of proving the corresponding result in [76]. Put $n=[t]$. By (6) it is easy to deduce the representation

$$S_n = \sum_{k\in\mathbb{Z}}(A_{n-k} - A_{-k})\xi_k.$$

We now split S_n into the two summands $\overline{S}_n$ and $\widetilde{S}_n$:

$$\overline{S}_n = \sum_{k=0}^\infty(A_{n+k} - A_k)\xi_{-k}, \quad \widetilde{S}_n = \sum_{k=1}^\infty(A_{n-k} - A_{-k})\xi_k.$$

Put

$$\gamma_k = W(k) - W(k-1),\ k=1,2,\dots, \quad \gamma_{-k} = \widetilde{W}(k+1) - \widetilde{W}(k),\ \ k=0,1,\dots,$$

where $W(s)$ and $\widetilde{W}(s)$ are two independent versions of the standard Wiener process. Moreover, introduce the following notations:

$$\overline{G}_n = \sum_{k=0}^{\infty}(A_{n+k} - A_k)\gamma_{-k}, \quad \widetilde{G}_n = \sum_{k=1}^{\infty}(A_{n-k} - A_{-k})\gamma_k, \quad G_n = \overline{G}_n + \widetilde{G}_n.$$

These random variables are Gaussian analogs of the sums $\overline{S}_n$, $\widetilde{S}_n$, and S_n respectively.

It is well known (see [30]) that if $\mathbf{E}|\xi_0|^\alpha < \infty$ for some $\alpha > 2$ then the sequences $\{\xi_k, \gamma_k; k \geq 1\}$ and $\{\xi_{-k}, \gamma_{-k}; k \geq 0\}$ can be defined on a common probability space so that

$$\sum_{k=1}^{n}\xi_k - \sum_{k=1}^{n}\gamma_k = o(n^{1/\alpha}) \text{ a.s.}, \quad \sum_{k=0}^{n-1}\xi_{-k} - \sum_{k=0}^{n-1}\gamma_{-k} = o(n^{1/\alpha}) \text{ a.s. as } n \to \infty. \tag{14}$$

In the sequel we assume that the sequences of the random variables $\{\xi_k, \gamma_k; k \in \mathbb{Z}\}$ are defined on a common probability space in such a way.

Lemma 1 *As $n \to \infty$ the following is valid: $|S_n - G_n| = o(\Delta_{\alpha,n})$ a. s.*

Proof. Introduce the following notations:

$$\overline{S}_n^{(l)} = \sum_{k=0}^{l}(A_{n+k} - A_k)\xi_{-k}, \quad \overline{G}_n^{(l)} = \sum_{k=0}^{l}(A_{n+k} - A_k)\gamma_{-k}, \quad \sigma_k = \sum_{i=0}^{k-1}\xi_{-i}, \quad k \geq 1.$$

It is clear that $\xi_{-k} = \sigma_{k+1} - \sigma_k$. By the Abel formula (a discrete analog of the formula of integration by parts), we then obtain

$$\overline{S}_n^{(l)} = \sum_{i=1}^{l}\sigma_i(a_i - a_{n+i}) + \sigma_{l+1}(A_{n+l} - A_l).$$

In absolutely the same manner, we deduce the representation

$$\overline{G}_n^{(l)} = \sum_{i=1}^{l}\varrho_i(a_i - a_{n+i}) + \varrho_{l+1}(A_{n+l} - A_l),$$

where $\varrho_k = \sum_{i=0}^{k-1}\gamma_{-i}, k \geq 1$. In this case,

$$\overline{S}_n^{(l)} - \overline{G}_n^{(l)} = \sum_{i=1}^{l}\delta_i(a_i - a_{n+i}) + \delta_{l+1}(A_{n+l} - A_l),$$

where $\delta_m = \sigma_m - \varrho_m$. ¿From (14) and (5) it follows that

$$\delta_{l+1}(A_{n+l} - A_l) \to 0 \quad \text{a.s.} \quad \text{as} \quad l \to \infty.$$

Hence, with probability 1,

$$\overline{S}_n - \overline{G}_n = \sum_{i=1}^{\infty}\delta_i(a_i - a_{n+i}).$$

Further, the following inequality is valid:

$$|\overline{S}_n - \overline{G}_n| \le \sup_{k\ge n}(k^{-1/\alpha}\max_{m\le k}|\delta_m|)\Delta^+_{\alpha,n}.$$

Indeed, we have

$$|\overline{S}_n - \overline{G}_n| \le \sum_{k=1}^{n-1}|\delta_k||a_{n+k}-a_k| + \sum_{k=n}^{\infty}|\delta_k||a_{n+k}-a_k|$$

$$= \sum_{k=1}^{n-1}\frac{\delta_k}{n^{1/\alpha}}n^{1/\alpha}|a_{n+k}-a_k| + \sum_{k=n}^{\infty}\frac{\delta_k}{k^{1/\alpha}}k^{1/\alpha}|a_{n+k}-a_k|$$

$$\le \max\left\{\sup_{k\le n}\frac{|\delta_k|}{n^{1/\alpha}}, \sup_{k\ge n}\frac{|\delta_k|}{k^{1/\alpha}}\right\}\Delta^+_{\alpha,n} \le \sup_{k\ge n}\left(\frac{\max_{m\le k}|\delta_m|}{k^{1/\alpha}}\right)\Delta^+_{\alpha,n}.$$

Since

$$\sup_{k\ge n}\left(\frac{\max_{m\le k}|\delta_m|}{k^{1/\alpha}}\right) = o(1) \quad \text{a.s. as} \quad n\to\infty;$$

therefore, $|\overline{S}_n - \overline{G}_n| = o(\Delta^+_{\alpha,n})$ a.s. By analogy we obtain $|\widetilde{S}_n - \widetilde{G}_n| = o(\Delta^-_{\alpha,n})$ a.s. Hence,

$$|S_n - G_n| = o(\Delta_{\alpha,n}) \quad \text{a.s. as} \quad n\to\infty.$$

The proof of the lemma is complete.

Introduce the sequence of random variables:

$$B_0 = 0, \quad B_n = \sigma L_H^{-1/2}\Bigg(\sum_{k=1}^{n}(n-k+1)^{H-1/2}g(n-k+1)\gamma_k$$

$$+\sum_{k=0}^{\infty}((n+k+1)^{H-1/2}g(n+k+1)-(k+1)^{H-1/2}g(k+1))\gamma_{-k}\Bigg), \quad n\ge 1.$$

Lemma 2. *B_n is finite with probability 1 for every n.*

Proof. It suffices to show that

$$\sum_{k=0}^{\infty}((n+k+1)^{H-1/2}g(n+k+1)-(k+1)^{H-1/2}g(k+1))^2 < \infty.$$

Indeed, we have

$$\sum_{k=0}^{\infty}((n+k+1)^{H-1/2}g(n+k+1)-(k+1)^{H-1/2}g(k+1))^2$$

$$\le 2\sum_{k=0}^{\infty}((n+k+1)^{H-1/2}-(k+1)^{H-1/2})^2g^2(n+k+1)$$

$$+2\sum_{k=0}^{\infty}(k+1)^{2H-1}(g(n+k+1)-g(k+1))^2. \tag{15}$$

The first series on the right-hand side of (15) converges due to the estimate

$$((n+k+1)^{H-1/2}-(k+1)^{H-1/2})^2 = O(k^{2H-3}) \quad \text{as} \quad k\to\infty.$$

The second series on the right-hand side of (15) converges as well since

$$(g(n+k+1)-g(k+1))^2 = O(g^2(k+1)/(k+1)^2) \quad \text{as} \quad k\to\infty,$$

and, for sufficiently large arguments, each slowly varying function has a power majorant with an arbitrarily small positive exponent (for example, see [51]). The proof of the lemma is complete.

Lemma 3. *As $n\to\infty$ the following is valid: $|G_n - B_n| = O(\sqrt{\Delta_n \log n})$ a. s.*

Proof. We have $\Delta_n = \mathbf{D}(G_n - B_n)$. Since $G_n - B_n$ has Gaussian distribution, we obtain the inequality

$$P(|G_n - B_n| \geq 2\sqrt{\Delta_n \log n}) \leq \exp\{-(2\sqrt{\Delta_n \log n})^2/(2\Delta_n)\} = n^{-2}.$$

Using the Borel–Cantelli lemma, we arrive at the claim.

Lemma 4. *As $n\to\infty$, we have*

$$|B_n - B_H(n)| = O(\sigma L_H^{-1/2}\Upsilon_H^{1/2}\sqrt{\log n}) \quad \textit{a. s.},$$

where

$$\Upsilon_H = \int_0^{\infty}((s+1)^{H-1/2}-s^{H-1/2})^2(g(s+1))^2\,ds + \int_0^{\infty}s^{2H-1}(g(s+1)-g(s))^2\,ds,$$

if g increases;

$$\Upsilon_H = \int_0^{\infty}((s+1)^{H-1/2}-s^{H-1/2})^2(g(s))^2\,ds + \int_0^{\infty}s^{2H-1}(g(s+1)-g(s))^2\,ds,$$

if g decreases.

Proof. Put

$$B_n^* = \sum_{k=1}^{n}(n-k+1)^{H-1/2}g(n-k+1)\gamma_k$$

$$+\sum_{k=0}^{\infty}((n+k+1)^{H-1/2}g(n+k+1)-(k+1)^{H-1/2}g(k+1))\gamma_{-k},$$

$$B^*(n) = \int_0^{n}(n-s)^{H-1/2}g(n-s)\,dW(s)$$

$$+\int_0^\infty ((n+s)^{H-1/2}g(n+s)-s^{H-1/2}g(s))d\widetilde{W}(s).$$

The quantity B_n^* has the following representation as a stochastic integral:

$$\int_0^n (n-[s])^{H-1/2}g(n-[s])\,dW(s)$$

$$+\int_0^\infty ((n+[s]+1)^{H-1/2}g(n+[s]+1)-([s]+1)^{H-1/2}g([s]+1))d\widetilde{W}(s).$$

We now study the variance $\mathbf{D}(B_n^*-B^*(n))$. We have

$$\mathbf{D}(B_n^*-B^*(n))\le \int_0^n ((n-s)^{H-1/2}g(n-s)-(n-[s])^{H-1/2}g(n-[s]))^2\,ds$$

$$+2\int_0^\infty (s^{H-1/2}g(s)-([s]+1)^{H-1/2}g([s]+1))^2\,ds$$

$$+2\int_0^\infty ((n+s)^{H-1/2}g(n+s)-(n+[s]+1)^{H-1/2}g(n+[s]+1))^2\,ds. \tag{16}$$

Consider the first summand on the right-hand side of (16):

$$\int_0^n ((n-[s])^{H-1/2}g(n-[s])-(n-s)^{H-1/2}g(n-s))^2\,ds$$

$$\le 2\int_0^n (g(n-[s]))^2((n-[s])^{H-1/2}-(n-s)^{H-1/2})^2\,ds$$

$$+2\int_0^n (n-s)^{2H-1}(g(n-[s])-g(n-s))^2\,ds. \tag{17}$$

Estimate the first summand on the right-hand side of (17). Since $s\le n-[n-s]<s+1$, we obtain the inequality

$$\int_0^n (g(n-[s]))^2((n-[s])^{H-1/2}-(n-s)^{H-1/2})^2\,ds$$

$$= \int_0^n (g(n-[n-s]))^2((n-[n-s])^{H-1/2} - s^{H-1/2})^2\, ds$$

$$\leq \int_0^n (g(n-[n-s]))^2((s+1)^{H-1/2} - s^{H-1/2})^2\, ds. \tag{18}$$

Estimate the second summand on the right-hand side of (17):

$$\int_0^n (n-s)^{2H-1}(g(n-[s]) - g(n-s))^2\, ds = \int_0^n s^{2H-1}(g(n-[n-s]) - g(s))^2\, ds$$

$$\leq \int_0^n s^{2H-1}(g(s+1) - g(s))^2\, ds. \tag{19}$$

We consider now the second summand on the right-hand side of (16):

$$\int_0^\infty (([s]+1)^{H-1/2}g([s]+1) - s^{H-1/2}g(s))^2\, ds \leq 2\int_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2(g([s]+1))^2\, ds$$

$$+2\int_0^\infty s^{2H-1}(g(s+1) - g(s))^2\, ds. \tag{20}$$

In turn, the third summand on the right-hand side of (16) admits the estimate

$$\int_0^\infty ((n+s)^{H-1/2}g(n+s) - (n+[s]+1)^{H-1/2}g(n+[s]+1))^2\, ds$$

$$= \int_n^\infty (([s]+1)^{H-1/2}g([s]+1) - s^{H-1/2}g(s))^2\, ds$$

$$\leq 2\int_n^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2(g([s]+1))^2\, ds + 2\int_n^\infty s^{2H-1}(g([s]+1) - g(s))^2\, ds. \tag{21}$$

Further, we note that $|g([s]+1) - g(s)| \leq |g(s+1) - g(s)|$. Moreover, $g([s]+1) \leq g(s+1)$ if g increases, and $g([s]+1) \leq g(s)$ if g decreases. Combining these upper bounds for the right-hand sides in (18)–(21), we obtain the following inequality:

$$\mathbf{D}(B_n^* - B^*(n)) \leq 8\int_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2(g(s+1))^2\, ds + 8\int_0^\infty s^{2H-1}(g(s+1) - g(s))^2\, ds,$$

if g increases;

$$\mathbf{D}(B_n^* - B^*(n)) \le 8\int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s))^2\, ds + 8\int\limits_0^\infty s^{2H-1}(g(s+1) - g(s))^2\, ds,$$

if g decreases.

We prove that the integral

$$\int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s+1))^2\, ds$$

exists. Indeed, as $s \to \infty$, we have $((s+1)^{H-1/2} - s^{H-1/2})^2 = O(s^{2H-3})$ and $g(s+1)/g(s) \to 1$. Thus the integral $\int_1^\infty s^{2H-3}(g(s))^2\, ds$ exists (see the properties of slowly varying functions in [13]) that implies existence of the integral

$$\int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s+1))^2\, ds.$$

Finally, by Borel–Cantelli lemma we infer the claim.

In the sequel we need the following (see [81])

Lemma 5. *Let $\{\xi(t); 0 \le t \le 1\}$ be a centered Gaussian process. Moreover, $\xi(0) = 0$ and $\mathbf{E}(\xi(t) - \xi(s))^2 \le C|t-s|^{2H}$ for some $H > 0$. Then, for all $x \ge 0$,*

$$P(\sup_{t\in[0,1]} |\xi(t)| > x) \le 4\exp\bigl\{-C_H^{-2} x^2 (8C)^{-1}\bigr\},$$

where $C_H = \sum_{k=1}^\infty 2^{-kH} k^{1/2}$.

Lemma 6. *As $t \to \infty$ the following is valid:*

$$B_H(t) - B_H([t]) = O\bigl(\Phi_H^{1/2} C_H \sqrt{\log t}\bigr) \quad a.\ s.,$$

where

$$\Phi_H = \frac{1}{4H}(g(1))^2 + \int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s+1))^2\, ds + \int\limits_0^\infty s^{2H-1}(g'(s))^2\, ds,$$

if g increases;

$$\Phi_H = \frac{1}{4H}(g(0))^2 + (g(0))^2\int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2\, ds + \int\limits_0^\infty s^{2H-1}(g'(s))^2\, ds,$$

if g decreases, and $C_H = \sum_{k=1}^\infty 2^{-kH} k^{1/2}$.

Proof. Let $t \in [n, n+1]$, $n \geq 0$. Put

$$\xi(t) = \int_0^t (t-s)^{H-1/2} g(t-s) dW(s) + \int_0^\infty ((t+s)^{H-1/2} g(t+s) - s^{H-1/2} g(s)) d\widetilde{W}(s)$$

$$- \int_0^n (n-s)^{H-1/2} g(n-s) dW(s) - \int_0^\infty ((n+s)^{H-1/2} g(n+s) - s^{H-1/2} g(s)) d\widetilde{W}(s).$$

Estimate the moment $\mathbf{E}(\xi(v+n) - \xi(u+n))^2, u \leq v$, where $u, v \in [0,1]$. We have

$$\mathbf{E}(\xi(v+n) - \xi(u+n))^2 \leq \int_{u+n}^{v+n} (v+n-s)^{2H-1} (g(v+n-s))^2 \, ds$$

$$+ \int_0^{u+n} ((v+n-s)^{H-1/2} g(v+n-s) - (u+n-s)^{H-1/2} g(u+n-s))^2 \, ds$$

$$+ \int_0^\infty ((v+n+s)^{H-1/2} g(v+n+s) - (u+n+s)^{H-1/2} g(u+n+s))^2 \, ds. \tag{22}$$

For the first summand on the right-hand side of (22) the following equality holds:

$$\int_{u+n}^{v+n} (v+n-s)^{2H-1} (g(v+n-s))^2 \, ds = \int_0^{v-u} s^{2H-1} (g(s))^2 \, ds.$$

¿From this it follows that

$$\int_0^{v-u} s^{2H-1} (g(s))^2 \, ds \leq \frac{1}{2H} (g(1))^2 (v-u)^{2H}, \tag{23}$$

if g increases, and

$$\int_0^{v-u} s^{2H-1} (g(s))^2 \, ds \leq \frac{1}{2H} (g(0))^2 (v-u)^{2H}, \tag{24}$$

if g decreases.

Consider the second summand on the right-hand side of (22):

$$\int_0^{u+n} ((v+n-s)^{H-1/2} g(v+n-s) - (u+n-s)^{H-1/2} g(u+n-s))^2 \, ds$$

$$\leq 2 \int_0^{u+n} (((v-u+s)^{H-1/2} - s^{H-1/2})^2 (g(v-u+s))^2 \, ds$$

$$+2 \int_0^{u+n} s^{2H-1} (g(v-u+s) - g(s))^2 \, ds. \tag{25}$$

Finally, the third summand on the right-hand side of (22) admits the upper bound

$$\int_0^{\infty} ((v+n+s)^{H-1/2} g(v+n+s) - (u+n+s)^{H-1/2} g(u+n+s))^2 \, ds$$

$$\leq 2 \int_{u+n}^{\infty} ((v-u+s)^{H-1/2} - s^{H-1/2})^2 (g(v-u+s))^2 \, ds$$

$$+2 \int_{u+n}^{\infty} s^{2H-1} (g(v-u+s) - g(s))^2 \, ds. \tag{26}$$

Combining the right-hand sides of (25) and (26), we obtain an upper bound for the sum of the second and third summands on the right-hand side of (22):

$$2 \int_0^{\infty} ((v-u+s)^{H-1/2} - s^{H-1/2})^2 (g(v-u+s))^2 \, ds + 2 \int_0^{\infty} s^{2H-1} (g(v-u+s) - g(s))^2 \, ds. \tag{27}$$

Consider the first integral in (27):

$$\int_0^{\infty} ((v-u+s)^{H-1/2} - s^{H-1/2})^2 (g(v-u+s))^2 \, ds$$

$$= (v-u)^{2H} \int_0^{\infty} ((s+1)^{H-1/2} - s^{H-1/2})^2 (g((v-u)(s+1)))^2 \, ds.$$

Whence we have the estimates

$$\int_0^{\infty} ((v-u+s)^{H-1/2} - s^{H-1/2})^2 (g(v-u+s))^2 \, ds$$

$$\leq (v-u)^{2H} \int_0^{\infty} ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s+1))^2 \, ds, \tag{28}$$

if g increases, and

$$\int\limits_0^\infty ((v-u+s)^{H-1/2} - s^{H-1/2})^2 (g(v-u+s))^2\, ds$$

$$\le (v-u)^{2H}(g(0))^2 \int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2\, ds, \tag{29}$$

if g decreases.

Estimate the second integral in (27):

$$\int\limits_0^\infty s^{2H-1}(g(v-u+s)-g(s))^2\, ds \le (v-u)^2 \int\limits_0^\infty s^{2H-1}(g'(s))^2\, ds. \tag{30}$$

Combining the upper bounds in (23), (24), (28)–(30) we have

$$\mathbf{E}(\xi(v+n)-\xi(u+n))^2$$

$$\le (v-u)^{2H}\left(\frac{1}{2H}(g(1))^2 + 2\int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s+1))^2\, ds\right) + 2(v-u)^2 \int\limits_0^\infty s^{2H-1}(g'(s))^2\, ds$$

$$\le (v-u)^{2H}\left(\frac{1}{2H}(g(1))^2 + 2\int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2 (g(s+1))^2\, ds + 2\int\limits_0^\infty s^{2H-1}(g'(s))^2\, ds\right),$$

if g increases, and

$$\mathbf{D}(\xi(v+n)-\xi(u+n))$$

$$\le (v-u)^{2H}\left(\frac{1}{2H}(g(0))^2 + 2(g(0))^2 \int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2\, ds\right) + 2(v-u)^2 \int\limits_0^\infty s^{2H-1}(g'(s))^2\, ds$$

$$\le (v-u)^{2H}\left(\frac{1}{2H}(g(0))^2 + 2(g(0))^2 \int\limits_0^\infty ((s+1)^{H-1/2} - s^{H-1/2})^2\, ds + 2\int\limits_0^\infty s^{2H-1}(g'(s))^2\, ds\right),$$

if g decreases. Finally, by Lemma 5 and by the Borel–Cantelli lemma (letting $x = 4\sqrt{2}(\Phi_H \log n)^{1/2} C_H$), we complete the proof.

The claim of Theorem 1 follows from the lemmas above.

2.3. Proof of Proposition 2 and Corollary 1.

Proof of Proposition 2. At first we prove item (i).

Lemma 7. *The following is valid:* $|a_k| = O(g(k)k^{H-3/2})$ *as* $k \to \infty$.

Proof. Indeed,

$$|(k+1)^{H-1/2} g(k+1) - k^{H-1/2} g(k)| = |g(k+1)((k+1)^{H-1/2} - k^{H-1/2})$$

$+k^{H-1/2}(g(k+1)-g(k))| \leq g(k+1)|(k+1)^{H-1/2}-k^{H-1/2}|+k^{H-1/2}|g(k+1)-g(k)|.$

Hence,

$$|(k+1)^{H-1/2}g(k+1)-k^{H-1/2}g(k)| = O(g(k)k^{H-3/2}) \quad \text{as} \quad k \to \infty.$$

From this the claim of the lemma follows.

Lemma 8. *The following is valid:*

$$\sum_{k=0}^{n}|a_k| = O(g(n)n^{H-1/2}) \quad as \quad n \to \infty.$$

Proof. Consider the case when g increases. Obviously, there exists a constant C such that $|a_k| \leq Cg(k+1)(k+1)^{H-3/2}$ for all $k \geq 0$. Moreover, the sequence $g(k)k^{H-3/2}$ is monotone decreasing for all sufficiently large $k \geq N$. So,

$$\sum_{k=0}^{n}|a_k| \leq C\sum_{k=1}^{n+1}g(k)k^{H-3/2} = \sum_{k=1}^{N}g(k)k^{H-3/2} + \sum_{k=N+1}^{n+1}g(k)k^{H-3/2}$$

$$\leq \sum_{k=1}^{N}g(k)k^{H-3/2} + \int_N^{n+1}g(x)x^{H-3/2}dx. \tag{31}$$

Consider the last summand on the right-hand side of (31):

$$\int_N^{n+1}g(x)x^{H-3/2}dx = g(n+1)\frac{(n+1)^{H-1/2}}{H-1/2} - g(N)\frac{N^{H-1/2}}{H-1/2} - \int_N^{n+1}\frac{x^{H-1/2}}{H-1/2}g'(x)dx.$$

Hence,

$$\int_N^{n+1}g(x)x^{H-3/2}\left(1+\frac{xg'(x)}{(H-1/2)g(x)}\right)dx = g(n+1)\frac{(n+1)^{H-1/2}}{H-1/2} - g(N)\frac{N^{H-1/2}}{H-1/2}.$$

We can choose the number N so that $\left|\frac{xg'(x)}{(H-1/2)g(x)}\right| \leq 1/2$ for $x \geq N$. Whence we obtain

$$\int_N^{n+1}g(x)x^{H-3/2}\,dx \leq 2g(n+1)\frac{(n+1)^{H-1/2}}{H-1/2}.$$

The case when g decreases is settled in a similar manner. The proof of the lemma is complete.

We now estimate $\Delta_{\alpha,n}$. First of all, we note that

$$\Delta_{\alpha,n} = n^{1/\alpha}\sum_{k=-n}^{n-1}|a_k-a_{k+n}| + \sum_{k=n}^{\infty}k^{1/\alpha}|a_k-a_{k+n}|. \tag{32}$$

Estimate the first summand on the right-hand side of (32). We have

$$n^{1/\alpha}\sum_{k=-n}^{n-1}|a_k-a_{k+n}|\le 2n^{1/\alpha}\sum_{k=0}^{2n-1}|a_k|=O(g(n)n^{H-1/2+1/\alpha}).$$

Consider now the second summand on the right-hand side of (32):

$$\sum_{k=n}^{\infty}k^{1/\alpha}|a_k-a_{k+n}|\le\sum_{k=n}^{2n-1}k^{1/\alpha}|a_k|+\sum_{k=2n}^{\infty}|a_k|(k^{1/\alpha}-(k-n)^{1/\alpha})$$

$$\le(2n-1)^{1/\alpha}ng(n)n^{H-3/2}+\sum_{k=2n}^{\infty}C_1k^{H-3/2}g(k)n(1/\alpha)(k-n)^{1/\alpha-1}$$

$$\le(2n-1)^{1/\alpha}ng(n)n^{H-3/2}+\sum_{k=n}^{\infty}C_1(k+n)^{H-3/2}g(k+n)n(1/\alpha)k^{1/\alpha-1}$$

$$\le(2n-1)^{1/\alpha}ng(n)n^{H-3/2}+\sum_{k=n}^{\infty}C_1k^{H-3/2}g(k)n(1/\alpha)k^{1/\alpha-1},\tag{33}$$

where C_1 is some constant (see Lemma 7). Consider the last summand on the right-hand side of (33). We have (see the proof of Lemma 8)

$$\sum_{k=n}^{\infty}k^{H-3/2}g(k)k^{1/\alpha-1}\le\int_{n-1}^{\infty}x^{H-3/2+1/\alpha-1}g(x)dx=O(g(n)n^{H-3/2+1/\alpha})\quad\text{as}\quad n\to\infty.$$

Thus,

$$\sum_{k=n}^{\infty}k^{1/\alpha}|a_k-a_{k+n}|=O(g(n)n^{H-1/2+1/\alpha})\quad\text{as}\quad n\to\infty.$$

Whence we deduce that

$$\Delta_{\alpha,n}=O(g(n)n^{H-1/2+1/\alpha}).$$

We now prove item (ii) of Proposition 2. We have $|a_k|=O(g(k)k^{H-3/2})$, $k\to\infty$ (the proof is absolutely the same as in the case $H>1/2$). Also, $\sum_{k=0}^{\infty}|a_k|<\infty$. We then estimate $\Delta_{\alpha,n}$. As in the case $H>1/2$ we separately study the sums

$$n^{1/\alpha}\sum_{k=-n}^{n-1}|a_k-a_{k+n}|,\quad\sum_{k=n}^{\infty}k^{1/\alpha}|a_k-a_{k+n}|.\tag{34}$$

For the first sum in (34) the upper bound is elementary:

$$n^{1/\alpha}\sum_{k=-n}^{n-1}|a_k-a_{k+n}|\le 2n^{1/\alpha}\sum_{k=0}^{2n-1}|a_k|=O(n^{1/\alpha})\quad\text{as}\quad n\to\infty.$$

Consider the second sum in (34). We have

$$\sum_{k=n}^{\infty} k^{1/\alpha}|a_k - a_{k+n}| \le \sum_{k=n}^{2n-1} k^{1/\alpha}|a_k| + \sum_{k=2n}^{\infty} |a_k|(k^{1/\alpha} - (k-n)^{1/\alpha})$$

$$\le (2n-1)^{1/\alpha} \sum_{k=n}^{2n-1} |a_k| + (2n)^{1/\alpha} \sum_{k=2n}^{\infty} |a_k| = O(n^{1/\alpha}) \quad \text{as} \quad n \to \infty.$$

Therefore, $\Delta_{\alpha,n} = O(n^{1/\alpha})$ as $n \to \infty$.

We now prove item (iii) of Proposition 2. We consider only the case when g increases. The case when g decreases is settled in the same manner as in item (ii). Consider the first sum in (34):

$$n^{1/\alpha} \sum_{k=-n}^{n-1} |a_k - a_{k+n}| \le 2n^{1/\alpha} \sum_{k=0}^{2n-1} |a_k| = O(g(n)n^{1/\alpha}) \quad \text{as} \quad n \to \infty.$$

Estimate the second sum in (34):

$$\sum_{k=n}^{\infty} k^{1/\alpha}|a_k - a_{k+n}| = \sum_{k=n}^{2n-1} k^{1/\alpha}|a_k| + \sum_{k=2n}^{\infty} |a_k|(k^{1/\alpha} - (k-n)^{1/\alpha})$$

$$\le C(2n-1)^{1/\alpha} \sum_{k=n}^{2n-1} g(k)/k + C \sum_{k=n}^{\infty} g(k)/k((n+k)^{1/\alpha} - k^{1/\alpha})$$

$$\le C(2n-1)^{1/\alpha} \sum_{k=n}^{2n-1} g(k)/k + Cn \sum_{k=n}^{\infty} g(k)k^{1/\alpha-2} = O(g(n)n^{1/\alpha}) \quad \text{as} \quad n \to \infty,$$

where C is a constant (see Lemma 7). So, in the conditions of item (iii),

$$\Delta_{\alpha,n} = O(g(n)n^{1/\alpha}) \quad \text{as} \quad n \to \infty, \text{ if } g \text{ increases};$$

$$\Delta_{\alpha,n} = O(n^{1/\alpha}) \quad \text{as} \quad n \to \infty, \text{ if } g \text{ decreases}.$$

We prove in all three cases above that $\Delta_n = O(g^2(n)n^{\max\{2\beta,2\gamma\}})$ as $n \to \infty$. Introduce the notations:

$$\alpha_0 = a_0 - \sigma L_H^{-1/2} g(1),$$

$$\alpha_n = a_n - \sigma L_H^{-1/2}((n+1)^{H-1/2}g(n+1) - n^{H-1/2}g(n)), \quad n \ge 1,$$

$$\beta_n = A_n - \sigma L_H^{-1/2}(n+1)^{H-1/2}g(n+1), \quad n \ge 0.$$

Note that

$$\Delta_n = \sum_{m=0}^{\infty} (\beta_{n+m} - \beta_m)^2 + \sum_{m=0}^{n-1} \beta_m^2. \qquad (35)$$

At first we consider the second summand on the right-hand side of (35):

$$\sum_{m=0}^{n-1} \beta_m^2 \le C \sum_{k=1}^{n} k^{2\beta-1} g^2(k) = O(n^{2\beta} g^2(n)) \quad \text{as} \quad n \to \infty,$$

where C is a constant (see (7)). We now consider the first summand on the right-hand side of (35):

$$\begin{aligned}\sum_{m=0}^{\infty} (\beta_{n+m} - \beta_m)^2 &\le \sum_{m=0}^{n-1} (\beta_{n+m} - \beta_m)^2 + \sum_{m=n}^{\infty} (\beta_{n+m} - \beta_m)^2 \\ &\le 2 \sum_{m=0}^{2n-1} \beta_m^2 + \sum_{m=n}^{\infty} (\beta_{n+m} - \beta_m)^2. \end{aligned} \tag{36}$$

By analogy with the above,

$$\sum_{m=0}^{2n-1} \beta_m^2 = O(n^{2\beta} g^2(n)) \quad \text{as} \quad n \to \infty.$$

Further, we have

$$|\beta_{n+m} - \beta_m| \le \sum_{k=m+1}^{m+n} |\alpha_k| \le C_1 n g(m) m^{\gamma - 3/2},$$

where C_1 is a constant (see (7)). Hence,

$$\sum_{m=n}^{\infty} (\beta_{n+m} - \beta_m)^2 \le C_1^2 n^2 \sum_{m=n}^{\infty} g^2(m) m^{2\gamma-3} = O(g^2(n) n^{2\gamma}) \quad \text{as} \quad n \to \infty.$$

Therefore,

$$\Delta_n = O(g^2(n) n^{\max\{2\beta, 2\gamma\}}) \quad \text{as} \quad n \to \infty,$$

which yields the claim.

Corollary 1 follows from Proposition 2 and Theorem 1.

2.4. Proof of Theorem 2. Put

$$A_{k,n}(t) = n^{-H} \sum_{j=-k+1}^{-k+[nt]} a_j, \quad V_n^2 = \mathbf{D} S_n.$$

Then the following representation is valid:

$$Z_{n,H}(t) = \sum_{k \in \mathbb{Z}} A_{k,n}(t) \xi_k.$$

We need some auxiliary statements.

Lemma 9. [64, p. 456] *For all $k \in \mathbb{Z}$, the following is valid:*

$$|a_{k+1} + \cdots + a_{k+n}| \leq \left(4V_n \sum_{k\in\mathbb{Z}} |a_k|^2 \left(1 + \frac{1}{2V_n}\right)\right)^{1/2}.$$

Lemma 10. *For all $1 \geq t \geq \tau \geq 0$ we have*

$$\mathbf{E}Z_{n,H}(t)Z_{n,H}(\tau) \to \mathbf{E}B_H^0(t)B_H^0(\tau) \quad as \quad n \to \infty.$$

Proof. Indeed,

$$\mathbf{E}(Z_{n,H}(t) - Z_{n,H}(\tau))^2 = n^{-2H} \sum_{i\in\mathbb{Z}} \left(\sum_{j=-i+1}^{-i+[nt]-[n\tau]} a_j \right)^2$$

$$= \frac{([nt] - [n\tau])^{2H}}{n^{2H}} \frac{V^2_{[nt]-[n\tau]}}{([nt] - [n\tau])^{2H}}.$$

Whence we obtain

$$\mathbf{E}(Z_{n,H}(t) - Z_{n,H}(\tau))^2 \to \sigma^2(t-\tau)^{2H} = \mathbf{E}\big(B_H^0(t) - B_H^0(\tau)\big)^2.$$

Using the convergence of the second moments of the one-dimensional projections of the random processes $Z_{n,H}(t)$ we deduce the equality

$$2\mathbf{E}Z_{n,H}(t)Z_{n,H}(\tau) = \mathbf{E}(Z_{n,H}(t))^2 + \mathbf{E}(Z_{n,H}(\tau))^2 - \mathbf{E}(Z_{n,H}(t) - Z_{n,H}(\tau))^2 \to 2\mathbf{E}B_H^0(t)B_H^0(\tau).$$

The proof of the lemma is complete.

Lemma 11. *For all $1 \geq t \geq \tau \geq 0$ there is a positive constant K depending only on $\{a_i\}$ such that*

$$\sum_{k\in\mathbb{Z}} (A_{k,n}(t) - A_{k,n}(\tau))^2 \leq K \frac{([nt] - [n\tau])^{2H}}{n^{2H}}.$$

Proof We have

$$\sum_{k\in\mathbb{Z}} (A_{k,n}(t) - A_{k,n}(\tau))^2 = \frac{([nt] - [n\tau])^{2H}}{n^{2H}} \frac{V^2_{[nt]-[n\tau]}}{([nt] - [n\tau])^{2H}}.$$

Since $V_n^2/n^{2H} \to \sigma^2$ as $n \to \infty$, we complete the proof.

Lemma 12 [93, p. 86]. *Let $\{X_k\}_{k=1,\ldots,n}$ be independent centered random variables and $\mathbf{E}|X_k|^\alpha < \infty$ for all k and some $\alpha \geq 2$. Put*

$$S_n = \sum_{k=1}^n X_k, \quad M_{\alpha,n} = \sum_{k=1}^n \mathbf{E}|X_k|^\alpha, \quad B_n = \sum_{k=1}^n \mathbf{E}X_k^2.$$

Then

$$\mathbf{E}|S_n|^\alpha \le c(\alpha)\bigl(M_{\alpha,n} + B_n^{\alpha/2}\bigr),$$

where $c(\alpha)$ is a positive constant depending only on α.

Lemma 13. *The following inequality is valid:*

$$\mathbf{E}|Z_{n,H}(t) - Z_{n,H}(\tau)|^\alpha \le C\left(\frac{[nt]-[n\tau]}{n}\right)^{\alpha H},$$

where C is a constant depending on the distributions of ξ_0, α, and $\{a_i\}$.

Proof. By Lemma 12 and the Fatou theorem we obtain

$$\mathbf{E}\Bigl|\sum_{k\in\mathbb{Z}}(A_{k,n}(t) - A_{k,n}(\tau))\xi_k\Bigr|^\alpha$$

$$\le c(\alpha)\Bigl(\sum_{k\in\mathbb{Z}}|A_{k,n}(t) - A_{k,n}(\tau)|^\alpha \mathbf{E}|\xi_0|^\alpha + \Bigl(\sum_{k\in\mathbb{Z}}(A_{k,n}(t) - A_{k,n}(\tau))^2\Bigr)^{\alpha/2}\Bigr)$$

$$\le c(\alpha)(1+\mathbf{E}|\xi_0|^\alpha)\Bigl(\sum_{k\in\mathbb{Z}}(A_{k,n}(t) - A_{k,n}(\tau))^2\Bigr)^{\alpha/2}.$$

Note that we applied here the following elementary inequality:

$$\sum_{k\in\mathbb{Z}}|b_k|^\gamma \le \Bigl(\sum_{k\in\mathbb{Z}}|b_k|\Bigr)^\gamma, \quad \gamma \ge 1.$$

Using Lemma 11, we complete the proof.

Lemma 14. *Let $1 \ge t_3 \ge t_2 \ge t_1 \ge 0$. Then the following inequality is valid:*

$$\mathbf{E}|Z_{n,H}(t_3) - Z_{n,H}(t_2)|^{\alpha/2}|Z_{n,H}(t_2) - Z_{n,H}(t_1)|^{\alpha/2} \le C(t_3-t_2)^{\alpha H},$$

where C is a positive constant depending on α, $\{a_i\}$, and the distribution of ξ_0.

Proof. By the Cauchy–Bunyakovskiĭ inequality we obtain

$$\mathbf{E}|Z_{n,H}(t_3) - Z_{n,H}(t_2)|^{\alpha/2}|Z_{n,H}(t_2) - Z_{n,H}(t_1)|^{\alpha/2}$$

$$\le (\mathbf{E}|Z_{n,H}(t_3) - Z_{n,H}(t_2)|^\alpha)^{1/2}(\mathbf{E}|Z_{n,H}(t_3) - Z_{n,H}(t_2)|^\alpha)^{1/2}$$

$$\le \frac{C_1}{n^{\alpha H}}([nt_3]-[nt_2])^{\alpha H/2}([nt_2]-[nt_1])^{\alpha H/2} \le \frac{C_1}{n^{\alpha H}}([nt_3]-[nt_1])^{\alpha H}$$

$$\le C_1(t_3-t_1+1/n)^{\alpha H},$$

where C_1 is the constant in Lemma 13. If $t_3 - t_1 \ge 1/n$ then, using the previous inequality, we obtain the following upper bound:

$$\mathbf{E}|Z_{n,H}(t_3) - Z_{n,H}(t_2)|^{\alpha/2}|Z_{n,H}(t_2) - Z_{n,H}(t_1)|^{\alpha/2} \le C_1 2^{\alpha H}(t_3-t_1)^{\alpha H}.$$

If $t_3 - t_1 < 1/n$ then

$$\mathbf{E}(|Z_{n,H}(t_3) - Z_{n,H}(t_2)|^{\alpha/2}|Z_{n,H}(t_2) - Z_{n,H}(t_1)|^{\alpha/2}) = 0,$$

since either the pair t_1 and t_2 or the pair t_3 and t_2 lies on an interval of the form $[(i-1)/n, i/n)$. The proof of the lemma is complete.

From Lemma 14 it follows that the family of distributions of the random processes $\{Z_{n,H}\}_{n\geq 1}$ is tight if $\alpha H > 1$ (see [12]).

Lemma 15 [76]. *Let $\{b_{ni}; n \geq 1, i \in \mathbb{Z}\}$ be an array of real numbers, and let $\{\zeta_{ni}; n \geq 1, i \in \mathbb{Z}\}$ be an array of random variables satisfying the following conditions:*

L1. $\lim_{n\to\infty} \sum_{i\in\mathbb{Z}} b_{ni}^2 = 1$;

L2. $\lim_{n\to\infty} \sup_{i\in\mathbb{Z}} |b_{ni}| = 0$;

L3. *For every $n \geq 1$ the sequence $\{\zeta_{ni}, i \in \mathbb{Z}\}$ consists of i.i.d. random variables with mean zero and variance* 1.

L4. $\lim_{K\to\infty} \sup_{n\geq 1} \mathbf{E}\zeta_{n0}^2 I(|\zeta_{n0}| > K) = 0$.

Then the sums $\sum_{i\in\mathbb{Z}} b_{ni}\zeta_{ni}$ converge in distribution to a standard Gaussian random variable as $n \to \infty$.

Lemma 16. *The finite-dimensional distributions of random processes $Z_{n,H}(t)$ converge to the corresponding finite-dimensional distributions of random process $B_H^0(t)$ as $n \to \infty$.*

Proof. It suffices to prove that $\sum_{i=1}^{l} c_i Z_{n,H}(t_i)$ converges in distribution to $\sum_{i=1}^{l} c_i B_H^0(t_i)$ for every finite set of numbers $\{c_i; i = 1 \ldots l\}$. So, we observe first that

$$z_n \equiv \sum_{i=1}^{l} c_i Z_{n,H}(t_i) = \sum_{k\in\mathbb{Z}} \sum_{i=1}^{l} c_i A_{k,n}(t_i)\xi_k.$$

Further (see Lemma 10)

$$\mathbf{E}z_n^2 = \sum_{i,j=1}^{l} c_i c_j \mathbf{E}(Z_{n,H}(t_i) Z_{n,H}(t_j)) \to \sum_{i,j=1}^{l} c_i c_j \mathbf{E}\big(B_H^0(t_i) B_H^0(t_j)\big)$$

$$= \mathbf{E}\left(\sum_{i=1}^{l} c_i B_H^0(t_i)\right)^2 \quad \text{as} \quad n \to \infty.$$

Use the notation

$$\delta^2 = \mathbf{E}\left(\sum_{i=1}^{l} c_i B_H^0(t_i)\right)^2, \quad z_{1n} = z_n/\delta.$$

Then

$$\mathbf{E}z_{1n}^2 = \sum_{k\in\mathbb{Z}} \left(\sum_{i=1}^{l} c_i/\delta A_{k,n}(t_i)\right)^2 \to 1 \quad \text{as} \quad n \to \infty.$$

Now, under the conditions of Lemma 15 put

$$b_{nk} = \sum_{i=1}^{l} c_i/\delta A_{k,n}(t_i), \quad \zeta_{nk} = \xi_k.$$

It is easy to see that, in this case, Condition L1 is valid.

Using Lemma 9 we find, as $n \to \infty$, that

$$\sup_k \left| \sum_{i=1}^{l} (c_i A_{k,n}(t_i)) \right| = O(n^{-H/2}).$$

Hence, Condition L2 is valid. It is easy to verify that Conditions L3 and L4 hold as well.

So, the random variables z_n converge in distribution to a normal random variable with mean zero and variance δ^2. The proof of the lemma is complete.

The claim of Theorem 2 follows from Lemmas 14 and 16.

2.5. Proof of Theorem 3. Put

$$\gamma_k^{(n)} = W(k/n) - W((k-1)/n),\ k \geq 1, \quad \gamma_{-k}^{(n)} = \widetilde{W}((k+1)/n) - \widetilde{W}(k/n),\ k \geq 0.$$

Denote by $\{G_l^{(n)}, l \geq 0\}$ the corresponding partial sums of moving averages in (1) for these Gaussian random variables $\{\gamma_k^{(n)}, k \in \mathbb{Z}\}$; and by

$$\Gamma_{n,H}(t) = \frac{G_{[nt]}^{(n)}}{n^{H-1/2}},$$

the corresponding normalized partial sum process.

We split the proof into the auxiliary lemmas below.

Lemma 17 [30]. *Let $\{\xi_k; k \geq 1\}$ be i.i.d. random variables, $\mathbf{E}\xi_1 = 0, \mathbf{D}\xi_1 = 1$ and $\mathbf{E}|\xi_1|^\alpha < \infty$ for some $\alpha > 2$. Then there exists a Wiener process W such that*

$$\mathbf{P}\Big(\sup_{k \leq n} |\sum_{j \leq k} \xi_k/\sqrt{n} - W(k/n)| \geq x(n) n^{-1/2}\Big) \leq C_\xi n x^{-\alpha}(n),$$

where $x(n) \to \infty$ as $n \to \infty$, and the constant C_ξ depends only on the distribution of ξ_0.

Put

$$\xi_i^{(n)} = \frac{\xi_i}{\sqrt{n}},\ i \in \mathbb{Z}, \quad S_k^{(n)} = \frac{S_k}{\sqrt{n}},\ k \geq 0;$$

$$\overline{S}_k^{(n)} = \sum_{i=0}^{\infty} (A_{k+i} - A_i)\xi_{-i}^{(n)},\ k \geq 0; \quad \underline{S}_0^{(n)} = 0,\ \underline{S}_k^{(n)} = \sum_{i=1}^{k} A_{k-i}\xi_i^{(n)},\ k \geq 1;$$

$$\overline{G}_k^{(n)} = \sum_{i=0}^{\infty} (A_{k+i} - A_i)\gamma_{-i}^{(n)},\ k \geq 0; \quad \underline{G}_0^{(n)} = 0,\ \underline{G}_k^{(n)} = \sum_{i=1}^{k} A_{k-i}\gamma_i^{(n)},\ k \geq 1.$$

Further, we have

$$S_{[nt]}^{(n)} = \overline{S}_{[nt]}^{(n)} + \underline{S}_{[nt]}^{(n)}, \quad G_{[nt]}^{(n)} = \overline{G}_{[nt]}^{(n)} + \underline{G}_{[nt]}^{(n)}.$$

In Lemmas 18–20 below we assume the conditions of the first part of Theorem 3 to be fulfilled.

Lemma 18. *The following equality is valid:*

$$\mathbf{P}\Big(\sup_{t\in[0,1]} \big|\underline{S}_{[nt]}^{(n)} - \underline{G}_{[nt]}^{(n)}\big| n^{-H+1/2} \ge C_1 n^{-\frac{\alpha-2}{2(\alpha+1)}}\Big) \le C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}}.$$

Proof Put

$$\sigma_i^{(n)} = \sum_{j=1}^{i} \xi_i^{(n)}, \quad \varrho_i^{(n)} = \sum_{j=1}^{i} \gamma_i^{(n)}.$$

Using the Abel formula, we obtain the representations

$$\underline{S}_{[nt]}^{(n)} = \sum_{i=1}^{[nt]} \sigma_i^{(n)} a_{[nt]-i}, \quad \underline{G}_{[nt]}^{(n)} = \sum_{i=1}^{[nt]} \varrho_i^{(n)} a_{[nt]-i}.$$

Hence,

$$\big|\underline{S}_{[nt]}^{(n)} - \underline{G}_{[nt]}^{(n)}\big| = \left|\sum_{i=1}^{[nt]} (\sigma_i^{(n)} - \varrho_i^{(n)}) a_{[nt]-i}\right| \le \sup_{1\le i\le n} \big|\sigma_i^{(n)} - \varrho_i^{(n)}\big| \sum_{i=0}^{n-1} |a_i|$$

$$\le \sup_{1\le i\le n} \big|\sigma_i^{(n)} - \varrho_i^{(n)}\big| C_1 n^{H-1/2}.$$

By Lemma 17 we have

$$\mathbf{P}\big(\sup_{t\in[0,1]} \big|\underline{S}_{[nt]}^{(n)} - \underline{G}_{[nt]}^{(n)}\big| n^{-H+1/2} \ge C_1 n^{H-1/2} n^{-H} x)$$

$$\le \mathbf{P}(\sup_{1\le i\le n} \big|\sigma_i^{(n)} - \varrho_i^{(n)}\big| \ge x n^{-1/2}) \le C_\xi n x^{-\alpha}.$$

Letting $x = n^{\frac{3}{2(\alpha+1)}}$ we arrive at the claim of the lemma.

Lemma 19. *For all $n \ge N$ the following inequality is valid:*

$$\mathbf{P}\Big(\sup_{t\in[0,1]} \big|\overline{S}_{[nt]}^{(n)} - \overline{G}_{[nt]}^{(n)}\big| n^{-H+1/2} \ge n^{-\frac{\alpha-2}{2(\alpha+1)}} (2^{H+3/2} C_1 + C(1 - 2^{H-3/2+1/\alpha})^{-2})\Big)$$

$$\le C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}} (1 + \alpha/(\alpha-1)).$$

Proof. We have

$$\overline{S}_{[nt]}^{(n)} = \sum_{i=0}^{n-1} (A_{[nt]+i} - A_i)\xi_{-i}^{(n)} + \sum_{i=n}^{\infty} (A_{[nt]+i} - A_i)\xi_{-i}^{(n)}. \tag{37}$$

We now represent the second summand on the right-hand side (37) as

$$\sum_{i=n}^{\infty}(A_{[nt]+i}-A_i)\xi_{-i}^{(n)}=\sum_{k=1}^{\infty}\sum_{i=n2^{k-1}}^{n2^k-1}(A_{[nt]+i}-A_i)\xi_{-i}^{(n)}.$$

Put

$$S_i^{(k-1)}(n)=\sum_{j=n2^{k-1}}^{i}\xi_{-j}^{(n)},\quad G_i^{(k-1)}(n)=\sum_{j=n2^{k-1}}^{i}\gamma_{-j}^{(n)},\quad i=n2^{k-1},\dots,n2^k-1.$$

Further, by the Abel formula we obtain

$$\sum_{i=n2^{k-1}}^{n2^k-1}(A_{[nt]+i}-A_i)\xi_{-i}^{(n)}=\sum_{i=n2^{k-1}}^{n2^k-2}(a_{i+1}-a_{[nt]+i+1})S_i^{(k-1)}(n)$$

$$+(A_{[nt]+n2^k-1}-A_{n2^k-1})S_{n2^k-1}^{k-1}(n).$$

Thus,

$$\begin{aligned}
&\Big|\sum_{i=n2^{k-1}}^{n2^k-1}(A_{[nt]+i}-A_i)\big(\xi_{-i}^{(n)}-\gamma_{-i}^{(n)}\big)\Big|\\
&\le\sup_{n2^{k-1}\le i\le n2^k-1}\big|S_i^{(k-1)}(n)-G_i^{(k-1)}(n)\big|(A_{n2^k-1}-A_{n2^{k-1}}\\
&-(A_{[nt]+n2^k-1}-A_{[nt]+n2^{k-1}})+A_{[nt]+n2^k-1}-A_{n2^k-1})\\
&=\sup_{n2^{k-1}\le i\le n2^k-1}\big|S_i^{(k-1)}(n)-G_i^{(k-1)}(n)\big|(A_{[nt]+n2^{k-1}}-A_{n2^{k-1}}).
\end{aligned}$$

Since $a_n\le Cn^{H-3/2}$; therefore,

$$A_{[nt]+n2^{k-1}}-A_{n2^{k-1}}\le Cn(n2^{k-1})^{H-3/2}=Cn^{H-1/2}2^{(k-1)(H-3/2)}.$$

So,

$$\begin{aligned}
&\Big|\sum_{i=n2^{k-1}}^{n2^k-1}(A_{[nt]+i}-A_i)\big(\xi_{-i}^{(n)}-\gamma_{-i}^{(n)}\big)\Big|\\
&\le\sup_{n2^{k-1}\le i\le n2^k-1}\big|S_i^{(k-1)}(n)-G_i^{(k-1)}(n)\big|Cn^{H-1/2}2^{(k-1)(H-3/2)}.
\end{aligned}$$

Hence (see Lemma 17),

$$\mathbf{P}\bigg(\sup_{t\in[0,1]}\Big|\sum_{i=n2^{k-1}}^{n2^k-1}(A_{[nt]+i}-A_i)\big(\xi_{-i}^{(n)}-\gamma_{-i}^{(n)}\big)\Big|n^{-H+1/2}\ge Cn^{-1/2}2^{(k-1)(H-3/2)}x(2^{k-1})^{1/\alpha}k^{\frac{1+\varepsilon}{\alpha}}\bigg)$$

$$\le\mathbf{P}\big(\sup_{n2^{k-1}\le i\le n2^k-1}\big|S_i^{(k-1)}(n)-G_i^{(k-1)}(n)\big|\ge\frac{x}{\sqrt{n}}(2^{k-1})^{1/\alpha}k^{\frac{1+\varepsilon}{\alpha}}\big)\le C_\xi\frac{n}{x^\alpha k^{1+\varepsilon}}.$$

Therefore,

$$\mathbf{P}\bigg(\sup_{t\in[0,1]}\bigg|\sum_{i=n}^{\infty}(A_{[nt]+i}-A_i)\big(\xi_{-i}^{(n)}-\gamma_{-i}^{(n)}\big)\bigg|n^{-H+1/2}$$
$$\geq C\frac{x}{\sqrt{n}}\sum_{k=1}^{\infty}2^{(k-1)(H-3/2+1/\alpha)}k^{\frac{1+\varepsilon}{\alpha}}\bigg)\leq C_\xi\frac{n}{x^\alpha}\sum_{k=1}^{\infty}\frac{1}{k^{1+\varepsilon}}.$$

Putting $\varepsilon=\alpha-1$ and $x=n^{\frac{3}{2(\alpha+1)}}$ and using the elementary facts

$$\sum_{k=1}^{\infty}\frac{1}{k^\alpha}\leq\frac{\alpha}{\alpha-1},\quad\sum_{k=1}^{\infty}k2^{(k-1)(H-3/2+1/\alpha)}=(1-2^{H-3/2+1/\alpha})^{-2},$$

we obtain the inequality

$$\mathbf{P}\bigg(\sup_{t\in[0,1]}\bigg|\sum_{i=n}^{\infty}(A_{[nt]+i}-A_i)\big(\xi_{-i}^{(n)}-\gamma_{-i}^{(n)}\big)\bigg|n^{-H+1/2}\geq Cn^{-\frac{\alpha-2}{2(\alpha+1)}}(1-2^{H-3/2+1/\alpha})^{-2}\bigg)$$
$$\leq C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}}\alpha/(\alpha-1). \tag{38}$$

We now study the first summand on the right-hand side of (37). Put

$$\sigma_i^{(n)}=\sum_{j=0}^{i}\xi_{-j}^{(n)},\quad\varrho_i^{(n)}=\sum_{j=0}^{i}\gamma_{-j}^{(n)},\quad i=0,\dots,n-1.$$

Using the Abel formula once again we have

$$\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\xi_{-i}^{(n)}=\sum_{i=0}^{n-2}\sigma_i^{(n)}(a_{i+1}-a_{[nt]+i+1})+\sigma_{n-1}^{(n)}(A_{[nt]+n-1}-A_{n-1});$$

$$\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\gamma_{-i}^{(n)}=\sum_{i=0}^{n-2}\varrho_i^{(n)}(a_{i+1}-a_{[nt]+i+1})+\varrho_{n-1}^{(n)}(A_{[nt]+n-1}-A_{n-1}).$$

Therefore,

$$\bigg|\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\big(\xi_{-i}^{(n)}-\gamma_{-i}^{(n)}\big)\bigg|\leq 4\sup_{0\leq i\leq n-1}\big|\sigma_i^{(n)}-\varrho_i^{(n)}\big|\sum_{i=0}^{2n-1}|a_i|.$$

By the conditions of Theorem 3,

$$\sum_{i=0}^{2n-1}|a_i|\leq C_1 2^{H-1/2}n^{H-1/2}.$$

So,

$$\mathbf{P}\bigg(\sup_{t\in[0,1]}\bigg|\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\big(\xi_{-i}^{(n)}-\gamma_{-i}^{(n)}\big)\bigg|n^{-H+1/2}\geq xC_1 2^{H+3/2}n^{H-1/2}n^{-H}\bigg)$$

$$\le \mathbf{P}\Big(\sup_{0\le i\le n-1}\big|\sigma_i^{(n)}-\varrho_i^{(n)}\big|\ge xn^{H-1/2}n^{-H}\Big)\le C_\xi n x^{-\alpha}.$$

Letting $x=n^{\frac{3}{2(\alpha+1)}}$ we obtain the upper bound

$$\mathbf{P}\left(\sup_{t\in[0,1]}\left|\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\big(\xi_{-i}^{(n)}-\gamma_{-i}^{(n)}\big)\right|n^{-H+1/2}\ge C_1 2^{H+3/2}n^{-\frac{\alpha-2}{2(\alpha+1)}}\right)\le C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}}. \quad (39)$$

Finally, combining inequalities (38) and (39) we infer the claim of Lemma 19.

Lemmas 18 and 19 imply the following

Lemma 20. *There exists a probability space such that*

$$\mathbf{P}\big(\sup_{t\in[0,1]}|Z_{n,H}(t)-\Gamma_{n,H}(t)|\ge n^{-\frac{\alpha-2}{2(\alpha+1)}}(C_1(2^{H+3/2}+1)+C(1-2^{H-3/2+1/\alpha})^{-2})\big)$$

$$\le C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}}(2+\alpha/(\alpha-1))$$

for all $n\ge N$.

In Lemmas 21–23 we assume the conditions of the second part of Theorem 3 to be fulfilled.

Lemma 21. *The following inequality is valid:*

$$\mathbf{P}\big(\sup_{t\in[0,1]}\big|\underline{S}_{[nt]}^{(n)}-\underline{G}_{[nt]}^{(n)}\big|n^{-H+1/2})\ge C_1 n^{-\frac{\alpha H-1}{\alpha+1}}\big)\le C_\xi n^{-\frac{\alpha H-1}{\alpha+1}}.$$

Proof Put

$$\sigma_i^{(n)}=\sum_{j=1}^{i}\xi_i^{(n)},\quad \varrho_i^{(n)}=\sum_{j=1}^{i}\gamma_i^{(n)}.$$

Then

$$\big|\underline{S}_{[nt]}^{(n)}-\underline{G}_{[nt]}^{(n)}\big|=\left|\sum_{i=1}^{[nt]}\big(\sigma_i^{(n)}-\varrho_i^{(n)}\big)a_{[nt]-i}\right|$$

$$\le \sup_{1\le i\le n}\big|\sigma_i^{(n)}-\varrho_i^{(n)}\big|\sum_{i=0}^{[nt]-1}|a_i|\le \sup_{1\le i\le n}\big|\sigma_i^{(n)}-\varrho_i^{(n)}\big|C_1.$$

Whence using Lemma 17, we deduce the upper bound

$$\mathbf{P}\big(\sup_{t\in[0,1]}\big|\underline{S}_{[nt]}^{(n)}-\underline{G}_{[nt]}^{(n)}\big|n^{-H+1/2}\ge C_1 n^{-H}x\big)$$

$$\le \mathbf{P}\big(\sup_{1\le i\le n}\big|\sigma_i^{(n)}-\varrho_i^{(n)}\big|\ge xn^{-1/2}\big)\le C_\xi n x^{-\alpha}.$$

Letting $x=n^{\frac{H+1}{\alpha+1}}$ we arrive at the claim of the lemma.

Lemma 22. *The following inequality holds:*

$$\mathbf{P}\big(\sup_{t\in[0,1]}\big|\overline{S}^{(n)}_{[nt]}-\overline{G}^{(n)}_{[nt]}\big|n^{-H+1/2}\geq 4C_1n^{-\frac{\alpha H-1}{\alpha+1}}+Cn^{-\frac{\alpha-2}{2(\alpha+1)}}(1-2^{H-3/2+1/\alpha})^{-2}\big)$$
$$\leq C_\xi n^{-\frac{\alpha H-1}{\alpha+1}}+C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}}\alpha/(\alpha-1)$$

for all $n\geq N$.

Proof. Consider the second summand on the right-hand side of (37) for which the Gaussian approximation is deduced by analogy with that in the proof of Lemma 19:

$$\mathbf{P}\left(\sup_{t\in[0,1]}\bigg|\sum_{i=n}^{\infty}(A_{[nt]+i}-A_i)\big(\xi^{(n)}_{-i}-\gamma^{(n)}_{-i}\big)\bigg|n^{-H+1/2}\geq Cn^{-\frac{\alpha-2}{2(\alpha+1)}}(1-2^{H-3/2+1/\alpha})^{-2}\right)$$
$$\leq C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}}\alpha/(\alpha-1) \tag{40}$$

for all $n\geq N$. We now consider the first summand on the right-hand side of (37). As in the proof of Lemma 19, put

$$\sigma^{(n)}_i=\sum_{j=0}^{i}\xi^{(n)}_{-j},\quad \varrho^{(n)}_i=\sum_{j=0}^{i}\gamma^{(n)}_{-j},\quad i=0,\dots,n-1.$$

By the Abel formula we obtain the representations

$$\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\xi^{(n)}_{-i}=\sum_{i=0}^{n-2}\sigma^{(n)}_i(a_{i+1}-a_{[nt]+i+1})+\sigma^{(n)}_{n-1}(A_{[nt]+n-1}-A_{n-1});$$
$$\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\gamma^{(n)}_{-i}=\sum_{i=0}^{n-2}\varrho^{(n)}_i(a_{i+1}-a_{[nt]+i+1})+\varrho^{(n)}_{n-1}(A_{[nt]+n-1}-A_{n-1}).$$

Therefore,

$$\bigg|\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\big(\xi^{(n)}_{-i}-\gamma^{(n)}_{-i}\big)\bigg|\leq 4\sup_{0\leq i\leq n-1}\big|\sigma^{(n)}_i-\varrho^{(n)}_i\big|\sum_{i=0}^{2n-1}|a_i|.$$

Since $\sum_{i=0}^{2n-1}|a_i|\leq C_1$, we have

$$\mathbf{P}\left(\sup_{t\in[0,1]}\bigg|\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\big(\xi^{(n)}_{-i}-\gamma^{(n)}_{-i}\big)\bigg|n^{-H+1/2}\geq 4xC_1n^{-H}\right)$$
$$\leq\mathbf{P}\big(\sup_{0\leq i\leq n-1}\big|\sigma^{(n)}_i-\varrho^{(n)}_i\big|n^{-H+1/2}\geq xn^{-H}\big)\leq C_\xi nx^{-\alpha}.$$

Letting $x=n^{\frac{H+1}{\alpha+1}}$ we obtain the estimate

$$\mathbf{P}\left(\sup_{t\in[0,1]}\bigg|\sum_{i=0}^{n-1}(A_{[nt]+i}-A_i)\big(\xi^{(n)}_{-i}-\gamma^{(n)}_{-i}\big)\bigg|n^{-H+1/2}\geq 4C_1n^{-\frac{\alpha H-1}{\alpha+1}}\right)\leq C_\xi n^{-\frac{\alpha H-1}{\alpha+1}}. \tag{41}$$

Combining (40) and (41), we complete the proof.

As a consequence of Lemmas 21 and 22 we obtain the following

Lemma 23. *There exists a probability space such that*

$$\mathbf{P}\big(\sup_{t\in[0,1]} |Z_{n,H}(t) - \Gamma_{n,H}(t)| \geq 5C_1 n^{-\frac{\alpha H-1}{\alpha+1}} + Cn^{-\frac{\alpha-2}{2(\alpha+1)}}(1-2^{H-3/2+1/\alpha})^{-2}\big)$$

$$\leq 2C_\xi n^{-\frac{\alpha H-1}{\alpha+1}} + C_\xi n^{-\frac{\alpha-2}{2(\alpha+1)}}\alpha/(\alpha-1)$$

for all $n \geq N$.

Introduce the sequence of random variables

$$B_0^{(n)} = 0, \quad B_k^{(n)} = \sum_{i=1}^{k}(k-i+1)^{H-1/2}\gamma_i^{(n)}$$

$$+\sum_{i=0}^{\infty}((k+i+1)^{H-1/2} - (i+1)^{H-1/2})\gamma_{-i}^{(n)}, \quad k \geq 1.$$

Lemma 24. *The following inequality is valid:*

$$\mathbf{P}\big(\sup_{t\in[0,1]} n^{-H+1/2}\big|G_{[nt]}^{(n)} - B_{[nt]}^{(n)}\big| \geq n^{-H}\sqrt{2\delta_n(H+1)\log n}\big) \leq n^{-H}.$$

Proof. We have

$$\mathbf{P}\big(\max\big\{n^{-H+1/2}\big|G_1^{(n)} - B_1^{(n)}\big|, \ldots, n^{-H+1/2}\big|G_n^{(n)} - B_n^{(n)}\big|\big\} \geq n^{-H}\psi_n\big)$$

$$\leq \sum_{k=1}^{n}\mathbf{P}\big(\big|G_k^{(n)} - B_k^{(n)}\big| \geq n^{-1/2}\psi_n\big) \leq n\exp\big(-\psi_n^2/2\delta_n\big).$$

Putting

$$\psi_n = \sqrt{2\delta_n(H+1)\log n}$$

we complete the proof.

Lemma 25. *The following inequality is valid:*

$$\mathbf{P}\big(\sup_{t\in[0,1]}\big|B_{[nt]}^n n^{-H+1/2} - B_H^0([nt]/n)\big| \geq 8n^{-H}\sqrt{2\Upsilon_H^0(H+1)\log n}\big) \leq n^{-H}.$$

Proof. First of all we note that the value $B_k^{(n)}/n^{H-1/2}$ can be represented as the stochastic integral

$$\int_0^{k/n}(k/n - [ns]/n)^{H-1/2}dW(s) + \int_0^{\infty}((k/n + ([ns]+1)/n)^{H-1/2}$$

$$-(([ns]+1)/n)^{H-1/2})d\widetilde{W}(s).$$

Then (see the proof of Lemma 4),

$$n^{2H}\mathbf{E}\big(B_H^0(k/n)-n^{1/2-H}B_k^n\big)^2\le 64\Upsilon_H^0,\quad 0\le k\le n.$$

Putting $\psi_n=8\sqrt{2\Upsilon_H^0(H+1)\log n}$, we complete the proof.

Lemma 26. *The following inequality is valid:*

$$\mathbf{P}\big(\sup_{t\in[0,1]}\big|B_H^0(t)-B_H^0([nt]/n)\big|\ge 2\sqrt{2}\sigma n^{-H}C_H\sqrt{(H+1)\log n}\big)\le 4n^{-H},$$

where $C_H=\sum_{k=1}^\infty 2^{-kH}k^{1/2}$.

Proof. We have

$$\sup_{t\in[0,1]}\big|B_H^0(t)-B_H^0([nt]/n)\big|=\max_{k=1\ldots n}\big\{\sup_{t\in[\frac{k-1}{n},\frac{k}{n}]}\big|B_H^0(t)-B_H^0([nt]/n)\big|\big\}.$$

Using the stationarity of the increments of the process $B_H^0(t)$ as well as its H-homogeneity (i.e., $B_H^0(nt)\overset{d}{\to}n^HB_H^0(t)$) and the upper bound $\mathbf{E}(B_H^0(t)-B_H^0(s))^2\le\sigma^2|t-s|^H$ (see Lemma 5), we deduce the following estimate:

$$\mathbf{P}\big(\sup_{t\in[0,1]}\big|B_H^0(t)-B_H^0([nt]/n)\big|\ge n^{-H}\psi_n\big)$$

$$\le\sum_{k=1}^n\mathbf{P}\big(\sup_{t\in[\frac{k-1}{n},\frac{k}{n}]}\big|B_H^0(t)-B_H^0([nt]/n)\big|\ge n^{-H}\psi_n\big)$$
$$\le n\mathbf{P}\big(\sup_{t\in[0,\frac{1}{n})}\big|B_H^0(t)\big|\ge n^{-H}\psi_n\big)=n\mathbf{P}\big(\sup_{t\in[0,1]}\big|B_H^0(t)\big|\ge\psi_n\big)$$
$$\le 4n\exp\big(-C_H^{-2}\psi_n^2/\sigma^2\big).$$

Putting $\psi_n=2\sqrt{2}\sigma C_H\sqrt{(H+1)\log n}$, we complete the proof.

Lemmas 24–26 yield

Lemma 27. *For all* $H\in(0,1)$ *we have the inequality*

$$\mathbf{P}\big(\sup_{t\in[0,1]}\big|n^{-H+1/2}G_{[nt]}^{(n)}-B_H^0(t)\big|$$

$$\ge n^{-H}\sqrt{(H+1)\log n}(8\sqrt{2\Upsilon_H^0}+\sqrt{2\delta_n}+2\sqrt{2}\sigma C_H)\big)\le 6n^{-H}.$$

The claim of Theorem 3 follows from Lemmas 20, 23, and 27.

2.6. Proof of Proposition 3. Introduce the notations:

$$\alpha_0=a_0-\sigma L_H^{-1/2},\quad \alpha_n=a_n-\sigma L_H^{-1/2}((n+1)^{H-1/2}-n^{H-1/2}),\quad n\ge 1;$$

$$\beta_n = A_n - \sigma L_H^{-1/2}(n+1)^{H-1/2}, \quad n \ge 0.$$

Prove item (i). Note that in the case under consideration we have

$$\Delta_n = \sum_{m=0}^{\infty}(\beta_{n+m} - \beta_m)^2 + \sum_{m=0}^{n-1} \beta_m^2. \tag{42}$$

Denote by $\Delta_n^{(1)}$ the first sum on the right-hand side of (42); and by $\Delta_n^{(2)}$, the second sum. We have

$$\Delta_n^{(1)} \le 2\sum_{m=0}^{\infty} \beta_{n+m}^2 + 2\sum_{m=0}^{\infty} \beta_m^2 = 2\sum_{m=n}^{\infty} \beta_m^2 + 2\sum_{m=0}^{\infty} \beta_m^2.$$

Therefore,

$$\Delta_n \le 4\sum_{m=0}^{\infty} \beta_m^2 < \infty.$$

We now prove item (ii). We have

$$\Delta_{[nt]}^{(2)} = \alpha_0^2 + \sum_{l=1}^{[nt]-1} \beta_l^2 \le \alpha_0^2 + \psi_n = O(n^{2\beta}) \quad \text{as} \;\; n \to \infty,$$

where $\psi_n = O(\sum_{l=1}^{n-1} l^{2\beta-1}) = O(n^{2\beta})$ (see the conditions of Proposition 3). Represent the value $\Delta_{[nt]}^{(1)}$ as follows:

$$\Delta_{[nt]}^{(1)} = \sum_{m=0}^{\infty}(\beta_{[nt]+m} - \beta_m)^2 = \sum_{m=0}^{n-1}(\beta_{[nt]+m} - \beta_m)^2 + \sum_{m=n}^{\infty}(\beta_{[nt]+m} - \beta_m)^2. \tag{43}$$

Estimate the first summand on the right-hand side of (43):

$$\sum_{m=0}^{n-1}(\beta_{[nt]+m} - \beta_m)^2 \le 4\sum_{m=0}^{n+[nt]-1} \beta_m^2 \le 4\sum_{m=0}^{2n-1} \beta_m^2 = O(n^{2\beta}) \quad \text{as} \;\; n \to \infty.$$

We now estimate the second summand on the right-hand side of (43):

$$|\beta_{[nt]+m} - \beta_m| \le \sum_{l=m+1}^{[nt]+m} |\alpha_l| \le C[nt]m^{\gamma-3/2}, \;\; m \ge 1;$$

$$\sum_{m=n}^{\infty}(\beta_{[nt]+m} - \beta_m)^2 = O\left(n^2\sum_{m=n}^{\infty} m^{2\gamma-3}\right) = O(n^{2\gamma}) \quad \text{as} \;\; n \to \infty,$$

where C is a positive constant. So, $\delta_n = O(n^{\max\{2\gamma, 2\beta\}})$ as $n \to \infty$ which was to be proved.

References

[1] Adell, J. A. and de la Cal, J., Coupling Methods in Poisson Approximation of Binomial Processes [Preprint], Universidad de Zaragoza, Zaragoza (1994).

[2] Araujo, A. and Giné, E. (1980). The central limit theorem for real and Banach valued random variables. *Wiley*, New York.

[3] Arcones, M. A. and Giné, E. Limit theorems for U-processes, Ann. Probab., 1993, Vol. 21, No. 3, p. 1494–1542.

[4] Arcones, M. A. and Giné, E. On the law of the iterated logarithm for canonical U-statistics and processes, Stochastic Process. Appl.,1995, Vol. 58, No. 2, p. 217–245.

[5] Barbour, A. D. and Hall, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.

[6] Barbour, A. D. Asymptotic expansions based on smooth functions in the central limit theorem. Probab. Theory. Related Fields, **72**, No. 2, 289–303 (1986).

[7] Barbour, A. D. (1987). Asymptotic expansions in the Poisson limit theorem. *Ann. Probab.* **15** 748–766.

[8] Barbour, A. D, Holst L., and Janson, S., 1992, Poisson approximation. *The Clarendon Press; Oxford University Press (Oxford Stud. Probab.; 2).*

[9] Barbour, A. D., Chen, L. H. Y. and Choi, K. P. (1995). Poisson approximation for unbounded functions, I: Independent summands. *Statist. Sinica* **5** 749–766.

[10] Bentkus, V. Yu. Asymptotic expansions of moments in the central limit theorem in Banach spaces. Litovsk. Mat. Sb., 1986, Vol 26, 1, p. 10–26. (in Russian)

[11] Bhattacharya, R.N. and Ranga Rao, R. *Normal Approximation and Asymptotic Expansions*, Wiley, New York, 1976.

[12] Billingsley, P. *Convergence of probability measures*, New York–London–Sydney–Toronto. J. Willey, 1968

[13] Blum J. R., Hanson D. L., Koopmans L. H. On the strong law of large numbers for a class of stochastic processes *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 1963, Bd 2, H. 1. S. 1–11.

[14] Borisov, I. S. (1983). Problem of accuracy of approximation in the central limit theorem for empirical measures. *Siberian Math. J.* **24** (6), 833-843.

[15] Borisov, I. S. (1990). Exponential inequalities for the distributions of von Mises and U-statistics. *Probab. Theory and Math. Statist.*, Proc. 5th Vilnius Conf., Vilnius, Lithuania, 1989, **1**, 166-178.

[16] Borisov, I. S. (1991). Approximation of von Mises statistics with multidimensional kernels. *Siberian Math. J.* **32** (4), 554-566.

[17] Borisov, I. S. (1993). Strong Poisson and mixed approximations of sums of independent random variables in Banach spaces. *Siberian Adv. Math.* **3** 1–13.

[18] Borisov, I. S. (1996). Poisson approximation of the partial sum process in Banach spaces. *Siberian Math. J.* **37** 627–634.

[19] Borisov, I. S., 2000. A note on Poisson approximation of rescalled set-indexed empirical processes. *Statist. Probab. Lett.* **46** (2), 101-103.

[20] Borisov, I. S., Mironov, D. V., 2000. Asymptotic representation of the likelihood ratio for multivariate samples with discontinuous densities. *Theory Probab. Appl.*, **45** (2), 345-356 (in Russian).

[21] Borisov, I. S. and Sakhanenko, L. A. The central limit theorem for generalized canonical Von Mises statistics. Siberian Adv. Math. 2000, Vol 10, 4, p. 1–14

[22] Borisov, I. S. and Ruzankin, P. S. (2002). Poisson approximation for expectation of unbounded functions of independent random variables. *Ann. Probab.* **30** (4), 1657-1680.

[23] Borisov, I. S. and Bystrov, A. A. Constructing a stochastic integral of a nonrandom function without orthogonality of the noise. Theory Probab. Appl, 2006. Vol 50, 1, p. 53–74

[24] Borisov, I. S. and Bystrov, A. A. Limit theorems for the canonical Von Mises statistics with dependent data. Siberian Math. J. 2006, Vol 47, 6, p. 980–989

[25] Borisov, I. S. and Volodko, N. V. Orthogonal series and limit theorems for canonical U- and V-statistics of stationary connected observations. Siberian Adv. Math., 2008, Vol 18, 3, p. 163–175

[26] Borovkov, A. A. and Sakhanenko, A. I. Remarks on convergence of random processes in nonseparable metric spaces and on the nonexistence of a Borel measure for processes in $C(0, \infty)$, *Theory Probab. Appl.*, **18**, No. 4, 774–778 (1973).

[27] Borovkov, A. A., 1976. *Probability Theory* (in Russian; "Nauka" Publ. House, Moscow).

[28] Borovkov, A. A. Asymptotic Methods in Queueing Theory. John Wiley and Sons, Chichester etc. 1984

[29] Borovkov, K. A. (1988). Refinement of Poisson approximation. *Theory Probab. Appl.* **33** 343–347.

[30] Borovkov, A. A., Mogulskiĭ, A. A., and Sakhanenko A. I., *Limit theorems for random processes*, in: Probability Theory 7 (in Russian), VINITI, Moscow, 1995, p. 5–199. (Itogi Nauki Tekh. Ser. Sovrem. Probl. Mat. Fundam. Napravleniya; Vol. 82).

[31] Borovkov, A.A., 1998. *Probability Theory* (Gordon and Breach Science Publishers, the Netherlands).

[32] Chen, L. H. Y. (1975). An approximation theorem for convolution of probability measures. *Ann. Probab.* **3** 992–999.

[33] Cox, D.C. and Kemperman, J.H.B. (1983). Sharp bounds on the absolute moments of a sum of two i.i.d. random variables *Ann. Probab.* **11** (3), 765-771.

[34] Cramér, H. On the theory of random processes, Ann. Math., Vol. 41, p. 215–230.

[35] Davydov, Yu. A., The invariance principle for stationary processes, Teor. Veroyatn. Primen., 1970, vol 24, 3, p. 487–498.

[36] Deheuvels, P. and Pfiefer, D. (1988). Poisson approximation of multinomial distributions and point processes. *J. Multivariate Anal.* **25**, 65-89.

[37] Deheuvels, P. and Mason, D.M. (1992). Functional laws of the iterated logarithm for increments of empirical and quantile processes *Ann. Probab.* **20**, 1248-1287.

[38] Deheuvels, P. and Pfeifer, D. (1986). A semigroup approach to Poisson approximation. *Ann. Probab.* **14** 663–676.

[39] Deheuvels, P., Karr, A., Pfeifer, D. and Serfling, R. (1988). Poisson approximations in selected metrics by coupling and semigroup methods with applications. *J. Statist. Plann. Inference* **22** 1–22.

[40] Deheuvels, P., Mason, D. M., 1990. Nonstandard functional laws of the iterated logarithm for tail empirical and quantile processes. *Ann.Probab.* **18**, 1693-1722.

[41] Deheuvels, P., Mason, D. M., 1995. Nonstandard local empirical processes indexed by sets. *J.Statist. Plann. Inference*, **45**, 91-112.

[42] Dehling, H., Denker, M., and Philipp, W. The almost sure invariance principle for the empirical process of U-statistic structure. Ann. Inst. Henri Poincaré, Probab. Stat., 1987, Vol 23, 2, p. 121–134

[43] Dehling, H., Taqqu, M. S. The empirical process of some long-range dependent sequences with an application to U-statistics. Ann. Statist., 1989, Vol 17, 4, p. 1767–1783

[44] Denker, M., Grillenberger, C., and Keller, G. A note on invariance principles for v. Mises' statistics. Metrika, 1985, Vol 32, p. 197–214

[45] Dobrushin, R. L. Prescribing a system of random variables by conditional distributions. – *Theory Probab. Appl.*, 1970, **15**, No. 3, p. 458–486 .

[46] Dudley, R.M., 1968. Distances of probability measures and random variables. *Ann. Math. Statist.* **39** (5), 1563-1572.

[47] Dynkin, E. B., Mandelbaum, A. Symmetric statistics, Poisson point processes and multiple Wiener integrals. Ann. Statist., 1983, Vol 11, 3, p. 739–745

[48] Eagleson, G. K. Orthogonal expansions and U-statistics. Austral. J. Statist., 1979, Vol 21, 3, p. 221–237

[49] Einmahl, J.H.J. (1987). Multivariate Empirical Processes. CWI Tract **33** Cent. Wisk. Inform., Amsterdam.

[50] Einmahl, J.H.J., 1997. Poisson and Gaussian approximation of weighted local empirical processes. *Stoch. Proces. Their Appl.* **70**, 31-58.

[51] Feller, W., 1966, An introduction to probability theory and its applications. *New York, London, Sydney*: *John Wiley and Sons, Inc.*.

[52] Filippova, A. A. Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and statistical applications, Theory Probab. Appl., 1962, Vol. 7, p. 24–58.

[53] Gikhman, I. I. and Skorokhod, A. V. An Introduction to the Theory of Stochastic Processes. Moscow, Nauka, 1965 (in Russian)

[54] Gikhman, I. I. and Skorokhod, A. V. Theory of Random Processes, Vol. 1, Moscow, Nauka, 1971

[55] Giné, E. M., Mason, D. M., and Zaitsev, A. Yu. (2001). The L_1-norm density estimator process. *Preprint.*

[56] Götze, F. and Hipp, C. Asymptotic expansions in the central limit theorem under moment conditions. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 1978, Vol 42, 1, p. 67–87.

[57] Harremoës, P. and Ruzankin, P. S., 2004, Rate of convergence to Poisson law in term of information divergence. *IEEE Trans. Inform. Theory*, **50** (9), 2145–2149.

[58] Hipp, C Edgeworth expansions for integrals of smooth functions. Ann. Probab., **5**, No. 6, 1004–1011 (1977).

[59] Hoeffding, W. A class of statistics with asymptotically normal distribution, Ann. Math. Statist., 1948, Vol. 19, p. 293–325.

[60] Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **27** 713–721.

[61] Hoeffding, W. The strong law of large numbers for U-statistics, 1961, Univ. North Carolina Inst. Statist. Mimeo. Series, Vol. 302.

[62] Horváth, L., 1990. A note on the rate of Poisson approximation of empirical processes. *Ann.Probab.* **18** (2), 724-726.

[63] Hsing, T., On the asymptotic distributions of partial sums of functionals of infinite-variance moving averages, Ann. Probab., 1999, vol 27, 3, p. 1579–1599.

[64] Ibragimov, I. A. and Linnik, Yu. V., Independent and Stationarily Connected Variables, 1971, Wolters-Noordhoff Publishing Company, Groningen, The Netherlands.

[65] Ibragimov, R. and Sharakhmetov, Sh. (1997). On the exact constant for the Rosentahl inequality. *Theory Probab. Appl.* **42** (2), 294-302.

[66] Ibragimov, R. and Sharakhmetov, Sh. (2001). The exact constant in the Rosentahl inequality for random variables with mean zero. *Theory Probab. Appl.* **46** (1), 134-138 (in Russian).

[67] Itô, K. Multiple Wiener integral J. Math. Soc. Japan, 1951, Vol. 3, p.157–164.

[68] Ivanoff, B. G. and Merzbach, E. Poisson convergence for set-indexed empirical processes. – *Stat. Probab. Lett.*, 1997, **32**, No. 1, p. 81-86.

[69] Kemperman, J.H.B. (1972). On a class of moments problems. in: Proceedings of 6th Berkley Sympos. Math. Statist. Probab., Univ. California, 1970, **2**, Berkley, Los Angelos, Univ. California Press, 1972.

[70] Khintchine, A. (1933). *Asymptotische Gesetze der Wahrschaintlichkeitsrechnung. Ergebnisse der Mathematik und ihrer grenzgebiete.* Springer, Berlin.

[71] Khintchine, A.Ya. (1936). *Asymptotic Laws of Probability Theory* (in Russian, translation from German; United Sci. and Techn. Publ. House of NKTP USSR, Moscow, Leningrad).

[72] Kolmogorov, A. N. Curves in a Hilbert space which are invariant with respect to a one-parameter group of motions, Dokl. Akad. Nauk SSSR, 1940, Vol. 26, p. 6–9 (in Russian).

[73] Kolmogorov, A. N. The Wiener spiral and some interesting curves in Hilbert space, 1940, Dokl. Akad. Nauk SSSR, Vol. 26, p. 115–118 (in Russian).

[74] Kolmogorov, A. N. Stationary sequences in Hilbert space, Bull. Mosk. Univ. Mat., 1941, Vol. 2, No. 6, p. 1–40.

[75] KOLMOGOROV, A. N. AND FOMIN, S. V. Elements of the Theory of Functions and Functional Analysis. Nauka, Moscow, 1972. (in Russian)

[76] KONSTANTOPOULOS T. AND SAKHANENKO, A. I., *Convergence and Convergence Rate to Fractional Brownian Motion for Weighted Random Sums,* [Preprint], Univ. Texas at Austin, Texas (2000).

[77] KOROLYUK, V. S. AND BOROVSKIKH, YU. V., *Theory of U-Statistics*, Mathematics and Its Appl. (Dordrecht), Kluwer Acad. Publ., Dordrecht, 1994.

[78] KRUOPIS, YU. (1986a). Precision of approximation of the generalized binomial distribution by convolutions of Poisson measures, *Lithuanian Math. J.* **26** 37–49.

[79] KRUOPIS, YU. (1986b). Approximations of distributions of sums of lattice random variables. I, II. *Lithuanian Math. J.* **26** 465–467; 692–704.

[80] KUTATELADZE, S.S. (1995). *Fundamentals of Functional Analysis.* Kluwer Texts in Mathematical Sciences, **12**, Dordrecht: Kluwer Academic Publishers.

[81] LEADBETTER, M. R., LINDGREN, G., AND ROOTZEN, H., Extremes and Related Properties of Random Sequences and Processes, 1983, Springer-Verlag, New York; Heidelberg; Berlin.

[82] LE CAM, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.*, **10**, 1181–1197.

[83] LE CAM, L. (1965). On the distribution of sums of independent random variables. In: *Bernoulli, Bayes, Laplace (anniversary volume)* (Springer, Berlin, Heidelberg, New York), 179-202.

[84] LE CAM, L. Remargues sur le théorème limit central dans les espaces localement convexes. – In: *Les Probabilitités sur les Structures Algébriques,* Paris, C.N.R.S., 1970, p. 233–249.

[85] MAJOR, P. *Multiple Wiener–Itô integrals*, Lecture Notes in Math., 1981, Vol. 849, Springer-Verlag, Berlin.

[86] MAJOR, P., 1990. A note on the approximation of the uniform empirical processes. *Ann. Probab.* **18** (1), 129-139.

[87] MAJOR, P. Asymptotic distribution for weighted U-statistics, Ann. Probab., 1994, Vol. 22, No. 3, p. 1514–1535.

[88] MANDELBROT, B. AND VAN NESS, J., Fractional Brownian motions, fractional noise and applications, SIAM Rev., 1968, vol 10, p. 422–437.

[89] VON MISES, R. On the asymptotic distribution of differentiable statistical functions, Ann. Math. Statist., 1947, Vol. 18, p.309–348.

[90] Novak, S. Yu. (1998). On the limiting distribution of extrema. *Siberian Adv. Math.* **8** 70–95.

[91] de la Peña, V. H. Decoupling and Khintchine's inequalities for U-statistics, Ann. Probab., 1992, Vol. 20, p. 1877–1892.

[92] de la Peña, V. H., Montgomery-Smith S. J, and Szulga J. Contraction and decoupling inequalities for multilinear forms and U-statistics, Ann. Probab., 1994, Vol. 22, No. 4, p. 1745–1765.

[93] Petrov, V. V.. Sums of Independent Random Variables. Nauka, Moscow (1972). [in Russian]

[94] Philipp, W. The central limit problem for mixing sequences of random variables. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 1969. Bd 12, H. 2, S. 155–171.

[95] Pinelis, I.F. and Utev, S.A. (1985). Estimation of the moments of sums of independent random variables. *Theory Probab. Appl.* **29** (3), 574-577.

[96] Pinelis, I.F. and Utev, S.A. (1989). Exact exponential bounds for sums of independent random variables. *Theory Probab. Appl.* **34** (2), 340-346.

[97] Pinsker, M. S., 1960, Information and information stability of random variables and processes. *Izdat. Akad. Nauk SSSR* (in Russian).

[98] Prohorov, Yu. V. (1953). Asymptotic behavior of the binomial distribution. *Uspekhi Mat. Nauk* **8** 135–142 (in Russian).

[99] Prohorov, Yu. V., 1956. Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1** (2), 157-214.

[100] Prokhorov, Yu. V. (1960). An extremal problem in probability theory. *Theory Probab. Appl.* **4** (2), 201-203.

[101] Prokhorov, Yu. V. (1962). Extremal problems in limit theorems. *Proceedings of the 6th USSR Conference on Probab. Theory and Math. Stat.*, Vilnius, 77-84 (in Russian).

[102] Rachev, S. T., 1984. The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory Probab. Appl.* **29** (4), 647-676.

[103] Reiss, R.-D., 1993. *A Course on Point Processes* (Springer, New York).

[104] Ronzhin, A. F. Functional limit theorems for U-statistics. Math. Notes, 1986, Vol 40, 5, p. 886–893.

[105] Roos, B., 1986, Sharp constants in the Poisson approximation. *Statist. Probab. Lett.*, **52** (2), 155–168.

[106] Roos, B. (1995). A semigroup approach to Poisson approximation with respect to the point metric. *Statist. Probab. Letters* **24**, 305–314.

[107] Roos, B. (1998). Metric multivariate Poisson approximation of the generalized multinomial distribution. *Theory Probab. Appl.* **43**, 306–315.

[108] Rosenthal, H. P. On the span in L^p of sequences of independent random variables. in: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 1972, **2**, pp. 149–167.

[109] Rubin, H., Vitale, R. A. Asymptotic distribution of symmetric statistics. Ann. Statist., 1980, Vol 8, 1, p. 165–170.

[110] Schay, G., 1974. Nearest random variables with given distributions. *Ann. Probab.* **2** (1), 163-166.

[111] Sen, P. K. Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* 1972. Bd 25, H. 1. S. 71–82.

[112] Serfling, R.J., 1975. A general Poisson approximation theorem. *Ann. Probab.* **3** (4), 726-731.

[113] Shiryaev, A. N., Probability, 1995, Springer-Verlag, New York, NY.

[114] Stoyanov, J. Counterexamples in Probability, 1997, John Wiley & Sons, Chichester.

[115] Strassen, V., 1965. The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** (2), 423-439.

[116] Tikhomirov, A. N. On the accuracy of normal approximation of the probability of hitting a ball of sums of weakly dependent Hilbert space valued random variables.I. *Theory Probab. Appl*, 1991, **36**, 4, p. 738–751.

[117] Utev, S. A. (1984). Extremal problems in moment inequalities. *Theory Probab. Appl.* **29** (2), 411-412.

[118] Utev, S. A. (1985). Extremal problems in moment inequalities. *Proceedings of Institute of Math.*, **5**, Nauka, Novosibirsk, 56-75 (in Russian).

[119] Utev S. A. Sums of random variables with φ-mixing. *Proceedings of Institute of Math.*, **13**, Nauka, Novosibirsk, 1989, p. 78–100 (in Russian).

[120] Von Bahr, B. On the convergence of moments in the central limit theorem, *Ann. Math. Statist.*, **36**, No. 3, 808–818 (1965).