NKFIH K 108615 Final report

Random graphs and combinatorial processes

In the paper [1] we examine a random structure consisting of objects with positive weights and evolving in discrete time steps. It generalizes certain random graph models. We prove almost sure convergence for the weight distribution and show scale-free asymptotic behaviour. Martingale theory and renewal-like equations are used in the proof.

In the paper [2] we examine the asymptotics of a renewal-like recursion and a similar integral equation. The motivation comes from probability theory. More precisely, in a random model of publication activity the asymptotic distribution of the weights of the authors satisfy such equations. We prove that the solution is polinomially decaying under suitable conditions.

In [8] we deal with a random graph model evolving in discrete time steps by duplicating and deleting the edges of randomly chosen vertices. We prove the existence of an a.s. asymptotic degree distribution, with stretched exponential decay; more precisely, the proportion of vertices of degree d tends to some positive number $c_d > 0$ almost surely as the number of steps goes to infinity, and $c_d \sim (e\pi)^{1/2} d^{1/4} e^{-2\sqrt{d}}$ holds as $d \to \infty$.

In [17] we deal with a random graph model where at each step, a vertex is chosen uniformly at random, and it is either duplicated or its edges are deleted. Duplication has a given probability. We analyse the limit distribution of the degree of a fixed vertex, and derive a.s. asymptotic bounds for the maximal degree. The model shows a phase transition phenomenon with respect to the probabilities of duplication and deletion.

Consider an increasing group of individuals who are organized by pairwise collaborations. A successful collaboration attracts newcomers, who start collaborating with one or both participants. However, the new connections can weaken and exhaust the attracting pair's collaboration, which eventually ceases. In the paper [29] we investigate the corresponding random graph process in the framework of general time-dependent branching processes. We have results on the life history of an edge. We compute the survival function and the expectation of the life span of an edge, and the distribution of the number of edges considered its direct descendants. We analyze the evolution of the graph process: the probability of extinction, the growth rate of the numbers of edges and vertices, and the number of birth or death events. We also deal with the history of a vertex: how its degree varies with time.

In paper [31] we considered a question on the limit and the speed of convergence of certain dense preferential attachment graphs. This graph model is based on Pólya urns and can be viewed as a dense analogue of the well-known Barabási-Albert graph. We proved an upper bound of $O(n^{-1/3} \log^2 n)$ for the distance of the preferential attachment graph and a certain random graph associated to the limit, in the so-called jumble distance of multigraphs. The tools of the proof mostly come from the theory of martingales and urn models.

Random walks and their limits

The study of properties of random walks has been extensively investigated in the literature started with Pólya (1921), Dvoretzky and Erdős (1951), Erdős and Taylor (1960), etc. In this project we have studied the properties of the so-called anisotropic random walks on the plane. In the paper [15] we investigated recurrence vs transience properties, local times, range, etc. In [32] we compared the fixed and random column configurations for transport phenomena earlier investigated by Heyde (1982, 1993) and den Hollander (1994).

An important particular case of anisotropic random walk is the random walk on the comb structure that can be obtained from square lattice on the plane by removing all horizontal lines except the x-axis. In [16] the largest square covered by a random walk on comb was investigated. For R_n, the largest integer for which the square $[-R_n, R_n]^2$ is covered by the random walk at time n, upper and lower bounds are given of order $n^{1/4}$ with some logarithmic factors.

In the paper [26] we investigated the distance between two or more independent random walks on some graphs after n steps. We have shown that for d-dimensional simple random walks, if $K > 1 + 3/d$, the maximal distance between K random walkers after n steps is of order $n^{1/2}$, asymptotically. Similar problems were investigated for random walks on the 2-dimensional comb. Krishnapur and Peres (2004) have shown that 2 random walkers meet only finitely often with probability 1. We have shown that the maximal distance between them after n steps is roughly of order $n^{1/4}$. In the case of 3 or more walkers the distance is of order $n^{1/2}$.

In [20] and [33] we investigated the properties of random walks on a spider that is a collection of half lines on the plane, starting from one point. Strong approximation of this random walk by the so-colled Brownian spider is proved. Weak and strong limit theorems for the heights of random walk on the half lines are investigated. In the paper [33] weak and strong limit theorems are proved for the local and occupation times. A particular case is the so-called skew random walk and skew Brownian motion for which the results are also new.

Extremal theory of independent and dependent processes

Let $X_1, X_2, \ldots$ be i.i.d. random variables with $P(|X_1| > t) \sim ct^{-\alpha}$ as $t \to \infty$ for some constant $c > 0$ and $0 < \alpha < 2$, let $S_n = \sum_{k=1}^{n} X_k$. Then under a simple "balancing" condition the normed partial sum $n^{-1/\alpha} S_n$, suitably centered, converges weakly to a stable limit distribution with parameter α. Unlike in the case of the central limit theorem for finite variances, the behavior of the partial sum S_n is influenced strongly by its individual terms; for example, removing from S_n its d largest terms for a fixed $d \geq 1$ ("trimming"), changes the limit distribution. This fact causes considerable difficulties in extending standard statistical methods from finite to infinite variances. However, Csörgő, Horváth and Mason (1986) proved that if from the sum S_n we remove its largest d_n terms where $d_n \to \infty$, $d_n/n \to 0$, then the remaining sum will satisfy the central limit theorem and this fact was used in Berkes, Horváth and Schauer (2011) to extend the classical CUSUM technique for change point detection to the case of processes with infinite variances. In [3] we showed that the result of Csörgő, Horváth and Mason remains valid, with random centering, for modulus trimming as well (i.e when we remove from S_n the d_n terms with the largest moduli), clarifying and old paradox from the 1980's. In [4] we extended the trimming CLT for autoregressive processes with stable innovations, providing the first dependent trimming theorem. The latter result was applied in [9] for change point theory of autoregressive processes.

Trimming also plays an important role in analysis. Khinchin (1935) proved that the "digits" a_k in the continued fraction expansion $[a_1, a_2, \dots]$ of a uniform random number α in $(0, 1)$ satisfy the weak law of large numbers with norming $n \log n$, but not the strong law. Diamond and Vaaler (1986) proved that if from the sum $S_n = \sum_{k=1}^n a_k$ we remove the largest term, the strong law will hold, so the obstacle to the strong law is a single large term a_k in the sum. In [18] we proved that if from the sum S_n we remove the d_n largest terms, where $d_n \to \infty$, $d_n/n \to 0$ (a "small" portion of the terms), then the order of magnitude of the remaining sum will get much smaller and it will satisfy the central limit theorem. This fact promises new insight into the metric theory of continued fractions and explains, in particular, curious phenomena for the discrepancy of the sequence $\{n\alpha\}$ observed recently in Setokuchi and Takashima (2014).

Strong approximation of Bernoulli shift processes

A deep result of probability theory is the approximation theorem of Komlós, Major and Tusnády (1975) stating that partial sums S_n of independent, identically distributed random variables $X_1, X_2, \dots$ with mean 0 and finite variance can be approximated by a Wiener process with an a.s. remainder term $o(n^{1/p})$ or $O(\log n)$ according as X_1 has a finite p-th moment ($p > 2$) or a finite moment generating function in a neighborhood of 0. Both remainder terms are optimal. This theorem provides a powerful tool for proving limit theorems for i.i.d. random variables and thus it would have great theoretical and practical importance to extend it for dependent random variables. However, no such results were proved until recently, when in [5] we extended the L^p version of the KMT results for a class of Bernoulli shift processes. Recently, Merlevede and Rio (2015) obtained an extension for Markov chains.

The St. Petersburg paradox

The St. Petersburg paradox (Bernoulli 1738) concerns the fair entry fee in a game where the winnings are distributed as $P(X = 2^k) = 2^{-k}$ ($k = 1, 2, \dots$). The difficulty of the problem is due to the fact that the tails of X are not regularly varying in the sense of Karamata and the sequence S_n of accumulated gains has a curious divergent behavior, with a class of semistable distributions as the subsequential limits of normed partial sums. (See Martin-Löf (1985) and Csörgő and Dodunekova (1991).) In [25] we gave a new extremal representation of the limiting semistable laws of the St. Petersburg game, leading to important asymptotic information on the tails of St. Petersburg sums and their limits. In [23] we proved a strong approximation for partial sums of St. Petersburg variables with the corresponding semistable process, providing the first example for a strong invariance principle in which the approximating continuous time process is not a stable process.

The subsequence principle

It is a well known fact that sufficiently thin subsequences of any (dependent) sequence of random variables behave like independent random variables. The following heuristic principle was formulated by Chatterji (1972):

Subsequence Principle. *Let T be a probability limit theorem valid for all sequences of i.i.d. random variables belonging to an integrability class L defined by the finiteness of a norm $\| \cdot \|_L$. Then if (X_n) is an arbitrary (dependent) sequence of random variables satisfying $\sup_n \|X_n\|_L < +\infty$ then there exists a subsequence (X_{n_k}) satisfying T in a randomized form.*

In a deep paper Aldous (1977) verified this principle for all weak (distributional) and strong (almost sure) limit theorems subject to some formal conditions, thereby unifying a theory going back to nearly a century in the form of a single theorem. Note, however, that many important applications of this principle in analysis are not covered by Aldous' result. For example, the famous result of Menshov (1936) stating that every orthonormal system (f_n) has a subsequence (f_{n_k}) such that $\sum_{k=1}^\infty c_k f_{n_k}$ converges a.e. provided $\sum_{k=1}^\infty c_k^2 < \infty$ requires the selection of a subsequence (f_{n_k}) whose weighted sums have a strong convergence property simultaneously for all weight sequences (c_k) and this does not follow from the theory of Aldous (1981). Similarly, an i.i.d. sequence is a symmetric structure and thus one can expect that limit theorems for thin subsequences of general random variable sequences remain valid after any permutation of their terms, but Aldous' theorem does not yield uniformity over permutations. In our papers [12], [13], [19] we proved limit theorems for lacunary series providing such a uniformity. In [13] we proved a permutation-invariant form of Aldous' theorem, leading to new applications in analysis. In [12] we proved uniform limit theorems for weighted sums $\sum_{k=1}^N c_k f_{n_k}$ in the case when (f_n) are not square integrable and the limit distribution of the normed partial sums of (f_{n_k}) is a stable law. Finally, in [19] we solved a long standing open problem in Banach space theory by giving a necessary and sufficient condition that a sequence (x_n) in $L^p(0,1)$, $p \geq 1$ has a subsequence (x_{n_k}) spanning a subspace isomorphic to Hilbert space, another uniform limit theorem escaping the general theory.

Strong dependence in analysis and number theory

Kac (1946) proved that if f is a 1-periodic Lipschitz function with $\int_0^1 f(x)dx = 0$, then $\sum_{k=1}^N f(2^k x)$ satisfies the central limit theorem and in later years many other asymptotic properties of independent random variables have been established for the sequence $f(2^k x)$. On the other hand, as Erdős and Fortet (1949) showed, the CLT fails for $f(n_k x)$ if $n_k = 2^k - 1$, and thus for exponentially growing n_k the almost independent behavior of $f(n_k x)$ provided by the subsequence principle is modified by the number theoretic properties of n_k. For sequences n_k growing subexponentially (and in particular for the whole sequence) almost independence breaks down completely and the sequence becomes strongly dependent, making its asymptotic study much harder. In our papers [7] and [10] we proved a few results in this case, specifically, we investigated the almost everywhere convergence of series

$$\sum_{k=1}^{\infty} c_k f(kx) \tag{1}$$

where f is a periodic measurable function with zero integral on its period interval. By Carleson's theorem, in the case $f(x) = \sin x$ the series (1) converges almost everywhere provided $\sum_{k=1}^{\infty} c_k^2 < \infty$ and Gaposhkin (1968) showed that this remains valid if the Fourier series of f is absolutely convergent, in particular if f is a Lip α function with $\alpha > 1/2$. Apart from these results, no satisfactory convergence criteria for (1) exist even for function classes like $C(0,1)$, $L_p(0,1)$, BV (functions with bounded variation on $(0,1)$), etc. In [7], [10], we gave optimal or near optimal criteria for some of these classes, settling long standing open problems in analysis. In [10] we proved that (1) converges a.e. if $f \in$ BV and

$$\sum_{k=1}^{\infty} c_k^2 (\log\log k)^{\gamma} < \infty$$

for $\gamma > 4$ and this is not valid for $\gamma \le 2$. The case $2 < \gamma \le 4$ remained open and was settled recently by Lewko and Radziwill (2017), who showed that the series converges a.e. in this case. In [7] we gave a sharp criterion for the a.e. convergence of (1) for the class C_α of functions with Fourier coefficients $O(k^{-\alpha})$ $(1/2 < \alpha < 1)$, namely

$$\sum_{k=1}^{\infty} c_k^2 \exp\left(\frac{K(\log k)^{1-\alpha}}{\log\log k}\right) < \infty$$

for a suitable constant $K > 0$. The proofs in [7], [10] depend on a new estimate on GCD sums

$$\sum_{k,\ell=1}^{N} \frac{(n_k, n_\ell)^{\alpha}}{[n_k, n_\ell]^{\alpha}}$$

for arbitrary sequences (n_k) of integers, where $0 < \alpha < 1$, see [10], Theorem 1. Our estimate improves results of Dyer and Harman (1986) and is sharp except the case $\alpha = 1/2$, completing a long development in number theory starting with works of Erdős, Gál and Koksma in the 1930/1940's. The case $\alpha = 1/2$ was settled recently by Bondarenko and Seip (2015, 2016), leading to breakthrough results on the maxima of the Riemann zeta function on the line $1/2 + it$.

As a further application of the GCD sum estimate in [10], we proved that for $f \in$ BV we have for any increasing sequence (n_k) that

$$\left|\sum_{k=1}^{N} f(n_k x)\right| = O((N\log N)^{1/2}(\log\log N)^{\gamma}) \qquad \text{a.e.} \tag{2}$$

for $\gamma > 5/2$, but not if $\gamma \le 1/2$. This result provides the best known result in an old problem of analysis. Erdős and Koksma (1949) and Cassels (1950) proved that for any increasing sequence (n_k) we have

$$\left|\sum_{k=1}^{N} f(n_k x)\right| = O(N^{1/2}(\log N)^{5/2+\varepsilon}) \quad \text{a.e.} \tag{3}$$

uniformly for all periodic, mean zero $f \in \mathrm{BV}$ with $\mathrm{Var}\, f \le 1$. R. C. Baker (1981) improved the exponent 5/2 in (3) to 3/2 and Berkes and Philipp (1994) showed that the bound $O(N^{1/2}(\log N)^{1/2})$ is not valid even for a single f. The gap between the exponents 1/2 and 3/2 remains open. As the result (2) shows, for fixed $f \in \mathrm{BV}$ the gap between the exponents 1/2 and 3/2 can be closed, the critical exponent is 1/2.

Function series with random gaps

By a classical result of Salem and Zygmund (1947), under the Hadamard gap condition $n_{k+1}/n_k \ge q > 1$ $(k = 1, 2, \dots)$, the sequence $\sin n_k x$ satifes the central limit theorem with respect to the Lebesgue measure, but for general (n_k) very little is known about the asymptotic behavior of $\sum_{k=1}^{N} \sin n_k x$. The same applies for dilated sums $\sum_{k=1}^{N} f(n_k x)$ for periodic measurable f and thus it is natural to consider random sequences (n_k). A simple random model is when the gaps $n_{k+1} - n_k$ are independent, identically distributed positive random variables, i.e. when (n_k) is an increasing random walk. In [11] we studied the case when (n_k) is an absolutely continuous random walk, i.e. when n_1 is random variable with density. We showed that for this model the discrepancy D_N of $\{n_k x\}_{1 \le k \le N}$ and its L^p version $D_N^{(p)}$ not only satisfy the law of the iterated logarithm, but we also described the precise asymptotic behavior of the empirical process of the sequence $\{n_k x\}_{1 \le k \le N}$. In [24] we gave a random construction to show that the order of magnitude of the discrepancy of $\{n_k x\}_{1 \le k \le N}$ can be any smooth function between $N^{-1}(\log N)(\log\log N)^{1+\varepsilon}$ and $N^{-1/2}(\log\log N)^{1/2}$. According the well known results of W. Schmidt (1972) and Philipp (1975), both bounds are nearly optimal.

Mathematical models in theoretical biology

Game theory focuses on payoffs and typically ignores time constraints that play an important role in the evolutionary processes where the repetition of games can depend on the strategies, too. In [27] we introduce a stochastic model, where each pairwise interaction has two consequences: each player receives a payoff and must wait a random time before playing the next game. Thus the introduced matrix game under time constraint is defined by two matrices: a payoff matrix and an average time duration matrix. We adapt Maynard Smith's concept of evolutionary stability for this class of games and give a full characterization of evolutionary stable strategies (ESS).

In [28] we consider and infinitely large asexual population without mutations and direct interactions. The activities of an individual determine the fecundity and the survival probability of individuals, moreover each activity takes time. We view this population model as a simple combination of life history and optimal foraging models. The phenotypes are given by probability distributions on these activities. We concentrate on the following phenotypes defined by optimization of different objective functions: selfish individual (maximizes the average offspring number during life span), survival phenotype (maximizes the probability of non-extinction of descendants) and Darwinian phenotype (maximizes the phenotypic growth rate). By using general time-dependent branching processes we find that the objective functions above can achieve their maximum at different activity distributions, in general. The novelty of our work is that we let natural selection act on the different objective functions. Using the classical Darwinian reasoning, we show that in our selection model the Darwinian phenotype outperforms all other phenotypes.

Starting from the results of [27], where we introduced the class of matrix games under time constraints and characterized the concept of (monomorphic) ESS, in [34] we are interested in how the ESS is related to the existence and stability of equilibria for polymorphic populations. We point out that, although the ESS may no longer be a polymorphic equilibrium, there is a connection between them. Specifically, the polymorphic state at which the average strategy of the active individuals in the population is equal to the ESS is an equilibrium of the polymorphic model. Moreover, in the case when there are only two pure strategies, a polymorphic equilibrium is locally asymptotically stable under the replicator equation for the pure-strategy polymorphic model if and only if it corresponds to an ESS. Finally, we prove that a strict Nash equilibrium is a pure-strategy ESS that is a locally asymptotically stable equilibrium of the replicator equation in n-strategy time-constrained matrix games.

Further results

In the paper [6] Chebyshev type upper bounds are presented for the probability that a random vector$X$ falls outside of an ellipsoid, in terms of the mean and variance of X. Our inequalities are better than those found in the literature.

The paper [22] is a contribution to the problem of estimating the deviation of two discrete probability distributions in terms of the supremum distance between their generating functions over the interval $[0, 1]$. Deviation can be measured by the difference of the kth terms, or by total variation distance. Our new bounds have better order of magnitude than those proved previously, and they are even sharp in certain cases.

In the paper [30] we prove that all multivariate random variables with finite variances are univariate functions of uncorrelated random variables and if the multivariate distribution is absolutely continuous then these univariate functions are piecewise linear. They can be independent of the correlations in the Gaussian case.

The subject of the paper [14] is to investigate the asymptotic properties of Bahadur-Kiefer processes for partial sums and renewals of dependent random variables. In the long range dependent case the limit of partial sums is fractional Brownian motion. It is shown that in this case the Bahadur-Kiefer process can be expressed as some increment of the fractional Brownian motion. In [21] investigations of the so-called Vervaat process are extended to the case of partial sums of long range dependent random variables. The Vervaat process is the integral of Bahadur-Kiefer process. It is shown that the limit of the appropriately normalized Vervaat process is the square of fractional Brownian motion.