STATISTICAL PROPERTIES OF EIGENVALUES OF THE HECKE OPERATORS

Peter Sarnak

0. Introduction.

Two basic questions concerning the Ramanujan τ-function concern the size and variation of these numbers :

(i) Ramanujan conjecture: $|\tau(p)| \leq 2p^{11/2}$ for all primes p.

(ii) "Sato-Tate" conjecture: $a_p = \dfrac{\tau(p)}{p^{11/2}}$ is equidistributed with respect to

$$d\mu(x) \;=\; \begin{cases} \dfrac{1}{2\pi}\,\sqrt{4-x^2}\;dx & \text{if } |x| \leqslant 2 \\[2mm] 0 & \text{otherwise} \end{cases}$$

as $p \to \infty$. We refer to the last as the semicircle distribution.

Concerning the above the following is known: (i) has been proved by Deligne [1]. However its generalization to a general GL(2) cusp form, as well as to more general groups is far from being solved. (ii) This conjecture is motivated by related questions for L-functions of elliptic curves [8]. It is conjectured to be true for $\tau(p)$ as well as for "typical" cusp forms in GL(2). It certainly does not hold for all cusp forms and we will consider this again later. Our aim here is to outline results which prove averaged versions of (i) and (ii) in general.

I have benefited immeasurably from discussions with R. Phillips and I. Piatetski-Shapiro and some of the results quoted here are from joint work with them.

1. Classical Hecke Operators.

We begin by considering the simplest example of Hecke

operators. Let $\Gamma = SL(2, \mathbb{Z})$ and $h = \{ z \mid \text{Im } z > 0 \}$. Let H be the Hilbert space $L^2(\Gamma/h)$, that is of all Γ invariant functions on h which are square summable over a fundamental domain F for Γ with respect to $d\mu(z) = \dfrac{dxdy}{y^2}$. The operators in question are then defined by

$$
\begin{cases}
T_n f(z) = \dfrac{1}{\sqrt{n}} \sum_{\substack{ad=n \\ b \bmod d}} f(\dfrac{az+b}{d}) \\[2ex]
T_\infty f(z) = -\Delta f(z) = -y^2(\dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2})f(z)
\end{cases}
\tag{1.1}
$$

for $n = 1, 2, \ldots$.

It is well known that $\{T_n\}$ forms a commutative family of self-adjoint operators. Furthermore H decomposes into Hecke invariant subspaces

$$
H = \{1\} \oplus E \oplus \text{Cusp}
$$

where $\{1\}$ spans the constant functions, E is spanned by Eisenstein series [3] and Cusp is orthogonal to these and consists of cuspidal functions. On Cusp we have a simultaneous orthonormal basis of $\{T_n\}$ which we denote by $u_j(z)$:

$$
\begin{cases}
T_p u_j = \rho_j(p) u_j \\[2ex]
T_\infty u_j = (\dfrac{1}{4} + r_j^2) u_j = \lambda_j u_j
\end{cases}
\tag{1.2}
$$

where $\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \ldots$. Thus we use the λ's to order the u_j's.

For these cusp forms u_j, very little is known about $\rho_j(p)$ or r_j. Very interesting computations of $\rho_1(p)$ for $p < 1000$ and r_j for small j appear in Stark [10] and Hejhal [3]. For these, the Ramanujan conjecture takes the form

$$
|\rho_j(p)| \leq 2
\tag{1.3}
$$

for all j and primes p.

We note that since the Ramanujan conjecture holds for the Eisenstein series $E(z, \dfrac{1}{2} + it)$, as one checks easily by a calculation, we can restate the Ramanujan conjecture purely in terms of the

spectrum of T_p. Thus the following is equivalent to (1.3). For p a prime,

$$\left|\begin{array}{l} |\langle T_p f, f\rangle| \le 2\langle f, f\rangle \quad \text{for all } f \in L^2(\Gamma/h) \text{ for which} \\ \qquad\qquad \langle f, 1\rangle = 0. \end{array}\right. \tag{1.3'}$$

Put another way $\sigma(T_p|_{\{1\}^\perp}) \subset [-2,2]$. Here $\sigma(T)$ is the spectrum of T. On the other hand $T_p 1 = (p^{1/2} + p^{-1/2})1$ and indeed

$$n(p): = \|T_p\| = p^{1/2} + p^{-1/2} > 2. \tag{1.4}$$

It is known that

$$|\rho_j(p)| \le 2 (p^{1/5} + p^{-1/5}). \tag{1.5}$$

(This was communicated to the author in a letter from S.J. Patterson 1981).

Definition 1.6. Let X be a topological space. We say that a sequence x_j in X is μ-equidistributed where μ is a Radon measure on X, if for all $f \in C_c(X)$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{j \le N} f(x_j) \to \int_X f(x) \, d\mu(x) . \tag{1.6}$$

The Sato-Tate conjecture for the numbers $\rho_j(p)$, states that for fixed j, $\rho_j(p)$ is μ-equidistributed, where μ is the semicircle distribution.

Our approach here is to study these questions concerning $\rho_j(p)$ in both variables j and p. Thus we consider seriously the operator $T_p|_{\text{Cusp}}$ i.e. the variation in j for fixed p. Our first result is a density result concerning the number of exceptions T_p may have to the Ramanujan conjecture. We recall Weyl's law, see Selberg [9]

$$N(K) = \# \{r_j \le K\} \sim \frac{1}{12} K^2. \tag{1.7}$$

For $\alpha \ge 2$ (and p fixed) we set

$$N(\alpha, K) = \#\{j \mid r_j \le K, \ |\rho_j(p)| \ge \alpha \}.$$

Theorem 1.1.
$$N(\alpha,K) \ll K^{2 - \frac{\log \alpha/2}{\log p}}.$$

In particular almost all $\rho_j(p)$ (in the sense of density in j) lie in $[-2,2]$.

Concerning the variation of the $\rho_j(p)$ in j and p, let

$$x_j = (\rho_j(2),\ \rho_j(3),\ \rho_j(5),\ldots)$$

so that $x_j \in X = \underset{p}{\Pi}\ [-n(p),\ n(p)]$.

Theorem 1.2. $\{x_j\}$, $j = 1,2,\ldots$ *is μ equidistributed in X where* $\mu = \underset{p}{\Pi}\ \mu_p$ *and*

$$d_{\mu_p}(x) = \begin{cases} \dfrac{(1+p)\ \sqrt{4-x^2}}{2\pi(n(p)^2 - x^2)} & \text{if } |x| < 2 \\ \\ 0 & \text{otherwise}. \end{cases}$$

The following Corollary was first proved by Phillips and Sarnak [7] by completely different methods. In that paper approximate eigenfunctions for T_p were constructed directly.

Corollary 1.3. *Let α_m, β_m, $m = 1,2,\ldots,k$ be numbers satisfying $-2 < \alpha_m < \beta_m < 2$ and let $p_1,p_2,\ldots,p_k$ be k primes. Then*

$$\lim_{K\to\infty} \frac{1}{K^2}\ \#\{r_j < K \,|\, \rho_j(p_m) \in [\alpha_m,\ \beta_m],\ m = 1,\ldots,k\} > 0.$$

It follows that any given finite sequence of numbers, satisfying the Ramanujan bound may be approximated by the eigenvalues of a cusp form.

In the above we study the behavior of $\rho_j(p)$ as a vector in p as $j \to \infty$. If, as expected, the Sato-Tate holds for each j, we might hope that the interchange of the two limits would agree. It is clear that

$$\lim_{p\to\infty} \mu_p = \mu \quad \text{the semicircle distribution!}$$

What this shows is that in this way of averaging the numbers $\rho_j(p)$, we do have equidistribution with respect to the semicircle. There are obvious advantages in averaging over j, since if for example we consider cusp forms for $\Gamma_o(N)$, $N > 1$, then there is a subset of the j´s (the number of which whose $r_j \leqslant K$, is of order K) for which the Sato–Tate conjecture is false. These are cusp forms coming from the Maass–Hecke construction [4]. Of course these disappear in our averaging and indeed we still find that the generic cusp form has the semicircle behaviour. These Maass–Hecke cusp forms have their eigenvalues equidistributed with respect to μ_p above, with p = 1! The measures μ_p therefore interpolate between this distribution at p = 1, and the semicircle at p = ∞ .

A final comment concerning the semicircle. As p → ∞ the operators T_p are presumably becoming random, at least that is what we are showing. For it is known that the eigenvalues of a random Hermitian matrix, whose size tends to ∞ , become distributed according to the semicircle distribution. This is due to Wigner (see [6]) and is known as the Wigner semicircle law.

We will discuss the general case in Section 4. We first turn to a general phenomenon which is at the heart of the above considerations.

2. A Weyl Law.

In this section we describe an extension of the classical Weyl theorem on eigenvalues of the Laplacian to the case where we have a family of operators commuting with the Laplacian. Let M be a compact Riemannian manifold and $\tilde{M} \overset{\Delta}{=} S$ its universal cover. Let G be the isometry group of S and so $\Gamma = \Pi_1(M)$ is a discrete subgroup of G. Δ will denote the Laplacian on M or S. Now suppose we are given a family of operators $T_1, T_2, \ldots$ on $L^2(M)$ for which the family $\Delta, T_1, T_2, \ldots$ is commutative. We take the T_j to be bounded, with say $\|T_k\| = n_k$. We may then simultaneously diagonalize the family:

$$\begin{cases} T_k u_j = \rho_j(k)\, u_j \\ T_\infty u_j = -\Delta\, u_j = \lambda_j u_j \end{cases} \qquad (2.1)$$

where $\{u_j\}_{j=1,2,\ldots}$ is an orthonomal basis for $L^2(M)$, and are

ordered by increasing λ_j. The asymptotics of λ_j is well known, this being Weyl's law

$$N(\lambda) = \#\{\lambda_j \leqslant \lambda\} \sim C\lambda^{n/2} \qquad (2.2)$$

where C is an appropriate non-zero constant and n = dim M. Let $B_k = \{z \in \mathbf{C} \mid |z| \leqslant n_k\}$ and

$$X = \prod_k B_k. \qquad (2.3)$$

For j = 1,2,... we obtain a point x_j in X where

$$x_j = (\rho_j(1), \ \rho_j(2), \ \rho_j(3) \ ...).$$

The question is how do these x_j's distribute themselves in X as j → ∞ ? To obtain an answer we assume further the T_k's are "Hecke like" operators. So we assume T_k to be selfadjoint (normal would suffice) and is of the form

$$T_k f(x) = \sum_{\ell=0}^{n_k} f(S_\ell^{(k)} x) \qquad (2.4)$$

where $S_\ell^{(k)} \in G$. The important assumption is that $T_k: L^2(\Gamma/S) \rightarrow L^2(\Gamma/S)$, which can be arranged with appropriate $S_\ell^{(k)}$ if the commensurator of Γ in G is non-trivial [11].

For ν_1 , ν_2 ,..., $\nu_r \in \mathbf{N}$ let

$$M(\nu_1, \ \nu_2, ..., \ \nu_r) = \text{the number of words of the type}$$
$$\begin{cases} \omega_1 \omega_2 \ ... \omega_r \equiv I \pmod \Gamma \text{ where } \omega_k \text{ is a} & (2.5) \\ \text{word in } S_1^{(k)}, S_2^{(k)} \ ... \ S_{n(k)}^{(k)} \text{ of length } \nu_k. \end{cases}$$

In this case, since we are assuming that the T_k's are self-adjoint, our space X in (2.3) is a product of intervals.

Theorem 2.1 *Let T_k be as above, then the sequence $\{x_j\}_{j=1,2,...} \in X$ is μ equidistributed, where μ is the measure given by the moments*

$$\int_X t_1^{\nu_1} \ ... \ t_k^{\nu_k} \ d\mu(t_1,...) = M(\nu_1,...,\nu_k) \ .$$

Notice that since X is compact, one sees easily that μ exists and is unique. We now examine some simple instances of the above theorem.

Example 2.2. Suppose that the original manifold M admits a non-trivial isometry $S : M \to M$ of order k (k may be infinite). Let $T : L^2 \to L^2$ be the unitary operator given by

$$Tf(x) = f(Sx).$$

T commutes with Δ and let u_j be as above with

$$T u_j = \omega_j u_j \qquad j = 1, 2, \ldots \quad .$$

Clearly $|\omega_j| = 1$. The theorem then asserts that ω_j is μ-equidistributed on the circle where

(i) μ puts mass $1/k$ at the k-th roots of 1 if $k < \infty$.

(ii) μ is $d\theta/2\pi$ on the circle, if $k = \infty$.

Example 2.3. $M = S^{\cdot} = R/\mathbf{Z}$, $\Delta = \dfrac{d^2}{dx^2}$, $u_j(x) = e^{2\pi i j x}$. Let $\alpha_1, \ldots, \alpha_k \in \mathbf{R}$ and $T_k(x) = x + \alpha_k$. In this case $\rho_j(k) = e^{2\pi i j \alpha_k}$. The theorem thus asserts that the sequence $j(\alpha_1, \alpha_2, \ldots, \alpha_k)$, $j = 1, 2, \ldots$ is μ-equidistributed in the k-torus. Clearly $M(\nu_1, \ldots, \nu_k) = 0$ if $1, \alpha_1, \alpha_2, \ldots, \alpha_k$ are linearly independent over Q, so that in this case the sequence is equidistributed with respect to Lebesgue measure. This is the well known result of Weyl [12].

The main application of the theorem is however to the Hecke operators in symmetric spaces. In the case of $\Gamma = SL(2, \mathbf{Z})$ as in Section 1, there are added complications in the proof of the above type of theorem due to the noncompactness. We will outline the proof in that case in the next section. The proof of Theorem 2.1 in the general case combines the ideas outlined in the next section, with the standard derivation of Weyl's law via differential equation methods - e.g. small time behavior of the fundamental solution to the wave equation on $\widetilde{M}$.

In the $\Gamma = SL(2, \mathbf{Z})$ case of Section 1, if we ignore the difficulties coming from the Eisenstein series (which in this case are not difficult to overcome) we can compute the number $M(\nu)$ for T_p

quite easily from the well known identity

$$T_p^n \, T_p = T_{p^{n+1}} + T_{p^{n-1}} \ .$$

We find

$$N(\nu) = \int_{-\infty}^{\infty} t^\nu \, d\mu_p(t) = \begin{cases} 0 \ , & \text{if } \nu \text{ is odd} \\ \sum_{j=0}^{n} \left(\binom{2n}{n-j} - \binom{2n}{n-j-1} \right) p^{-j}, & \text{if } \nu = 2n. \end{cases}$$

The inverse moment problem is easily solved giving the μ_p's in Theorem 1.2. The fact that μ is a product of the μ_p's follows from the multiplicative property of the Hecke operators.

3. Outline of Proofs.

We now outline proofs of the results in Section 1, details will appear elsewhere. The basic ingredient is the Selberg trace formula but it is not the full formula that is needed. Indeed such a formula cannot be used to prove Theorem 2.1. Basically what we need is the "singularity at 0" in the trace formula.

Consider the case of $\Gamma = SL(2, \mathbf{Z})$. Let $k(z, \zeta)$ be a point pair invariant [3], which we assume to have very small support. That is $k(z, \zeta) = 0$, if $d(z, \zeta) > \varepsilon$, where $d(z, \zeta)$ is the non-Euclidian distance from z to ζ. Let

$$K(z, \zeta) = \sum_{\gamma \in \Gamma} k(z, \gamma \zeta) \ . \tag{3.1}$$

We have the spectral expansion [3]

$$K(z, \zeta) = \sum_j h(r_j) u_j(z) \overline{u_j(\zeta)} + \frac{1}{4\pi} \int_{-\infty}^{\infty} h(t) E(z, \tfrac{1}{2} + it) \overline{E(\zeta, \tfrac{1}{2} + it)} \, dt. \tag{3.2}$$

For what follows we ignore the contribution from the Eisenstein series since in this case as was mentioned before they are known explicitly, and may be dealt with easily. It follows that

$$T_p^\nu \, K(z, \zeta) = \sum_j h(r_j)(\rho_j(p))^\nu u_j(z) \overline{u_j(\zeta)} + \ldots \tag{3.3}$$

and hence

$$[T_p^\nu K(z,\zeta)]_{z=\zeta} = \sum_j h(r_j)(\rho_j(p))^\nu |u_j(\zeta)|^2 + \ldots \quad . \qquad (3.4)$$

However one can calculate $[T_p^\nu K(z,\zeta)]_{z=\zeta}$ asymptotically as $\varepsilon \to 0$:

$$[T_p^\nu K(z,\zeta)]_{z=\zeta} = \frac{1}{p^{\nu/2}} \sum_{\gamma \in \Gamma} k(S_{i_1} S_{i_2} \ldots S_{i_\nu} z, \gamma\zeta)|_{z=\zeta}$$

so that unless ζ is the fixed point of some $\gamma^{-1} S_{i_1} S_{i_2} \ldots S_{i_\nu}$, the above is zero for ε small enough.

On integrating with respect to ζ one finds the main contribution comes from exactly those $S_{i_1} S_{i_2} \ldots S_{i_\nu} \equiv I \pmod{\Gamma}$. This, combined with (3.4) leads naturally to the asymptotics

$$\frac{1}{N(K)} \sum_{|r_j| \leqslant K} (\rho_j(p))^\nu \sim M(\nu) \quad . \qquad (3.5)$$

Theorems 2.1 and 1.2 follow from this type of argument. If one is more careful in the analysis in the case $\Gamma = SL(2,\mathbf{Z})$, and keeps track of all contributions above, one finds: (i) that the contribution from the continuous spectrum is controlled by the constant term of the Eisenstein series which is essentially the zeta function. (ii) the number of terms $\gamma S_{i_1} S_{i_2} \ldots S_{i_\nu}$ with fixed points in F is easily majorized by elementary bounds for class numbers of binary quadratic forms. This leads to the inequality:

$$K > p^k \Rightarrow \sum_{|r_j| \leqslant K} |\rho_j(p)|^{2k} \leqslant 2^k K^2 + p^{2k} 2^k \qquad (3.6)$$

Theorem 1.1 is an immediate consequence.

4. General Case.

The results in this section are joint with I. Piatetski-Shapiro. The first thing to observe is that the measures μ_p are none other than the spherical Plancherel measures for $SL_2(Q_p)$, see for example MacDonald [5]. He uses the variable θ where $x = 2 \cos \theta$. One may also see that this is so by carrying out the above proof using the adelic trace formula for $GL_2(\mathbb{Q})/GL_2(\mathbf{A}_\mathbb{Q})$ [2].

The case of a compact quotient such as that coming from a quaternion algebra and its generalizations, is particularly simple and an analogue of Theorem 1.2 may be proved in complete generality, i.e. for a reductive algebraic group defined over a number field. In this case the existence of a limiting distribution follows from Theorem 2.1 but the point is that one can avoid solving the inverse moment problem, since these limiting distributions are spherical Plancherel measures, which have been computed in complete generality - see MacDonald [5]. In the general noncompact case such as $G = SL(n,\mathbb{R})$, $\Gamma = SL(n,\mathbb{Z})$ there are technical problems coming from the continuous spectrum. We expect the same answer for the limiting distribution, but so far have not been able to verify it in general.

For $GL(n,\mathbb{Z})$ the eigenvalues of the p-th Hecke operators on u_j (cusp forms) may be parametrized by $\alpha_j^{(1)}(p)$, ..., $\alpha_j^{(n)}(p)$ where $\alpha_j^{(1)}...\alpha_j^{(n)} = 1$. The corresponding limiting distribution for these is the spherical Plancherel measure for $SL(n,\mathbb{Q}_p)$, and lives on the n-1 torus. As in Section 1, one takes the limit $p \to \infty$ of these measure and this turns out to be the measure

$$d\mu(\Theta_1,...,\Theta_{n-1}) = C_n \prod_{k<j} |e^{i\Theta_k} - e^{i\Theta_j}|^2 \, d\Theta_1 ... d\Theta_{n-1} \quad (4.1)$$

where

$$k,j = 1,2,...,n \text{ and } \Theta_1 + \Theta_2 ...+ \Theta_n = 0.$$

This gives a natural generalization of the semicircle or Sato-Tate distribution. Indeed the above results prove this conjecture in the average over the cusp forms (in the sense of Section 1). There are other theoretical ways of arriving at the measure in (4.1), we note in particular that it is the measure obtained by projecting Haar measure on $SU(n)$ to its maximal torus. If $n = 2$ then the measure (4.1) is $C_2 \sin^2\Theta \, d\Theta$ which is of course the semicircle distribution for the variable $\rho = 2 \cos \Theta$.

References.

[1].   Deligne, P., La conjecture de Weil I. *Publ. Math.* IHES, **43** (1974) 273-307.

[2].   Gelbart, S., <u>Automorphic forms on Adele group</u>, Anal of Math. Studies, **83**, 1975.

[3].   Hejhal, D., <u>The Selberg trace formula for PSL(2,IR)</u>, Vol.2, S.L.N. 1001, 1983.

[4].   Maass, H., Über eine neue Art von Nichtanalytischen Automorphismen ..., *Math. Ann.* 121, 1949 pp. 141-183.

[5].   MacDonald, I.G., <u>Spherical functions on groups of P-adic type</u>, TATA Inst. Series, 1971.

[6].   Mehta, M.L., <u>Random matrices</u>, Academic Press, 1967.

[7].   Phillips, R. and Sarnak, P., Preprint.

[8].   Serre, J.P., <u>Abelian ℓ-adic representations</u>, Benjamin, 1968.

[9].   Selberg, A., Gottingen lectures, 1954.

[10].   Stark, H., Fourier coefficients of Maass wave forms, in <u>Modular forms</u> ed. Rankin, Ellis Horwood, 1985.

[11].   Venkov, A.B., Spectral theory of automorphic forms, *Proc. Steklov. Inst.* 1982, No. 4 (English translation).

[12].   Weyl, H., Über die Gleichverteilung von Zahlen Mod. eins, *Math. Ann.* **77** , 1914, 313-352.

P.Sarnak,
Stanford University,
Stanford, CA, 94305 U.S.A.