Three Lectures on the Mobius Function Randomness and Dynamics

by

Peter Sarnak*

Lecture 1. The Mobius Flow

The Mobius function $\mu(n), n = 1, 2, 3, \ldots$ is defined by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n \text{ is not square-free} \\ (-1)^t & \text{if } n \text{ is a product of } t \text{ distinct primes.} \end{cases} \tag{1}$$

Its behavior is central in the theory of prime numbers. For example, the "orthogonality" of μ and the function 1, meaning that

$$\sum_{n=1}^{N} \mu(n) = o(N), \text{ as } N \to \infty, \tag{2}$$

is elementarily equivalent to the prime number theorem. The more quantitative statement; for $\epsilon > 0$

$$\sum_{n=1}^{N} \mu(n) = O_\epsilon(N^{1/2+\epsilon}) \tag{3}$$

is equivalent to the Riemann Hypothesis. The exponent $1/2$ in (3) is often taken as an indication that the $\mu(n)$'s are random in some sense and much has been written about it. However, as with any deterministic sequence at close enough inspection it is not random. For example, if $N = e^y$ then as a function of y

$$G(y) = N^{-1/2} \sum_{n \leq N} \mu(n) \tag{4}$$

*Department of Mathematics, Princeton University & Institute for Advanced Study

is no doubt an unbounded B^2-almost periodic function (see [Ng 1]). This is a reflection of the existence and location of the zeros of the Riemann Zeta Function.

A well-known but vague principle concerning the randomness of $\mu(n)$ is that summing it against any reasonable function $\xi(n)$ leads to significant cancellations. This is called the "Mobius randomness law" in [I-K] and in [G-T] it is expressed by saying μ doesn't correlate with any function of low "complexity." In general, this means that

$$\sum_{n\leq N} \mu(n)\,\xi(n) \;=\; o\left(\sum_{n\leq N} |\xi(n)|\right). \tag{5}$$

and in practice one often needs an explicit rate in the implied o notation, typically $O((\log N)^{-A})$ for some $A > 0$. We limit ourselves to ξ's which are bounded and not sparsely supported and ask only for orthogonality, that is

$$\sum_{n\leq N} \mu(n)\,\xi(n) \;=\; o(N) \text{ as } N \to \infty. \tag{6}$$

One can view ξ being orthogonal to μ as a strong form of ξ not approximating μ. In [I-L-S] it is shown that the above randomness principle is in some settings reliable up to square-root cancellations but that similar heuristics concerning randomness for sums over primes are not.

One of the aims in these lectures is make some aspects of the Mobius randomness principle precise by identifying the notions of reasonable or of low complexity. One might try doing so using the modern notion of computational complexity. That is ξ is of low complexity if it is in the class P (polynomial) meaning that each value of $\xi(n)$ can be computed in $O((\log n)^B)$ steps for some $B = B(\xi)$. This brings up the issue of whether μ itself is in P, which is apparently not known [A-M]. There is a wide spread belief in cryptographic and computer science circles that factoring is hard. There is no theoretical or scientific evidence for such a belief and I for one believe that the opposite is true. In particular that μ is P. In this direction it would be interesting to give an explicit ξ in P which is not orthogonal to μ or better still such that

$$\tfrac{1}{N} \sum_{n\leq N} \mu(n)\,\xi(n) \to \alpha > 0. \tag{7}$$

Such a ξ would serve as a probabilistic substitute for μ for various purposes. It is not hard to give ξ's in P for which (5) fails or that the limit in (6) is approached slowly. For example if $\xi(n) = -1$ if n is a prime and is 0 otherwise, then $\xi \in P$ [A-K-S] and

$$\frac{1}{N}\sum_{n\le N}\xi(n)\mu(n) \;=\; \frac{1}{N}\sum_{n\le N}|\xi(n)|\,1 \;\sim\; \frac{1}{\log N}\,. \tag{8}$$

Given these remarks, our view is that computational complexity is not the correct notion for examining μ-randomness.

The study of (6) when ξ is multiplicative, that is $\xi(mn)=\xi(m)\xi(n)$ if $(m,n)=1$, has received a lot of attention because of its close connection to the theory of L-functions. In this case (6) reduces to the behavior of ξ at the primes. For example, the general theorem of Wirsing [Wi] implies that if ξ is multiplicative and $-1\le\xi(n)\le 1$ then ξ is orthogonal to μ iff the non-negative series over the primes

$$\sum_p \frac{1+\xi(p)}{p}\,,\ \text{diverges.}$$

However, if we replace the multiplicative function $\xi(n)$ by $\xi(n+a)$ or more generally by $\xi(n+a_1)\xi(n+a_2)\cdots\xi(n+a_t)$ with $0<a_1<a_2\cdots<a_t$, then saying anything about orthogonality to μ is problematic. This is exemplified by considering the local correlations of μ with itself. The following conjecture of Chowla [Ch] (see also [Ng2]) is very plausible and its only drawback is that it is apparently very difficult having resisted any progress.

CONJECTURE 1: (Chowla):

Let $0\le a_1<a_2<\cdots<a_t$ and $k_1,k_2\cdots k_t$ in $\{1,2\}$ not all even, then as $N\to\infty$

$$\sum_{n\le N}\mu^{k_1}(n+a_1)\,\mu^{k_2}(n+a_2)\cdots\mu^{k_t}(n+a_t) \;=\; o(N)\,.$$

We will use this conjecture as a working hypothesis which suggests what is likely to be true in connection with the randomness principle. Our approach is to realize $\xi(n)$ as the return times or sampling sequence of a dynamical system. In this way the complexity of $\xi(n)$ is measured by that of the dynamics. This point of view has been developed by Furstenberg and we use the notions and notations from his paper [F1] and Glasner's book [GL]. A flow F is a pair (X,T_X) where X is a compact metric space and $T:X\to X$ is a continuous map. The sampling sequences associated with a flow F are the sequences $\xi(n)=f(T^nx)$ for some $x\in X$ and $f\in C(X)$. In this case we say that ξ is realized in the flow F. A given sequence $\xi(n)$ can be realized in many flows and we seek to realize it (and the sequences $\xi(n+a_1)\xi(n+a_2)\cdots\xi(n+a_t)$ derived from it) in the simplest possible flow. If $\xi(n)$ takes values in a finite set $\Lambda\subset\mathbb{R}$ then $\xi(n)$ may be realized in the full shift flow $F=(\Omega,T)$ where $\Omega=\Lambda^{\mathbb{N}}$ with its product topology and $T:\Omega\to\Omega$ is the shift $Tx(n)=x(n+1)$. Let $\xi=(\xi(1),\xi(2),\cdots)\in\Omega,\pi_1(x)=x_1$, then $\xi(n)$ is realized in F_Λ with $x=\xi$ and $f=\pi_1$. In fact ξ is realized in the potentially simpler flow $F_\xi=(X_\xi,T)$ where X_ξ is the closure in Ω of the orbit of ξ under T. We call this sub-flow of F_Λ the shift flow determined by ξ.

Note that if $\xi(n)$ is realized in F and $\eta(n)$ in G then $\xi(n+a)\eta(n+b)$ is realized in the product flow $F \times G = (X_F \times X_G, T_F \times T_G)$.

The basic measure of complexity of a flow F is its topological entropy $h(F)$ (see [A-K-M] for definitions and properties). It is a measure of the exponential growth rate of the number of distinct orbits in F. The flow is called deterministic if it has zero entropy. A sequence $\xi(n)$ is called deterministic if it is realized in a deterministic flow. The name comes from the similar considerations in the probabilistic setting of dynamics which we call a process. We introduce it since it will play a role later in our discussion. A process is a triple $F_\nu = (X, T, \nu)$ with X as before, $T : X \to X$ a Borel measurable transformation and ν a T invariant Borel (probability) measure i.e. $\nu(T^{-1}A) = \nu(A)$ for all measurable A. In this process setting the basic measure of complexity is the Kolmogorov-Sinai entropy, $h(F_\nu)$. A process F_ν is of zero entropy iff it is deterministic in the sense that for ν-almost all sampling sequences, $\xi(1)$ is determined by $\xi(2), \xi(3), \xi(4) \cdots$.

With this dynamical realization of sequences we can state our first result about the randomness of $\mu(n)$.

PROPOSITION 2: *$\eta(n) = \mu^2(n)$ and a-fortiori $\mu(n)$, is not deterministic.*

This follows from the positivity of the entropy of the square-free flow S discussed below which in turn is proved in Lecture 3. As far as orthogonality in (6) we make the following definition.

DEFINITION 3: *μ is orthogonal to a flow F if it is orthogonal to every sequence realized in F.*

In connection with flows of positive entropy and sequences that can be realized by them, the following result of Weiss [We] is worth keeping in mind. Let F be a subflow of positive entropy of the full shift F_Λ with $\Lambda = \{0, 1\}$, then there is a subset A of $\mathbb{N}$ of positive density such that the restriction of X_F to A consists of all sequences in Λ^A. Thus we can't expect μ to be orthogonal to a typical positive entropy flow (though Bourgain [Bo1] shows that given a sequence $\epsilon(n)$ satisfying certain conditions such as $\mu(n)$, one can construct a positive entropy flow orthogonal to ϵ). Our fundamental conjecture concerning the randomness of Mobius is the following:

CONJECTURE 4: μ is orthogonal to any deterministic flow F and in particular to every deterministic sequence.

Lecture 2 is devoted to progress towards Conjecture 4 whose formulation is such that this can be measured by establishing it for various classes of flows. In applications specific F's enter and each case for which it is proved is hard earned. If F is the trivial flow (i.e. a single point) then the Conjecture is no more than (2) which is the prime number theorem. If F is a finite flow (i.e. X_F is finite) then it is essentially the prime number theorem for arithmetic progressions. If F is the rotation of the circle ($X_F = \mathbb{R}/\mathbb{Z}$ and $T_\alpha x = x + \alpha$) then Conjecture 4 was proved by Davenport [Da] using the sieve and bilinear techniques of Vinogradov [Vi]. All further progress

on the Conjecture is based on these techniques in one of their modern renditions ([Va], [I-K]). Vinogradov's methods are well suited to our dynamical setting requiring effective equidistribution of the orbits of T_F and its powers as well as the same for joinings of F with itself. Conjecture 4 is true for equicontinuous or Kronecker flows (a Kronecker flow $K = (G, T_\alpha)$ is translation on a compact topological group G given by $T_\alpha x = x \cdot \alpha$ where $\alpha \in G$) indeed these cases reduce to Davenport's Theorem. The next simplest flows are distal flows. $F = (X_F, T_F)$ is distal if given distinct points x, y in X_F, $\inf_{n \geq 1} d(T^n x, T^n y) > 0$. These flows are deterministic ([Pa]), they contain the equi-continuous flows and Furstenberg [F2] shows how to build the general minimal such flow by repeated isometric extensions, starting from the trivial flow. An interesting class of distal flows are the homogeneous nilflows given by translations on compact nilmanifolds. As part of their program on linear equations in primes, Green and Tao ([GT]) establish Conjecture 4 for these nilflows. In Lecture 2 we establish Conjecture 4 for some nonhomogeneous distal skew products. These are the typical building blocks in Furstenberg's classification and proving Conjecture 4 for distal flows seems to be a ripe problem. For deterministic flows which are statistically more complex, for example mixing, Conjecture 4 is apparently much more difficult. A critical case and perhaps the most interesting one is that of a homogeneous flow coming from a unipotent translation on a compact (or finite volume) quotient of a semi-simple Lie group. These flows are mixing ([Ma], [Mo]) and Ratner's theorem ([Ra], [Wi]) yields the requisite rigidity and equidistribution of orbits, however her proof is ineffective and applying Vinogrdov's method is problematic. In the simplest setting of $G = SL_2$ [Ub-Sa] establish various results towards Conjecture 4 which we describe in Lecture 2.

The distribution of orbits in a general deterministic flow can be quite irregular and one may be skeptical about Conjecture 4. However the following shows that it is at least as realistic as Conjecture 1.

Theorem 5: *Conjecture 1 implies Conjecture 4.*

The proof of this is entirely combinatorial and is given in Lecture 3. We persist in maintaining Conjecture 4 as the central one even though it is much weaker than Conjecture 1. The point is that Conjecture 4 refers only to correlations of μ with deterministic sequences and avoids the difficulties associated with self correlations.

In Definition 3 it is important that we demand orthogonality for all sequences realized in F. If we relax the condition to most x in the measure theoretic context of a process $F_\nu = (X_F, T, \nu_F)$ then such orthogonality is valid universally, i.e. for every F_ν. Indeed, define the linear operator L_N on continuous functions on X_F by

$$L_N f(x) = \frac{1}{N} \sum_{n=1}^{N} \mu(n) f(T^n x) \tag{9}$$

L_N extends to a bounded linear operator defined on $L^2(X, \nu_F)$. Using the spectral theorem for the unitary operator U_T on $L^2(X, \nu_F)$ given by $U_T f(x) = f(Tx)$ and Davenport's exponential sum bound (see Lectures 2) we have that for any $h > 0$

$$\| L_N \|_2 \underset{2}{\ll} (\log N)^{-h}, \tag{10}$$

(the implied constant depends only on h!) In particular, for $g \in L^2(X, \nu_F)$, $L_N g \to 0$ in $L^2(X, \nu_F)$. Using the ergodic theorems in [Bo2] one can deduce further that $L_N g \to 0$ for ν_F-almost all x.

We return to Conjecture 4. Having realized the low complexity sequences in terms of deterministic flows it seems natural enough to realize $\mu(n)$ dynamically. This leads us to some interesting flows and related processes and it also allows for a partial interpretation of Conjecture 4 in terms of joinings. The sequence $\mu(n)$ defines a point which we denote by w in $\Omega^{(3)} = \{-1, 0, 1\}^{\mathbb{N}}$.

DEFINITION 6: *The Mobius flow M is the subflow of the full shift on $\Omega^{(3)}$ determined by w, that is $M = (X_M, T)$ with X_M the closure of the T-orbit of w in $\Omega^{(3)}$.*

M is easily defined but saying something nontrivial about it is another matter. It is closely related to two other flows; its square-free factor S and the full extension N of S to $\Omega^{(3)}$.

DEFINITION 7: *The square-free flow S is the subflow of the full shift on $\Omega^{(2)} = \{0, 1\}^{\mathbb{N}}$ determined by the point $\eta = (\mu^2(1), \mu^2(2), \cdots)$, that is $S = (X_S, T)$ with X_S the closure of the T-orbit of η in $\Omega^{(2)}$.*

The surjective map $\phi : X_M \to X_S$ taking $x = (x_1, x_2, \ldots)$ to $(x_1^2, x_2^2, \ldots)$ and taking w to η realizes S as a factor of M (that is ϕ is continuous and $T_S \phi = \phi T_M$). One can study S in some detail and its salient properties as well as a central process associated with it, are described next.

Call a subset A of $\mathbb{N}$ admissible if its reduction mod p^2 doesn't cover all of the residue classes mod p^2 for every prime p. This condition is translation invariant, that is if A is admissible so is $A + k$ for any $k \geq 0$. For $y \in \{0, 1\}^{\mathbb{N}}$ denote by support(y) the subset of $\mathbb{N}$ at which $y \neq 0$. The set of all $y \in \Omega^{(2)}$ whose support is admissible, is a closed T-invariant subset which we denote by $\mathcal{A}$. It is not hard to see that $X_S \subset \mathcal{A}$. The following theorem describes the dynamical properties of S.

Theorem 8.

(i) $X_S = \mathcal{A}$.

(ii) $T_s : X_S \to X_S$ is surjective, it is topologically ergodic and $h(S) = \frac{6}{\pi^2} \log 2$.

(iii) S is proximal (i.e. for any $x, y \in X_S$, $\inf_{n\geq 1} d(T_x^n, T_y^n) = 0$) and $\{(0,0,0\cdots)\}$ is the unique T-minimal subset of X_S.

(iv) S has no nontrivial Kronecker factors but it has a nontrivial joining with $K = (G, T)$ where G is the compact abelian group $\prod_p(\mathbb{Z}/p^2\mathbb{Z})$ and $Tx = x + (1, 1, \cdots)$. In particular, S is not weak-mixing.

The key point is part (i) identifying X_S with $\mathcal{A}$. Once we have this it is clear that if Q is the set of square-free numbers then the restriction of elements in X_S to Q contains all zero one sequences on Q. Thus Q is an explicit set satisfying Weiss' independent condition in the discussion after Definition 3. It is well-known that Q has density $6/\pi^2$. The entropy of S being $(6/\pi^2)\log 2$ says that Q is a largest such independent set.

Part (i) of Theorem 8 is proven by showing that the orbit of η is equidistributed with respect to a measure ν_S on $\Omega^{(2)}$ whose support is $\mathcal{A}$. To define ν_S we first define $\gamma(A)$ for A is a finite subset of $\mathbb{N}$ by

$$\gamma(A) = \prod_p \left(1 - \frac{t(A, p^2)}{p^2}\right) \tag{11}$$

where $t(A, p^2)$ is the number of distinct residue classes gotten by reducing $A \mod p^2$. The product in (11) is over all the primes. Clearly $\gamma(A) = \gamma(A + k)$ for $k \geq 0$ and $\gamma(A) > 0$ iff A is admissible. For disjoint finite subsets A and B of $\mathbb{N}$ set

$$\gamma(A; B) = \sum_{A\subset D\subset A\cup B} (-1)^{|D|-|A|}\gamma(D). \tag{12}$$

Then $\gamma(A; B) = \gamma(A + k, B + k)$ for $k \geq 0$ and it is less clear but true, that $\gamma(A; B) \geq 0$, with equality iff A is not admissible (see Lecture 3 for a proof of this important last point). For A and B as above define the cylinder set $C_{A;B}$ in $\Omega^{(2)}$ by

$$C_{A;B} = \{x : x_a = 1 \text{ if } a \in A,\ x_b = 0 \text{ if } b \in B\}. \tag{13}$$

Finally define the Borel probability measure ν_S on $\Omega^{(2)}$ by setting

$$\nu_S(C_{A;B}) := \gamma(A; B). \tag{14}$$

Clearly ν_S is T invariant and by the remarks above its support is $\mathcal{A}$. Let S_ν be the process $(\mathcal{A}, T, \nu_s)$ which we call the square-free process. We will prove the following in Lecture 3 using among other things elementary sieve methods (see for example [Ts]).

Theorem 9:

(i) *The T-orbit of η in $\Omega^{(2)}$ is ν_S equidistributed, that is for $f \in C(\Omega^{(2)})$*

$$\frac{1}{N}\sum_{n=1}^{N} f(T^n\eta) \to \int_{\mathcal{A}} f(y)d\nu_S(y)\,,$$

or using other terminology, η is ν_S-generic.

(ii) *The process S_ν is ergodic and deterministic (i.e. is of zero entropy).*

(iii) *The Kronecker process $K_\nu = (X_K, T, \nu)$ where K_K is the group $\prod_p(\mathbb{Z}/p^2\mathbb{Z}$, $T(x) = x+(1,1,\ldots)$ and ν is Haar measure on X_K, is a factor of S_ν.*

The quotient map realizing (iii) above has an explicit description. By ergodicity, for ν_S almost all $x \in \mathcal{A}$, the residue class mod p^2 omitted by support (x) is unique for each prime p. This gives a measurable map $t : x \to (x_4, x_9, \ldots)$ which is a morphism from S_ν onto K_ν. The determinism in part (ii) can be described using t. Given a ν_S-generic point $x \in \mathcal{A}$, x_1 is given in terms of Tx as follows: If for some prime p $\ t(Tx)_{p^2} = 1$ then $x_1 = 0$, otherwise $x_1 = 1$.

The second flow connected with the Mobius flow is the full extension of S to $\Omega^{(3)} = \{-1, 0, 1\}^{\mathbb{N}}$. Let $X_N \subset \Omega^{(3)}$ consist of all x's whose support is admissible. X_N is closed, T-invariant and defines a subflow N of the full shift $\Omega^{(3)}$. The map $\phi : X_N \to X_S$ which squares each coordinate gives a morphism of N onto S, so that S is a factor of N. N is topologically ergodic and has entropy equal to $(6/\pi^2)\log 3$. N inherits various properties from the factor S such as being nonmixing, proximal and having $\{0\}$ as its only minimal invariant subflow .

The statistical properties of interest are connected with a specific process on X_N. Let ν_N be the measure on $\Omega^{(3)}$ defined on cylinder sets W by

$$\nu_N(W) = 2^{-|\text{ support } W|}\,\nu_S(\phi(W)). \tag{15}$$

ν_N is shift invariant, its support is X_N and it defines a process $N_\nu = (X_N, T, \nu_N)$. Having constructed N_ν as a fully random extension of S_ν the following properties of N_ν follow from Theorem 9.

(i) N_ν is ergodic and $h(N_\nu) = \frac{6}{\pi^2} \log 3$.

(ii) S_ν is the Pinsker factor of N_ν that is N_ν is a completely positive (entropy) extension of S_ν and S_ν is its largest deterministic factor.

In order to examine the question of orthogonality in Conjecture 4 for the Mobius flow we need to study various joinings. Let Y_ν be a deterministic process and consider the possible joinings of N_ν and Y_ν. They need not be disjoint since N_ν has S_ν as a factor. However, since S_ν is the Pinsker factor of N_ν any common factor of N_ν and Y_ν must factor through S_ν. Let $\lambda \in J(N_\nu, Y_\nu)$ be a joining of these processes. Using that Y_ν and all its self-products are deterministic and that N_ν is a completely positive extension of S_ν together with the disjointness theorem [G-T-W], we conclude that; If $f \in C(X_N)$ and its conditional expectation $E_{\nu_N}(f|\mathcal{S}) = 0$ (where $\mathcal{S}$ is the σ-algebra on X_N corresponding to the factor S_ν) and $g \in C(X_Y)$ then $\lambda(f \otimes g) = 0$. Hence as in [F1] if $n_0 \in N$ is ν_N-generic and $y_0 \in X_Y$ is ν_Y-generic, then as $L \to \infty$.

$$\frac{1}{L} \sum_{m=1}^{L} f(T^m n_0)\, g(T^m y_0) \to 0. \tag{16}$$

In particular if Y is a uniquely ergodic deterministic flow with unique invariant measure ν_Y, then every point $y \in X_Y$ is ν_y-generic and we can replace y_0 by a general point y in (16).

Finally, we turn to the Mobius flow. M is a subflow of N and both have S as a factor. As a corollary to Theorem 8 we have

Theorem 10:

(i) *The entropy of M is positive and satisfies*

$$\frac{6}{\pi^2} \log 2 \leq h(M) \leq \frac{6}{\pi^2} \log 3\,.$$

(ii) *M is proximal and $\{0\}$ is its only minimal subflow.*

(iii) *M has a nontrivial joining with the Kronecker flow K and in particular M is not weak mixing.*

The expectation is that $X_M = X_N$ and hence that $M = N$ (see below) but this appears to be well beyond the reach of current techniques. Any result showing that M is a random extension of S such as $h(M) > h(S)$, would be very interesting. Apparently the only results in this direction are those towards Conjecture 4 discussed in Lecture 2. One can formulate a relative version of

Conjecture 4 (i.e. orthogonality of μ to any relatively deterministic extension of S) which implies that $h(M) > h(S)$.

The oldest conjecture that implies $X_M = X_N$ is the Hardy-Littlewood k-tuple conjecture ([H-L]) (a similar remark about a related conjecture of Chowla for the Liouville function $\lambda(n)$ following from Schinzel's hypothesis H, is made in [C-F-M-R-S]). We know even less at the statistical level as far as the Mobius flow goes. Again the expectation is clear enough.

CONJECTURE 11 *The point $w = (\mu(1), \mu(2), \ldots)$ in X_N is ν_N-generic.*

In Lecture 3 we show that Conjecture 11 and Conjecture 1 are equivalent. Conjecture 11 implies that $M = N$ and places the processes $N_\nu = M_\nu$ as central to the Mobius flow together with its dense orbit corresponding to w. According to (16) and Conjecture 11 we have that for any deterministic uniquely ergodic flow Y and any $f \in C(X_M)$ with $E(f|\mathcal{S}) = 0$, $g \in C(X_Y)$ and $y \in X_Y$

$$\frac{1}{L}\sum_{m=1}^{L} f(T^m w))\, g(T^m y) \to 0, \text{ as } L \to \infty. \tag{17}$$

Specializing further to $f(x) = x_{a_1}^{k_1} x_{a_2}^{k_2} \cdots x_{a_t}^{k_t}$ with $k_j \in \{1, 2\}$ not all even, so that $E_{\nu_M}(f|\mathcal{S}) = 0$, yields

$$\frac{1}{L}\sum_{m=1}^{L} \mu^{k_1}(a_1 + m)\, \mu^{k_2}(a_2 + m) \cdots \mu^{k_t}(a_t + m)\, g(T_y^m) \to 0. \tag{18}$$

In particular it follows that μ is orthogonal to any deterministic uniquely ergodic flow Y. While this doesn't recover the full implication in Theorem 5 it gives an explanation in terms of disjointness of a substantial part of the basic Conjecture 4: that the Mobius process M_ν as an extension of S_ν is relatively disjoint from any deterministic process.

References for Lecture 1:

[A-M] L. Adleman and K. McCurley, Lecture notes in Comp. Sci. **877**, 29-322, Springer (1994).

[A-K-M] R. Adler, A. Konheim and M. McAndrew *T.A.M.S.*, **114**, 309-319, (1965).

[A-K-S] M. Agrawal, N. Kayal and N. Sexana, *Annals of Math.*, **160**, 781-793, (2004).

[Bo1] J. Bourgain; private communication (2009).

[Bo2] J. Bourgain "An approach to pointwise ergodic theorems," *L.N.M.*, 1317, (1988).

[C-F-M-R-S] J. Cassaigue, S. Ferenczi, C. Mauduit, J. Rivat and A. Sarkozy, *Acta Arith.* LXXXVII.4, 367-390, (1999).

[Ch] S. Chowla, "The Riemann Hypothesis and Hilbert's tenth problem," Gordon & Breach, NY, (1965).

[Da] H. Davenport, *Quat. J. Math.*, **8**, 313-320, (1937).

[F1] H. Furstenberg, *Math. Syst. Theory*, Vol 1, No. 1, 1-49, (1961).

[F2] H. Furstenberg, *American. J. Math.*, **85**, 477-515, (1963).

[GL] E. Glasner, "Ergodic Theory via Joinings," *AMS Math Surveys*, Vol. 101, (2002).

[G-T-W] E. Glasner, J. Thouvenot and B. Weiss, *Erg. Th. and Dynamical Systems*, **20**, 1355-1370, (2000).

[G-T] B. Green and T. Tao, "The Mobius Function is strongly orthogonal to nilsequences," ARx1 08071736.

[H-L] G. Hardy and J. Littlewood, *Acta Math.*, **44**, 1-70, (1922).

[I-K] H. Iwaniec and E. Kowalski, "Analytic Number Theory," *AMS Coll. Ser.*, **53**, (2004).

[I-L-S] H. Iwaniec, W. Luo, P. Sarnak, *Pub. I.H.E.S.*, **91**, 55-131, (2001).

[Ma] B. Marcus, *Invent. Math.*, **46**, 201-209, (1978).

[Mo] S. Mozes, *Invent. Math.*, **107**, 235-241, (1992).

[Ng 1] N. Ng, *Proc. London Math. Soc.*, **89**, 361-389, (2004).

[Ng 2] N. Ng, "The Mobius function in short intervals," *CRM Proc. & Lett. Notes*, **46**, 247-267, (2008).

[Pa] W. Parry, "Entropy and generators in ergodic theory," Benj., New York, (1969).

[Ra] M. Ratner, *Annals of Math.*, **134**, 545-607, (1991).

[Ts] K. M. Tsang, *Mathematika*, **32**, 265-275, (1985).

[U-S] A. Ubis and P. Sarnak, "Unipotent orbits and primes," (in preparation).

[Va] R. Vaughan, *C.r. Acad. Sci.*, Paris, **A 285**, 981-983, (1977).

[Vi] I. M. Vinogralov, *Recueiv Math.*, (1937).

[We] B. Weiss, "Single orbit dynamics," CBMS Conf. Series, **95**, *AMS*, (2000).

[Wi] E. Wirsing, *Acta Math. Acad. Sci. Hung.*, **18**, 411-467, (1967).

[Wi] D. Witte, "Ratner's theorems on unipotent flows," University of Chicago Press, (2005).