

Abstracts of the lectures

IMRE CSISZÁR

Closures of exponential families

An *exponential family* \mathcal{E} of probability measures (pm's) on a measurable space $(\mathcal{X}, \mathcal{F})$, with dominating measure μ and canonical statistic $f : \mathcal{X} \mapsto R^d$, is

$$\mathcal{E} = \left\{ Q_\vartheta : \frac{dQ_\vartheta}{d\mu}(x) = e^{\langle \vartheta, f(x) \rangle - \Lambda(\vartheta)}, \vartheta \in \text{dom}(\Lambda) \right\}$$

where

$$\Lambda(\vartheta) = \ln \int e^{\langle \vartheta, f(x) \rangle} \mu(dx), \quad \text{dom}(\Lambda) = \{ \vartheta : \Lambda(\vartheta) < \infty \}.$$

The closures $cl_v(\mathcal{E})$ and $cl_{rI}(\mathcal{E})$ of \mathcal{E} in variation distance and reverse information divergence (Kullback-Leibler distance) consist of those pm's Q to which there exist pm's $Q_n \in \mathcal{E}$ such that $|Q_n - Q|$ respectively $D(Q||Q_n)$ approaches 0. Note that $cl_{rI}(\mathcal{E})$ is not necessarily rI-closed [4]. The rI-closure is relevant in the context of maximum likelihood estimation.

Theorem 1. *Given a sample from an unknown $Q \in \mathcal{E}$, let $a \in R^d$ denote the sample mean of the canonical statistic f . If the supremum $\Lambda^*(a)$ of the normalized log-likelihood function*

$$L(Q) = \langle \vartheta, a \rangle - \Lambda(\vartheta) \quad \text{if} \quad Q = Q_\vartheta$$

is finite, there exists a unique pm $Q_a^ \in cl_{rI}(\mathcal{E})$ such that*

$$D(Q_a^*||Q) \leq \Lambda^*(a) - L(Q) \quad \text{for each} \quad Q \in \mathcal{E}.$$

Thus, any ϵ -MLE (any $Q \in \mathcal{E}$ with $L(Q) \geq \Lambda^*(a) - \epsilon$) belongs to the "reverse I-divergence ball" with center Q_a^* and radius ϵ .

Remark 1. A sufficient condition for $\Lambda^*(a) < \infty$, also necessary when $\text{dom}(\Lambda) = R^d$, is $a \in cc(\mu_f)$. Here μ_f denotes the f -image of μ on R^d , and $cc(\mu_f)$ denotes its *convex core* [2], the intersection of all convex Borel sets in R^d of full μ_f -measure.

The *extension* $ext(\mathcal{E})$ of an exponential family \mathcal{E} is defined [3] as the union, for all faces F of $cc(\mu_f)$, of the exponential families

$$\mathcal{E}^F = \{ Q_{F,\vartheta} : \vartheta \in \text{dom}(\Lambda^F) \}$$

with canonical statistic f whose dominating measure is the restriction of μ to $f^{-1}(cl(F))$; here Λ^F denotes the analogue of Λ above.

Theorem 2. $cl_{rI}(\mathcal{E}) \subseteq cl_v(\mathcal{E}) \subseteq ext(\mathcal{E})$. If $dom(\Lambda) = R^d$, the equalities hold. Each pm in $textcl_v\mathcal{E}$ for which f is integrable belongs also to $cl_{rI}(\mathcal{E})$.

Remark 2. If the convex core of μ_f has no proper faces then $ext(\mathcal{E}) = \mathcal{E}$ and therefore \mathcal{E} is closed in variation distance. This happens, e.g., for the familiar continuous type families, thus mixed type families are of more interest for this talk; the discrete case when μ_f is concentrated on a finite set has been settled already in [1].

The following main result provides complete description of the variation closure of any exponential family, as well as of any subfamily $\mathcal{E}(\Theta) = \{Q_\vartheta : \vartheta \in \Theta\}$ with convex parameter set $\Theta \subseteq dom(\Lambda)$. For a set $F \subset R^d$, $ri(F)$ denotes its relative interior, and π_F denotes orthogonal projection to the subspace of R^d spanned by the differences of elements of F .

Theorem 3. [5]. *For any convex $\Theta \subseteq dom(\Lambda)$, the variation closure of $\mathcal{E}(\Theta)$ intersects exactly those components \mathcal{E}^F of $ext(\mathcal{E})$ for which either $F = cc(\mu_f)$ (thus $\mathcal{E}^F = \mathcal{E}$) or there exists a chain*

$$cc(\mu_f) = F_0 \subset \cdots \subset F_m = F, \quad F_{i+1} = F_i \cap H_i, \quad 0 \leq i < m$$

such that H_i is a supporting hyperplane of F_i whose normal vector is a direction of recession of $\pi_{F_i}(ri(\Theta))$. Then

$$cl_v(\mathcal{E}(\Theta)) \cap \mathcal{E}^F = \{Q_{F,\vartheta} : \vartheta \in cl(\pi_F(\Theta)) \cap dom(\Lambda^F)\}.$$

References

- [1] Barndorff-Nielsen, O., *Information and Exponential Families in Statistical Theory*. Wiley, New York, 1978.
- [2] Csiszár, I. and Matúš, F., Convex cores of measures on \mathbb{R}^d , *Studia Sci. Math. Hungar.* **38**, pp. 177–190, 2001.
- [3] Csiszár, I. and Matúš, F., Information projections revisited, *IEEE Trans. Inform. Theory* **49**, pp. 1474-1490, 2003.
- [4] Csiszár, I. and Matúš, F., On information closures of exponential families: a counterexample, *IEEE Trans. Inform. Theory* **50**, pp. 922-924, 2004.
- [5] Csiszár, I. and Matúš, F., Closures of exponential families, *Ann. Probability*, to appear.

Laws of large numbers for cooperative St. Petersburg gamblers

It was shown in an earlier paper [S. Csörgő and G. Simons: The two-Paul paradox and the comparison of infinite expectations, In: *Limit Theorems in Probability and Statistics* (I. Berkes, E. Csáki, M. Csörgő, eds.), Vol. I, pp. 427–455. János Bolyai Mathematical Society, Budapest, 2002] that it may be a very different situation when n distinct Pauls play one St. Petersburg game each rather than the classical case, in which one Paul plays n games. Here we consider sequences of arbitrary pooling strategies for n cooperative Pauls, each playing one St. Petersburg game with the same, possibly biased coin, and prove laws of large numbers for their individual winnings as $n \rightarrow \infty$.

PAUL DEHEUVELS

Multivariate Bahadur-Kiefer representations

The classical Bahadur-Kiefer representation in dimension 1 is obtained as follows. Given a sequence U_1, U_2, \dots of independent and identically distributed random variables with a uniform distribution on $(0, 1)$, for each $n \geq 1$, one sets $\mathbb{U}_n(t) = n^{-1} \#\{U_i \leq t : 1 \leq i \leq n\}$ and $\mathbb{V}_n(t) = \inf\{s \geq 0 : \mathbb{U}_n(s) \geq t\}$ for $0 \leq t \leq 1$. Introduce the empirical processes $\alpha_n(t) = n^{1/2}(\mathbb{U}_n(t) - t)$ and $\beta_n(t) = n^{1/2}(\mathbb{V}_n(t) - t)$. Denote by $\|f\| = \sup_{t \in I} |f(t)|$ the sup-norm of a bounded function f on I . Then, Bahadur (1967) and Kiefer (1969, 1970) showed that, with probability 1 as $n \rightarrow \infty$, $\|\alpha_n + \beta_n\| \rightarrow 0$, and established the rate of this convergence. Our aim is to provide multivariate extensions of these results in the setup of d -dimensional random vectors. To simplify notation our exposition is made here for $d = 2$, and we consider a sequence $(U_1^{(1)}, U_1^{(2)}), (U_2^{(1)}, U_2^{(2)}), \dots$ of independent random vectors with a uniform distribution on $(0, 1)^2$. We set accordingly, for each $n \geq 1$ and $j = 1, 2$, $\mathbb{U}_n^{(j)}(t) = n^{-1} \#\{U_i^{(j)} \leq t : 1 \leq i \leq n\}$ and $\mathbb{V}_n^{(j)}(t) = \inf\{s \geq 0 : \mathbb{U}_n^{(j)}(s) \geq t\}$ for $0 \leq t \leq 1$. Set further $\mathbb{U}_n(t_1, t_2) = n^{-1} \#\{U_i^{(1)} \leq t_1, U_i^{(2)} \leq t_2 : 1 \leq i \leq n\}$ and $\alpha_n(t_1, t_2) = n^{1/2}(\mathbb{U}_n(t_1, t_2) - t_1 t_2)$ for $0 \leq t_1, t_2 \leq 1$. We consider the *tied-down* empirical process defined by $\alpha_n^{(0)}(t_1, t_2) = \alpha_n(t_1, t_2) - t_1 \alpha_n(1, t_2) - t_2 \alpha_n(t_1, 1)$ for $0 \leq t_1, t_2 \leq 1$, and compare this process with the *empirical copula process* defined by $\beta_n^{(0)}(t_1, t_2) = n^{1/2}(\mathbb{U}_n(\mathbb{V}_n^{(1)}(t_1), \mathbb{V}_n^{(2)}(t_2)) - t_1 t_2)$. We show that, almost surely, $\|\alpha_n^{(0)} + \beta_n^{(0)}\| \rightarrow 0$ and provide the rate of this convergence. This result turns out to be a direct extension of the original Bahadur-Kiefer representation in dimension 1 mentioned above.

Limit theorems for non-symmetric U -statistics

We study the limit distribution of non-symmetric degenerate U -statistics

$$U_n = \sum_{1 \leq i < j \leq n} h(X_i, X_j),$$

where $(X_i)_{i \geq 1}$ is an i.i.d. sequence with values in some separable metric space S and $h : S \times S \rightarrow \mathbf{R}$ a measurable kernel. Under second moment conditions on the kernel, we establish convergence of $U_{[nt]}/n$ towards a stochastic integral $\int h(x, y)K(t, dx, dy)$, where $K(t, x, y)$ is a three-parameter Gaussian process. An important tool in these investigations is the empirical pair process

$$F_n(t, x, y) := \frac{1}{n} \sum_{1 \leq i < j \leq [nt]} I(X_i \leq x)I(X_j \leq y),$$

whose convergence to a Gaussian process is established. As a byproduct of our investigations we obtain weak convergence of nonsymmetric bilinear forms $\sum a_{ij}^{(n)} \phi(X_i) \psi(X_j)$, for certain weights $(a_{ij}^{(n)})$. (joint work with S. A. Utev)

Almost sure limit theorems

Let (B, ϱ) be a complete separable metric space and ζ_n , $n \in \mathbb{N}$, be a sequence of random elements in B . Let μ_ζ denote the distribution of ζ and let δ_x be the unit mass at x .

Theorem 1. Assume that there exist $C > 0$, $\varepsilon > 0$, an increasing sequence of positive numbers c_n with $\lim_{n \rightarrow \infty} c_n = \infty$, $c_{n+1}/c_n = O(1)$, and B -valued random elements ζ_{kl} , $k, l \in \mathbb{N}$, $k < l$, such that the random elements ζ_k and ζ_{kl} are independent for $k < l$, and

$$\mathbb{E}\{\varrho(\zeta_{kl}, \zeta_l) \wedge 1\} \leq C \{\log_+ \log_+ (c_l/c_k)\}^{-(1+\varepsilon)},$$

for $k < l$. Let $0 \leq d_k \leq \log(c_{k+1}/c_k)$, assume that $\sum_{k=1}^{\infty} d_k = \infty$. Let $D_n = \sum_{k=1}^n d_k$. Then for any probability distribution μ on the Borel σ -algebra of B the following two statements are equivalent

$$\frac{1}{D_n} \sum_{k=1}^n d_k \delta_{\zeta_k(\omega)} \Rightarrow \mu, \quad \text{as } n \rightarrow \infty, \quad \text{for almost every } \omega \in \Omega;$$

$$\frac{1}{D_n} \sum_{k=1}^n d_k \mu_{\zeta_k} \Rightarrow \mu, \quad \text{as } n \rightarrow \infty.$$

Theorem 1 is an extension of the basic results of Berkes and Csáki (2001). Theorem 1 is applied to prove a.s. versions of some functional limit theorems: convergence of Wiener processes, Donsker's theorem, empirical processes (Fazekas and Rychlik (2002)). A multiindex version of Theorem 1 is obtained (Fazekas and Rychlik (2003)). Also an appropriate version of Theorem 1 can be applied to prove almost sure limit theorems for random allocations (Fazekas and Chuprunov (2004)).

References

- Berkes, I. and Csáki, E. (2001). A universal result in almost sure central limit theory. *Stoch. Proc. Appl.* **94**(1), 105–134.
- Fazekas, I. and Rychlik, Z. (2002). Almost sure functional limit theorems. *Ann. Univ. M. Curie-Skłodowska*, **LVI**, no.1, 1–18.
- Fazekas, I. and Rychlik, Z. (2003). Almost sure central limit theorems for random fields. *Math. Nachr.* **259**, 12–18.
- Fazekas, I. and Chuprunov, A. (2004). Almost sure limit theorems for random allocations. Accepted by *Studia. Sci. Math. Hung.*
-

ANTÓNIA FÖLDES

About some of my favorite results with Bandi and Pali

I will talk about two old and little known results and about our most recent work.

The first of the old topics is dealing with the following question. How much can we modify a simple symmetric random walk by introducing some deterministic change in the usual definition to preserve the basic properties.

The second topic is about the length of the longest monotone block in a sample of discrete distribution. It turns out that there is a major difference between the length of longest monotone and the length of the longest strictly monotone block.

Finally I will discuss some results about the maximal local time of a d-dimensional simple random walk on subsets. The characterization of the maximal local time in two and higher dimension is originated by Erdős and Taylor (1960), and major developments were made recently by Dembo, Peres, Rosen and Zeitouni (2001). Here I will talk about the maximal local time on lines and circles in dimension two, and in subspaces and balls in higher dimensions.

Central Limit Problems in Space and Time

We investigate the central limit problem for interacting diffusions on the infinite dimensional sphere, the model was proposed by Stefano Olla. Temporal evolution of the system is given by a set paired stochastic equations,

$$\begin{aligned} dp_k &= (r_k - r_{k-1}) dt - \gamma p_k dt + \sqrt{\gamma} p_{k+1} dw_k - \sqrt{\gamma} p_{k-1} dw_{k-1}, \\ dr_k &= (p_{k+1} - p_k) dt, \end{aligned}$$

where $\gamma > 0$, $p_k, r_k \in \mathbb{R}$ for $k \in \mathbb{Z}$, and $w_k, k \in \mathbb{Z}$ is a family of standard Wiener processes. The deterministic part (drift) of these equations corresponds to a chain of coupled harmonic oscillators (lattice approximation of a vibrating string). Besides $\sum r_k$, also total energy

$$H := \frac{1}{2} \sum_{k \in \mathbb{Z}} (p_k^2 + r_k^2)$$

is preserved, and the Gibbs field with energy H is a stationary state of the process. We investigate the limit distribution of the fluctuation fields

$$\begin{aligned} R_\epsilon(t, \phi) &:= \sqrt{\epsilon} \sum_{k \in \mathbb{Z}} \phi(k\epsilon) r_k(t\epsilon^{-2}), \\ H_\epsilon(t, \phi) &:= \frac{\sqrt{\epsilon}}{2} \sum_{k \in \mathbb{Z}} \phi(k\epsilon) (p_k^2(t\epsilon^{-2}) + r_k^2(t\epsilon^{-2}) - 1) \end{aligned}$$

as $0 < \epsilon \rightarrow 0$ in this stationary regime. Due to the Hamiltonian part of the equations of motion, this problem is fairly singular. Nevertheless, the limit distribution of fluctuations can be identified as the measure of a couple of linear stochastic partial differential equations, namely

$$\partial_t \rho = \gamma^{-1} \partial_x^2 \rho + \gamma^{-1/2} \partial_x J_1, \quad \partial_t \chi = c \partial_x^2 \chi + \sqrt{2c} \partial_x J_2,$$

where $c = (1 + \gamma^2)/2\gamma$, J_1 and J_2 are independent white noise processes in space and time.

EDIT GOMBAY

On Sequential and Group-Sequential test in Clinical Trials

In large clinical trials data is often gathered over years, even decades, so repeated analyses during that period is very important for ethical and financial reasons. This demand introduces the model of censored data with random staggered entry, which is probabilistically a stochastic process with two-dimensional time variables: calendar time and time-on-test. Beginning with Tsiatis (1982) and Slud and Wei (1982) results for large sample approximations in discrete time, and then in continuous time appeared by several authors (e.g. Slud, Selke and Siegmund, Gu and Lai). These results enable researchers to analyse data with group-sequential technology, that is the performance of repeated fixed sample size tests. With the help of strong approximation methods we can now propose a family of fully sequential testing procedures, which allow continuous monitoring. The family of tests is indexed by weight functions, giving tests with different performance. Examples include parametric and non-parametric test, such as rank statistics using the so-called log-rank score

YUEYUN HU

Invariance principles for excursion lengths and heights

This talk is based on a joint work with E. Csáki: we prove strong invariance principles between ranked excursion lengths and heights of a simple random walk and those of a standard Brownian motion; some consequences concerning limiting distributions and strong limit theorems will also be presented.

GONCHIGDANZAN HURELBAATAR

Almost sure central limit theorem for the products of the partial sums

In this talk we present an almost sure limit theorem for the products of the partial sums of i.i.d positive random variables. Some results on the almost sure limit theorems for nonlinear functionals of independent random variables and its extensions to dependent random variables are also discussed.

An occupation time theorem for a class of one-dimensional diffusion processes

Let $\{B_t, P^x\}$ be a one-dimensional standard Brownian motion and let

$$\Gamma_+(t) = \int_0^t 1_{\{B_s > 0\}} ds \quad t \geq 0.$$

Thus $\Gamma_+(t)$ denotes the sojourn time up to time t on the half line $(0, \infty)$ and the following fact is well-known as Lévy's arc-sine law.

$$P^0\left(\frac{1}{t}\Gamma_+(t) \leq x\right) = P^0(\Gamma_+(1) \leq x) = \frac{2}{\pi} \arcsin \sqrt{x}, \quad 0 \leq x \leq 1.$$

Many authors have tried to extend this result for more general stochastic processes, and in this talk we are interested in linear diffusions. A typical, interesting example is the case of *skew Bessel diffusion processes* and for this case Barlow-Pitman-Yor [1] obtained the law of $\Gamma_+(t)$ explicitly. In connection with this result S. Watanabe [2] determined all possible limiting distributions as $t \rightarrow \infty$ of $\Gamma_+(t)/t$ for general linear diffusion processes. He proved that the class of random variables ξ obtained as a limit in law of $\Gamma_+(t)/t$ is given as follows:

- A) Constant random variables $\xi \equiv p$, $0 \leq p \leq 1$,
- B) $\{0,1\}$ -two valued random variables ξ ,
- C) $\Gamma_+(1)$ of the skew Bessel diffusions.

The first two cases may look less interesting but in fact they include important classes of diffusions and it might be of interest to find meaningful limiting distributions under other normalizations.

In this talk we shall study the case B) and will see that, under a suitable nonlinear normalization, we have a meaningful limiting distribution.

References

1. M.T.Barlow, J.W.Pitman and M.Yor: Une extension multidimensionnelle de la loi de l'arc sinus, *Séminaire de Prob.XXIII, ed J.Azéma, P.A.Meyer et M.Yor*, **LNM 1372**, Springer(1989), 294–314
2. S.Watanabe: Generalized arc-sine laws for one-dimensional diffusion processes and random walks, *Stochastic analysis(Ithaca,NY,1993)*, 157–172, Proc. Sympos. PureMath., **57**, Amer. Math. Soc., Providence, RI, 1995

Sharp estimates on the tail behaviour of some random integrals and their application in statistics

I met the problems discussed here in an investigation when I tried to adapt the methods of maximum likelihood estimates to some non-parametric problems. One can give a good and simple approximation for the error of the maximum likelihood estimate by means of an appropriate linearization in the maximum-likelihood equation. It can be shown that a really good approximation is obtained in such a way with the help of the observation that the coefficient of the second term in the Taylor expansion we apply in this linearization is bounded. This linearization argument can be adapted to the study of several interesting non-parametric problems, but in the non-parametric case some multiple random integrals have to be bounded instead of Taylor coefficients. This led to the following problem described below.

Let us have a sequence of iid. random variables ξ_1, \dots, ξ_n on a space (X, \mathcal{X}) with distribution μ , and let μ_n denote their empirical distribution. Given a real valued function $f(x_1, \dots, x_k)$ of k variables on the space (X, \mathcal{X}) consider the k -fold random integral

$$J_{n,k}(f) = \frac{n^{k/2}}{k!} \int' f(u_1, \dots, u_k) (\mu_n(du_1) - \mu(du_1)) \dots (\mu_n(du_k) - \mu(du_k)),$$

where prime means that the diagonals are omitted from the domain of integration. We want to give a good bound on the probability $P(|J_{n,k}(f)| > x)$ for all $x > 0$. More generally, given a nice class of functions $f \in \mathcal{F}$ of k variables we are interested in the probability $P\left(\sup_{f \in \mathcal{F}} |J_{n,k}(f)| > x\right)$ for $x > 0$.

The tail-behaviour of $J_{n,k}(f)$ is similar to that of the k -th power of a Gaussian random variable with expectation zero and variance of the same order as the k -th root of the variance of the random variable $J_{n,k}(f)$ provided that this variance is not too small. More explicitly, $P(|J_{n,k}(f)| > x) < Ce^{-B(x/\sigma)^{2/k}}$ with some universal constants $C > 0$ and $B > 0$ and $\sigma^2 = \int f^2(u_1, \dots, u_k) \mu(du_1) \dots \mu(du_k)$ if the absolute value of the function f is bounded by 1, and $0 < x < n^{k/2} \sigma^{k+1}$. Beside this, the variance of $J_{n,k}(f)$ has the same order as σ^2 . The same estimate holds with possibly different universal constants $C > 0$ and $B > 0$ for the probability $P\left(\sup_{f \in \mathcal{F}} |J_{n,k}(f)| > x\right)$ if \mathcal{F} is a nice class of functions of k variables whose elements are such functions which are bounded in supremum norm by 1 and in L_2 norm by σ . Such a result holds for instance if \mathcal{F} is a so-called Vapnik–Červonenkis class of functions. The only additional restriction we have to impose for the validity of such an estimate is the condition $(\frac{x}{\sigma})^{2/k} > D \log n$ with some $D > 0$ to exclude the possibility that the supremum of relatively small random variables be large.

References

- Major, P. and Rejtő, L. (1998) A note on nonparametric estimations. In the conference volume to the 65. birthday of Miklós Csörgő.759–774
- Major, P. (2003) An estimate about multiple stochastic integrals with respect to a normalized empirical measure. submitted to *Probability Theory and Related Fields*
- Major, P. (2003) An estimate on the maximum of a nice class of stochastic integrals. submitted to *Probability Theory and Related Fields*,
-

SRI GOPAL MOHANTY

Combinatorial methods in transient behaviour of queues

Random walk representation of Markovian queues to study the transient behaviour is a well-known technique that allows us to use combinatorial methods. Starting with M/M/1 model, we review the transition from the traditional analytic approach to a combinatorial approach and state a few expressions alternative to the standard one involving modified Bessel function. If the queue is nonMarkovian, say of the type M/G/1 or G/M/1, we refer to busy period solutions through an elegant combinatorial result as done by Takacs. At the end, we discuss how enumeration of random walk paths for deriving transient solutions becomes easier by considering a dual process corresponding to queues involving batch services.

PAP GYULA

Maximum likelihood estimation of parameters of a discrete time forward interest rate curve model

The most famous model for forward rate structures is due to Heath, Jarrow and Morton (1990). In their model, the forward rate processes belonging to different time points are driven by the same Wiener process, and it is not realistic. We follow the idea of Kennedy (1994), Goldstein (2000) and Santa-Clara and Sornette (2001) by introducing a random driving field instead of a single driving process. In this paper, we consider a discrete time forward interest rate model, where the forward rates corresponding to different times to

maturity are driven by a random field, which has been built up by a system of i.i.d. random variables in an autoregressive way. To make the models realistic, one has to assure that the market excludes arbitrage opportunities. We derive no-arbitrage conditions and assume that they are satisfied. Assuming moreover that the volatility parameter is deterministic and independent of time and of maturity, and that the i.i.d. random variables building the driving random fields are normally distributed, we show asymptotic normality of the maximum likelihood estimators of four parameters. We also show that the maximum likelihood estimators of two parameters are strongly consistent.

WALTER PHILIPP

A uniform Chung-Smirnov LIL for monotone subsequences

Let $(X_k, k \geq 1)$ be a sequence of independent random variables, each having uniform distribution over $[0, 1]$. Let \mathcal{A} be a collection of subsequences of the set of positive integers. For each $N \geq 1$ denote \mathcal{A}_N the collection of the restriction of these subsequences to $[1, 2, \dots, N]$, i.e.

$$\mathcal{A}_N = \{A \cap [1, 2, \dots, N] : A \in \mathcal{A}\}$$

For $0 \leq r \leq 1$ let $\mathcal{A}_N(r) \subset \mathcal{A}_N$ be the set of all (finite) sequences $A \in \mathcal{A}_N$ such that $\text{card}(A) \geq rN$. Denote

$$\psi(N, r) = \text{card } \mathcal{A}_N(r).$$

Suppose there is a constant C (not depending on N, r) such that

$$\log \psi(N, r) \leq Cr^{-1/2} \quad 0 \leq r \leq 1.$$

Then we have with probability 1

$$\sup_{0 \leq x \leq 1} \sup_{A \in \mathcal{A}} \left| \sum_{m_k \leq N, m_k \in A} (1\{X_{m_k} \leq x\} - x) \right| \ll (N \log \log N)^{1/2}.$$

If $\mathcal{A} = Z^+[z]$ the set of all polynomials with non-negative integer coefficients then $\psi(N, r) \ll r^{-2}$. In this case the conclusion of the theorem remains valid for weakly dependent random variables, such as stationary strongly mixing sequences and sequences $\{n_k \omega\}$, where (n_k) satisfies a Hadamard gap condition.

LÍDIA REJTÓ

Tree Clustering of Hidden Variables

DNA microarray data provide high dimensional data with small amount of repetitions. Another difficulty is that most cases we do not have independent repetitions gene expressions are measured under different conditions and under the same condition the measurements coming from different subjects. Traditional data analysis approaches do not work. We propose a parametric model to find hidden variables, (real expression levels) based on the measured ones. Using the estimated variables we construct cluster tree of genes. The simplest case is an extension of the k-mean clustering method. The problem is to find a tree of the cluster centers with bounded sum of length of edges and the L_2 distance of the data from the edges of the tree is the smallest. The tree shows the structure of the clusters and possible connections among different clusters and it may help to identify networks of genes.

JAY ROSEN

An almost sure invariance principle for the range of planar random walks

For a symmetric random walk in Z^2 with $2 + \delta$ moments and range $\mathcal{R}(n)$, we represent $|\mathcal{R}(n)|$, the cardinality of $\mathcal{R}(n)$, in terms of an expansion involving the renormalized intersection local times of a Brownian motion. We show that for each $k \geq 1$

$$(\log n)^k \left[\frac{1}{n} |\mathcal{R}(n)| + \sum_{j=1}^k (-1)^j \left(\frac{1}{2\pi} \log n + c_X \right)^{-j} \gamma_{j,n} \right] \rightarrow 0, \quad a.s.$$

where W_t is a Brownian motion, $W_t^{(n)} = W_{nt}/\sqrt{n}$, $\gamma_{j,n}$ is the renormalized intersection local time at time 1 for $W^{(n)}$, and c_X is a constant depending on the distribution of the random walk.

On perpetual integral functionals of Brownian motion with drift

This talk is based on a joint work with Marc Yor. We consider integrability properties of functionals

$$I_\infty(f) := \int_0^\infty f(B_t^{(\mu)}) dt,$$

where f is a non-negative, locally integrable function and $\{B_t^{(\mu)} : t \geq 0\}$ is a Brownian motion with drift $\mu > 0$. The results are summarized in the following table

Class	Criterion
$\mathcal{K}_1 : I_\infty(f) < \infty \text{ a.s.}$	$\Leftrightarrow \int^\infty f(x) dx < \infty$
$\mathcal{K}_2 : \forall n \in \mathbf{N}, \forall x$ $\mathbf{E}_x(I_\infty(f)^n) < \infty$	$\Leftarrow \int^\infty f(x) dx < \infty$ and $\exists m > 0 f(x) = O(x ^m)_{x \rightarrow -\infty}$
$\mathcal{K}_3 : \exists \gamma > 0, \forall x$ $\mathbf{E}_x(\exp(\gamma I_\infty(f))) < \infty$	$\Leftarrow \int^\infty f(x) dx < \infty$ and $f(x) = O(1)_{x \rightarrow -\infty}$
$\mathcal{K}_4 : \sup_x \mathbf{E}_x(I_\infty(f)) < \infty$	$\Leftrightarrow \int_{-\infty}^\infty f(x) dx < \infty$

Moreover, under some additional assumptions on f , we construct a diffusion such that $I_\infty(f)$ is identical in law to the first hitting time of a point for this diffusion. In particular, the following two functionals and their one-sided variants are discussed:

$$(a) \quad \int_0^\infty \exp(-2 B_t^{(\mu)}) dt, \quad (b) \quad \int_0^\infty (a + \exp(B_t^{(\mu)}))^{-2} dt, \quad \mu = 1/2.$$

It is seen, e.g., that the identity associated with the one-sided functional

$$\int_0^\infty \exp(-2 B_t^{(\mu)}) \mathbf{1}_{\{B_t^{(\mu)} > 0\}} dt$$

yields, via Lamperti's transformation, the Ciesielski–Taylor identity.

QI-MAN SHAO

Normal Comparison Inequalities and their Applications

Let X and Y be R^n -valued normally distributed random vectors. In this talk we shall review comparison inequalities between $P(X \in A)$ and $P(Y \in A)$ in general, and Slepian and reverse Slepian inequalities in particular. Applications to the Erdős and Révész law of the iterated logarithm for stationary Gaussian processes and the random pursuit problem for the fractional Brownian motion will be discussed.

ZHAN SHI

Law of the iterated logarithm, local time, random walk, random environment

Abstract: The talk consists of discussions upon some favourite topics: law of the iterated logarithm, local time, random walk, random environment. Joint work with A. Dembo, N. Gantert, Y. Peres.

SZABADOS TAMÁS

Stochastic integration based on strong approximation by simple random walks

Based on a construction of Brownian motion which uses a nested sequence of random walks that a.s. uniformly converges on bounded intervals, one can give a similar approximation for continuous local martingales. This strong approximation can be used for an a.s. converging approximation of the local time, of stochastic integrals and also of the Itô-Tanaka formula.

ZSOLT TALATA* AND IMRE CSISZÁR

Consistent Estimation of the Basic Neighborhood of Markov Random Fields

For Markov random fields with finite set of states, a modification of the Bayesian Information Criterion admits strongly consistent estimation of the smallest region that determines the conditional distributions. Phase transition or non-stationarity do not affect the result.

BÁLINT TÓTH

Hydrodynamic limit for two-component hyperbolic systems

I will present two recent results in the context of the title. (1) Proof of the hydrodynamic limit valid even beyond the appearance of shock waves for a two-component lattice gas system. (Joint work with Jozsef Fritz) (2) Universal hydrodynamic description of propagation of small perturbations of equilibrium states in hyperbolic systems with two components. (Joint work with Benedek Valko.)

TUSNÁDY GÁBOR

Decision forests

According to Kauffman's theory enzymes have two possible behavior in living organisms: they may be active or passive. Behaviors of enzymes at a given moment together form the actual state of the organism. The states are changing step by step following prescribed rules dictated by a system called Kauffman network. After a possible initial phase one previous state returns and then the whole process is systematically repeated. The repeated states together form an attractor of the organism. In microarray measurements the activity of clones is measured. In the present investigation we reduce the experimental results to two possible outputs: one clone may be active or passive likewise the enzymes themselves. The problem what we are discussing here is the reconstruction of the structure of enzymatic interactions of the living organism from microarray data. The task resembles recapitulating

the whole story of a film from unordered and perhaps not complete collections of its pieces. We shall use two basic ingredients in tackling the problem. In our earlier works we used an evolutionary strategy called Tierra which was proposed by Tom Ray for investigating complex systems. Here we apply the method together with the tree-structure of clones found in our earlier statistical analysis of microarray measurements.

MARC YOR

Renormalizing Brownian motion with its maximum, minimum, and local time processes

The limiting laws, as t tends to infinity, for Brownian motion perturbed by multiplicative weights, up to time t , depending on its maximum, minimum, and local time processes, are obtained and described. Connections with Pitman's theorem are detailed.