Notes on finite-range dependent random matrices

Greg W. Anderson, University of Minnesota

gwanders@umn.edu www.math.umn.edu/∼gwanders

Workshop on Random Matrices and Wireless Communications

Boulder, Colorado

July 2008

TALK OUTLINE

Part I

Discussion of joint work with O. Zeitouni:

"A CLT for regularized sample covariance matrices", Annals of Statistics (to appear)

Part II

Remarks on Khintchine-type inequalities in the dependent situation.

References (various takes on RM's with dependent entries)

A. Boutet de Monvel, A. Khorunzhy, and V. Vasilchuck, *Markov Proc. Rel. Fields* **2** (1996), pp. 607–636.

L. Pastur, Annales de l'I.H.P., section A, tome 64, n0 3(1996) 325–337

W. Hachem, P. Loubaton, J. Najim, *Markov Proc. Rel. Fields* **11** (2005), pp. 629–648.

R. R. Far, T. Oraby, W. Bryc, R. Speicher, IEEE Trans. on Information Th., Vol. 54, No. 2, Feb. 2008

P. Bickel and E. Levina, *Regularized estimation of large covariance matrices* to appear in Annals of Statistics

G. Anderson, O. Zeitouni, "A law of large numbers for finite-range dependent random matrices", CPAM 2008

References (various takes on Khintchine-type inequalities and applications)

P. Whittle, *Bounds for the moments of linear and quadratic forms in independent variables.* (Russian summary) Teor. Verojatnost. i Primenen. **5**(1960), 331-335.

Y. Q. Yin, *Limiting spectral distribution for a class of random matrices.* J. Multivariate Anal. **20**(1986), 50–68. (contains "Yin's lemma")

I. K. Matsak, A. N. Plichko, *The Khinchin inequality for k-multiple products of independent random variables. (Russian)* Mat. Zametki **44**(1988), no. 3, 378-384, 411; translation in Math. Notes **44**(1988), no. 3-4, 690694 (1989).

Z.D. Bai, J. W. Silverstein *On the Empirical Distribution of Eigenvalues of a class of Large Dimensional Random Matrices.* J. Multivariate Analysis 54(1995)175-192 (representative of many applications of Yin's lemma)

BACKGROUND FOR PART I

Classical problem
Given a stationary sequence $\{Z_j\}_{j=-\infty}^{\infty}$ of real random variables (with $\mathbf{E}Z_0 = 0$ for simplicity), estimate the covariance

$$R(j) = \mathrm{Cov}(Z_0, Z_i).$$

Basic strategy is to consider the sample covariance matrix for a matrix of independent samples from the sequence:

$$X = \begin{bmatrix} Z_1^{(1)} & \ldots & Z_p^{(1)} & (\leftarrow \text{first sample}) \\ \vdots & & \vdots & \vdots \\ Z_n^{(n)} & \ldots & Z_p^{(n)} & (\leftarrow n^{th} \text{ sample}) \end{bmatrix},$$

in the naive expectation that

$$X^*X \sim \mathbf{E}X^*X = [R(i-j)]_{i,j=1}^{p}.$$

Classically this approach was successfully considered for large n and small p.

But nowadays you might be in bad shape...

EMAIL FROM MY COUSIN AT MAYO

“ Oh yeah by the way, the “p” has increased again, this time to 7 million ... There is a new piece of DNA sequencing equipment on campus. I have just heard briefly about it and it generates 700 Gb of data per 7 patient samples and about 7 million useable pieces of information regarding the counts of DNA sequence occurences. Now we are starting to get big! ”

BACK ON TASK...

It is a big job to estimate $R(j)$ outside of large n small p regime. About this see the cited paper of Bickel and Levina for more information, references and background. We retreat from the big problem to the much smaller one of estimating the measure ν_Z to which the normalized counting measure of eigenvalues of $[R(i-j)]_{i,j=1}^p$ converges as $p \to \infty$. The latter we know by the Szego limit theorem. (More details on this later...)

It is well-known that outside of the large n small p regime, the normalized counting measure of eigenvalues of X^*X doesn't even come close to converging to ν_Z. How to cope?

REGULARIZATION

It is currently a hot topic to figure out how to overcome the problem of bad convergence by *regularization*. One rather unsophisticated and brutal method is *truncation*, i.e., putting to 0 all entries of the sample covariance matrix "far" from the diagonal.

ROUGH DESCRIPTION OF RESULT

We pursue the truncation strategy in a certain regime which allows large p and small n, managing to get convergence to ν_Z at the LLN level. Then we prove a corresponding CLT.

Now to some details...

FORMULATION OF THE MAIN RESULT

BASIC PARAMETERS

Throughout, let

$$p = \text{“sample vector length”}$$

be a positive integer. Let

$$b = b(p) = \text{“half-width of band”}$$

and

$$n = n(p) = \text{“number of samples”}$$

be positive integers depending on p.

Assumption I As $p \to \infty$, we have
$b \to \infty$, $n \to \infty$ and $b/n \to 0$, with $b \leq p$.

"INPUT" STATIONARY SEQUENCE

Let $\{Z_j\}_{j=-\infty}^{\infty}$ be a stationary sequence of real random variables variables, satisfying

Assumption II:

$$\mathbf{E}(|Z_0|^k) < \infty \text{ for all } k \geq 1 \tag{1}$$

$$\mathbf{E}Z_0 = 0 \tag{2}$$

$$\sum_{j_1} \cdots \sum_{j_r} |\mathbf{C}(Z_0, Z_{j_1}, \ldots, Z_{j_r})| < \infty \quad \text{ for all } r \geq 1. \tag{3}$$

Here, for random variables $U_1, \ldots, U_n$, we let $\mathbf{C}(U_1, \ldots, U_n)$ denote their joint (classical) cumulant. We call (3) above *joint cumulant summability*.

EXAMPLES OF STATIONARY SEQUENCES SATISFYING ASSUMPTION 2

Fix a summable function

$$h : \mathbb{Z} \to \mathbb{R}$$

and an i.i.d. sequence

$$\{W_\ell\}_{\ell=-\infty}^{\infty}$$

of mean zero real random variables with moments of all orders. Now convolve: put

$$Z_j = \sum_\ell h(j+\ell) W_\ell$$

for every j. The formula

$$\mathbf{C}(Z_{j_0}, \ldots, Z_{j_r}) \;=\; \sum_\ell h(j_0+\ell) \cdots h(j_r+\ell) \mathbf{C}(\underbrace{W_0, \ldots, W_0}_{r+1})$$

holds at least when h is finitely supported, then holds in general by a limiting argument, and easily implies joint cumulant summability.

RANDOM MATRICES

Let $\{\{Z_j^{(i)}\}_{j=-\infty}^{\infty}\}_{i=1}^{\infty}$ be an i.i.d. family of copies of $\{Z_j\}_{j=-\infty}^{\infty}$. Let $X = X^{(p)}$ be the n-by-p random matrix with entries

$$X(i,j) = X_{ij} = Z_j^{(i)}/\sqrt{n}.$$

Let $Y = Y^{(p)}$ be the p-by-p random symmetric matrix with entries

$$Y(i,j) = Y_{ij} = \begin{cases} (X^T X)_{ij} & \text{if } |i-j| \leq b, \\ 0 & \text{if } |i-j| > b, \end{cases} \qquad (Y \text{ arises by } \textit{banding})$$

and eigenvalues $\{\lambda_i^{(p)}\}_{i=1}^{p}$. Let

$$L^{(p)} = \frac{1}{p}\sum_{i=1}^{p} \delta_{\lambda_i^{(p)}}$$

be the corresponding normalized eigenvalue counting measure.

A PEEK AT THE THEOREMS

Theorem 1 (A.-Zeitouni) *Let Assumptions I and II hold. Then* $L^{(p)}$ *converges weakly in probability as* $p \to \infty$ *to* ν_Z.

We'll provide several equivalent detailed descriptions of ν_Z presently.

Theorem 2 (A.-Zeitouni) *Let Assumptions I and II hold. The process*

$$\left\{\sqrt{\frac{n}{p}}(\mathrm{trace} Y^k - \mathbf{E}\mathrm{trace} Y^k)\right\}_{k=1}^{\infty}$$

converges in distribution as $p \to \infty$ *to a zero mean Gaussian process* $\{G_k\}_{k=1}^{\infty}$ *with covariance specified by the formula*

$$\frac{1}{k\ell}\mathbf{E}G_k G_\ell = 2R_0^{(k+\ell)} + \sum_{i,j\in\mathbb{Z}} R_i^{(k-1)} Q_{ij} R_j^{(\ell-1)}.$$

We'll define Q_{ij} and $R_i^{(m)}$ presently.

THE MEASURE ν_Z

For integers j let $R(j) = \mathrm{Cov}(Z_0, Z_j)$. Since $\mathbf{C}(Z_0, Z_j) = \mathrm{Cov}(Z_0, Z_j)$, joint cumulant summability implies existence of a *spectral density*

$$f_Z : [0, 1] \to \mathbb{R}$$

associated with the sequence $\{Z_j\}$, defined to be the Fourier transform

$$f_Z(\theta) = \sum_{j \in \mathbb{Z}} e^{2\pi i j \theta} R(j).$$

By the Szegö limit theorem, the normalized eigenvalue counting measure of the matrix

$$R(|i - j|)_{i,j=1}^{p}$$

converges as $p \to \infty$ to the measure

$$\nu_Z := (\text{Lebesgue measure}) \circ f_Z^{-1}$$

on $\mathbb{R}$, i.e., the law of f_Z viewed as a random variable.

ANOTHER CHARACTERIZATION OF ν_Z

All moments of ν_Z are finite and are given by

$$\begin{aligned}\int_{\mathbb{R}} x^k \nu_Z(dx) &= \int_0^1 f_Z(\theta)^k d\theta = \underbrace{R \star R \star \cdots \star R}_{k}(0) \\ &= \sum_{\substack{i_1,\ldots,i_k \in \mathbb{Z} \\ i_1+\cdots+i_k=0}} \mathrm{Cov}(Z_0, Z_{i_1}) \cdots \mathrm{Cov}(Z_0, Z_{i_k}),\end{aligned}$$

where $\star$ denotes convolution:

$$(F \star G)(j) = \sum_{k \in \mathbb{Z}} F(j-k)G(k),$$

for any two summable functions $F, G : \mathbb{Z} \to \mathbb{R}$. Note that (4) could just as well serve as the definition of ν_Z.

THE COEFFICIENTS Q_{ij} AND $R_i^{(m)}$

For integers $m > 0$ and all integers i and j, we write

$$Q_{ij} = \sum_{\ell \in \mathbb{Z}} \mathbf{C}(Z_i, Z_0, Z_{j+\ell}, Z_\ell),$$

$$R_i^{(m)} = \underbrace{R \star \cdots \star R}_{m}(i), \quad R_i^{(0)} = \delta_{i0}.$$

Joint cumulant summability insures that Q_{ij} is well-defined and summable:

$$\sum_{i,j \in \mathbb{Z}} |Q_{ij}| < \infty.$$

The array Q_{ij} is also symmetric:

$$Q_{ij} = \sum_{\ell \in \mathbb{Z}} \mathbf{C}(Z_{i-\ell}, Z_{-\ell}, Z_j, Z_0) = Q_{ji},$$

by stationarity of $\{Z_j\}$ and symmetry of $\mathbf{C}(\cdot,\cdot,\cdot,\cdot)$ under exchange of its arguments.

BACK TO THE THEOREMS (details filled in)

Theorem 1 (Law of large numbers) *Let Assumptions I and II hold. Then* $L^{(p)}$ *as* $p \to \infty$ *converges weakly to* ν_Z*, in probability.*

In other words, Theorem 1 implies that L is a consistent estimator of ν_Z, in the sense of weak convergence.

Theorem 2 (Central limit theorem) *Let Assumptions I and II hold. The process*

$$\left\{\sqrt{\frac{n}{p}}(\mathrm{trace} Y^k - \mathbf{E}\mathrm{trace} Y^k)\right\}_{k=1}^{\infty}$$

converges in distribution as $p \to \infty$ *to a zero mean Gaussian process* $\{G_k\}_{k=1}^{\infty}$ *with covariance specified by the formula*

$$\frac{1}{k\ell}\mathbf{E}G_k G_\ell = 2R_0^{(k+\ell)} + \sum_{i,j\in\mathbb{Z}} R_i^{(k-1)} Q_{ij} R_j^{(\ell-1)}.$$

Note that the "correction" Q_{ij} vanishes if $\{Z_j\}$ is Gaussian.

REMARK

We tried to obtain a corresponding result on the “bias” in the CLT (an approximation to $\mathbf{E}\text{trace}(Y^k)$ at CLT scale) as one usually does, but this proved to be too hard. It is possible that the bias depends significantly on the rate at which b/n tends to 0. In other words, special cases might proliferate wildly. The problem to calculate the bias is wide open.

HINTS OF PROOF

The approach is via "method of moments", and heavily combinatorial.

We work with an exact representation for the classical joint cumulants of the random variables traceY^k.

To sort out the terms according to the order of their contribution the following combinatorial result is needed.

KEY COMBINATORIAL RESULT

Part(k): set partitions of $\{1, \ldots, k\}$.

Given $\Pi, \Sigma \in$ Part(k), let $\Pi \vee \Sigma \in$ Part(k) be the least upper bound of Π and Σ.

We call $\Pi \in$ Part(k) a *perfect matching* if every part of Π has cardinality 2.

Let Part$_{\geq 2}$(k) be the subfamily of Part(k) consisting of partitions Π such that every part has cardinality at least 2.

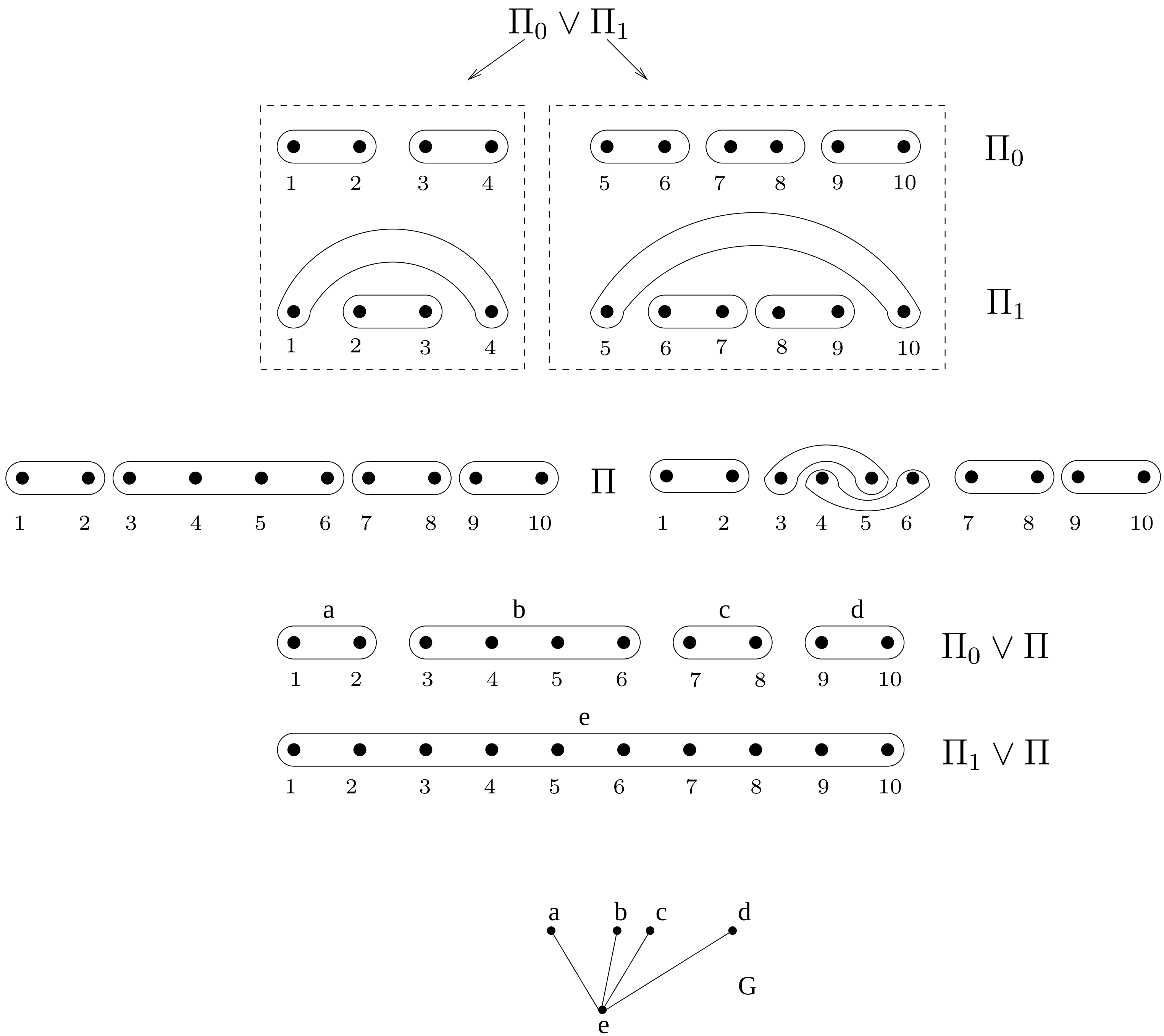
Proposition 1 *Let k be a positive integer. Let $\Pi_0, \Pi_1, \Pi \in \mathrm{Part}_{\geq 2}(2k)$ be given. Assume that Π_0 and Π_1 are perfect matchings. Assume that $\#\Pi_0 \vee \Pi_1 \vee \Pi = 1$. Then we have*

$$\#\Pi_0 \vee \Pi + \#\Pi_1 \vee \Pi \leq 1 + \#\Pi \leq k + 1,$$

and furthermore,

$$r > 1 \Rightarrow \#\Pi_0 \vee \Pi + \#\Pi_1 \vee \Pi \leq k + 1 - \lfloor r/2 \rfloor,$$

where $r = \#\Pi_0 \vee \Pi_1$.


$\Pi_0 \vee \Pi_1$
Π_0
Π_1
Π
a b c d
$\Pi_0 \vee \Pi$
e
$\Pi_1 \vee \Pi$
G

REMARKS ON THE PROPOSITION

Formally, the proposition bears a significant resemblance to the Riemann-Hurwitz formula for the genus of a covering of the sphere, or rather, to be more precise, the inequality deduced by using the fact that genus is nonnegative. In fact, if Π is also a perfect matching, then the proposition *is* an instance of Riemann-Hurwitz.

The proposition is proved by adapting a graph-theoretical argument in Anderson-Zeitouni PTRF 2006.

The proposition expresses a point of view which may be capable of efficiently reconciling the "bracelet-and-pendant-tree" picture and the "second order freeness" picture in Mingo-Speicher J. Functional Analysis 2006 (and many follow on papers). "Dictionaries" remain to be written.

Part II

Remarks on Khintchine-type inequalities in the dependent situation

GENERAL MOTIVATION

Stieltjes-based methods for studying RM in the independent entry situation rely heavily on Khintchine-type inequalities.

We are going to spend a few minutes reviewing the Khintchine inequality and some descendants, and then explain a strategy for pushing to the dependent situation which I have found useful, as part of an ongoing Stieltjes-based attack on LLN's and CLT's for random hermitian matrices with "finite-range-dependent-above-the-diagonal" entries.

THE KHINTCHINE INEQUALITY

Let $\{\epsilon_i\}_{i=1}^{\infty}$ be i.i.d. with

$$\Pr(\epsilon_1 = \pm 1) = 1/2.$$

Let $p \geq 1$ be a constant.

Theorem 3 (Khintchine 1923) *There exist constants*

$$0 < A_p \leq B_p$$

such that for all sequences $a_1, \ldots, a_n \in \mathbb{C}$

$$A_p \left(\sum_{i=1}^{n} |a_i|^2 \right)^{1/2} \leq \left\| \sum_{i=1}^{n} a_i \epsilon_i \right\|_p \leq B_p \left(\sum_{i=1}^{n} |a_i|^2 \right)^{1/2}.$$

Best constants are known (famous result of Haagerup with major preliminary contributions by Szarek).

GENERALIZATION

With $p \geq 2$ an integer, let $X_1, \ldots, X_n$ be real random variables with $\|X_i\|_p < \infty$. We say these are p-*uncorrelated* if for each index $i_0 = 1, \ldots, n$ and monomial M of degree $< p$ in the X_i for $i \neq i_0$ we have $\mathbf{E}(X_{i_0} M) = (\mathbf{E} X_{i_0})(\mathbf{E} M)$.

Theorem 4 *Let $p \geq 2$ be an even integer. Let $X_1, \ldots, X_n$ be real random variables with finite L^p-norms and vanishing means. Assume further that $X_1, \ldots, X_n$ are p-uncorrelated. Then we have*

$$\left\| \sum_{i=1}^{n} X_i \right\|_p \;\leq\; p \left(\sum_{i=1}^{n} \|X_i\|_p^2 \right)^{1/2} .$$

I don't know first occurrence of exactly this result in the literature. But there is considerable overlap with linear part of Whittle and with the Marcinkiewicz-Zygmund inequality. And anybody familiar with Yin's lemma would know how to generate a proof.

HIGHER KHINTCHINE-TYPE INEQUALITIES

Theorem 5 *Let $X_1, \ldots, X_n$ be independent real random variables with $\|X_i\|_p \leq 1$ and $\mathbf{E}X_i = 0$ for all i. For $\emptyset \neq I \subset \{1, \ldots, n\}$ put $X_I = \prod_{i \in I} X_i$. Let $\{A_I\}_{\emptyset \neq I \subset \{1,\ldots,n\}}$ be a collection of real numbers. Then for all integers $p \geq 2$ and $\ell \geq 1$ we have*

$$\left\| \sum_{\substack{I \subset \{1,\ldots,n\} \\ \#I = \ell}} A_I X_I \right\|_p \leq \frac{p^\ell \ell^\ell}{\ell!} \left(\sum_{\substack{I \subset \{1,\ldots,n\} \\ \#I = \ell}} A_I^2 \right)^{1/2} .$$

Again I don't know exact reference for this. But it overlaps with the main result of Matsak-Plichko, and with the quadratic part of the main result of Whittle. And for $\ell \leq 2$ one can generate a proof using "Yin's lemma". Also there is a cheap trick...

PROOF We may assume $\ell > 1$, because otherwise nothing to prove. Let N be very large and divisible by ℓ. For $I \subset \{1,\ldots,N\}$ but $I \not\subset \{1,\ldots,n\}$, put $A_I = 0$. For $i \in \{1,\ldots,N\}$ but $i \notin \{1,\ldots,n\}$, put $X_i = 0$. We then have an algebraic identity

$$\ell \binom{N-\ell}{N/\ell-1} \sum_{\substack{I\subset\{1,\ldots,N\}\\ \#I=\ell}} A_I X_I \;=\; \sum_{\substack{J\subset\{1,\ldots,N\}\\ \#J=N/\ell}} \sum_{j\in J} \sum_{\substack{K\subset\{1,\ldots,N\}\\ \#K=\ell-1\\ J\cap K=\emptyset}} A_{\{j\}\cup K} X_j X_K$$

which by the previous theorem, induction on ℓ, and the Minkowski inequality implies what we want except for the ugly constant

$$\frac{\binom{N}{N/\ell}}{\ell\binom{N-\ell}{N/\ell-1}} \frac{(\ell-1)^{\ell-1}p^\ell}{(\ell-1)!}.$$

Letting $N \to \infty$ we get the result as stated. QED

KHINTCHINE-TYPE INEQUALITIES FOR DEPENDENT RV's

The "Fibonacci" random variable

Let χ be a random variable such that

$$\Pr(\chi = \phi) = \frac{1}{1+\phi^2}, \quad \Pr(\chi = -1/\phi) = \frac{\phi^2}{1+\phi^2} \quad \left(\phi = \frac{1+\sqrt{5}}{2}\right).$$

Then

$$\begin{aligned}
\mathbf{E}\chi^2 &= 1, \\
\mathbf{E}\chi &= 0, \\
\chi^2 &= 1+\chi, \\
\|\chi\|_\infty &= \phi < 2, \\
\mathbf{E}\chi^m &\geq 1 \quad \text{for integers } m \geq 2.
\end{aligned}$$

Notion of ρ-dependence Let (T,ρ) be a finite metric space (finite nonempty set plus metric). Let $\{X_t\}_{t\in T}$ be a family of real random variables. We say that $\{X_t\}_{t\in T}$ is ρ-*dependent* if for all nonempty sets $T_1, T_2 \subset T$ such that $\min_{t_1\in T_1}\min_{t_2\in T_2}\rho(t_1,t_2) > 1$, the subfamilies $\{X_t\}_{t\in T_1}$ and $\{X_t\}_{t\in T_2}$ are independent. In short, "distant subfamilies are independent".

Common setup for next two propositions

- Let (T,ρ) be a finite metric space.
- Let $\{X_t\}_{t\in T}$ be a family of ρ-dependent integrable mean zero real random variables.
- Let $p \geq 2$ be an even integer.
- $\{\chi_t\}_{t\in T}$ be an i.i.d. family of copies of χ.

Proposition 2 *Assume that*

$$\max_{t\in T} \|X_t\|_p \le 1.$$

Then for every family $\{A_t\}_{t\in T}$ *of real constants we have*

$$\left\| \sum_{t\in T} A_t X_t \right\|_p \le \left\| \sum_{t\in T} |A_t| \left(\sum_{\substack{t'\in T \\ \rho(t,t')<p}} \chi_{t'} \right) \right\|_p .$$

Proposition 3 *Assume that*

$$\max_{t \in T} \|X_t\|_{2p} \leq 1.$$

Put

$$Y_{t,t'} = X_t X_{t'} - \mathbf{E}(X_t X_{t'}),$$

$$\phi_{t,t'} = \begin{cases} \chi_t \chi_{t'} & \text{if } t \neq t', \\ 2\chi_t & \text{if } t = t' \end{cases} \qquad \text{for } t, t' \in T.$$

Then for any family $\{A_{t,t'}\}_{t,t' \in T}$ *of real constants we have*

$$\left\| \sum_{t,t' \in T} A_{t,t'} Y_{t,t'} \right\|_p \leq \left\| \sum_{t_1,t_2 \in T} |A_{t_1,t_2}| \left(\sum_{\substack{t_3 \in T \\ \rho(t_1,t_3) < 2p}} \sum_{\substack{t_4 \in T \\ \rho(t_2,t_4) < 2p}} \phi_{t_3,t_4} \right) \right\|_p .$$

LAST COMMENTS

In both propositions we have control of the right sides by the higher Khintchine inequalities (Theorem 5 above).

Proofs in both cases are relatively simple. One "opens the brackets", taking note of the important fact that all terms coming up on the right are nonnegative. Then for every nonzero term on the left one finds some nonzero term on the right to dominate it.

Higher (cubic, quartic...) versions can be obtained.