Nonlinearity 1 (1988) 399–407. Printed in the UK

Semiclassical formula for the number variance of the Riemann zeros

M V Berry
H H Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK

Received 9 December 1987, in final form 20 March 1988
Accepted by J D Gibbon

Abstract. By pretending that the imaginary parts E_m of the Riemann zeros are eigenvalues of a quantum Hamiltonian whose corresponding classical trajectories are chaotic and without time-reversal symmetry, it is possible to obtain by asymptotic arguments a formula for the mean square difference $V(L; x)$ between the actual and average number of zeros near the xth zero in an interval where the expected number is L. This predicts that when $L \ll L_{\max} = \ln(E/2\pi)/2\pi \ln 2$ (where $x = (E/2\pi)(\ln(E/2\pi) - 1) + \frac{7}{8}$), V is the variance of the Gaussian unitary ensemble (GUE) of random matrices, while when $L \gg L_{\max}$, V will have quasirandom oscillations about the mean value $\pi^{-2}(\ln \ln(E/2\pi) + 1.4009)$. Comparisons with $V(L; x)$ computed by Odlyzko from 10^5 zeros E_m near $x = 10^{12}$ confirm all details of the semiclassical predictions to within the limits of graphical precision.

1. Introduction

It was realised long ago [1] that the truth of the Riemann hypothesis would be established if it could be shown that the imaginary parts E_m of the non-trivial zeros of $\zeta(z)$ are eigenvalues of a self-adjoint operator. Montgomery [2] suggested that the statistics of the E_m are those of the eigenvalues of an infinite complex Hermitian matrix drawn randomly from the Gaussian unitary ensemble (GUE) [3]. In a recent numerical study, Odlyzko [4] showed that while short-range statistics (such as the distribution of the spacings $E_{m+1} - E_m$ between neighbouring zeros) accurately conform to GUE predictions, long-range statistics (such as the correlations between distant spacings) do not, and are better described in terms of primes.

I have argued elsewhere [5, 6] that exactly this behaviour would be expected if the E_m were eigenvalues not of a random matrix but of the Hamiltonian operator obtained by quantising some still-unknown dynamical system without time-reversal symmetry, whose phase-space trajectories are chaotic. The theory is based on asymptotics of the semiclassical limit, in which Planck's constant $\hbar \rightarrow 0$. Its central result [7] is that statistics of eigenvalues separated by less than $O(\hbar)$ are universal (that is independent of the details of the Hamiltonian) and given (in the absence of time-reversal symmetry) by the GUE, while the statistics of eigenvalues with larger separations are non-universal and depend on the details of the closed trajectories of the corresponding classical system. The connection with $\zeta(z)$ comes from an analogy between the Von Mangoldt formula [1] for $\ln(\zeta(z))$ and an expression for the semiclassical eigenvalue density fluctuations in terms of closed orbits.

My purpose here is to illustrate how the semiclassical theory can give a uniformly accurate description of both the long-range (non-universal) and short-range (universal) statistics of the E_m, by studying the number variance of the zeros. This is the mean square difference between the actual number of E_m and the expected number, in an interval where the expected number is L. Although the semiclassical analogy has not led to an identification of the elusive Riemann operator (if indeed there is one), this illustration does suggest that certain tantalising hints [5] about the underlying dynamical system deserve to be taken seriously.

2. Number variance

The mean number of zeros with height less than E is [1]

$$\mathcal{N}(E) = (E/2\pi)\{\ln(E/2\pi) - 1\} + \tfrac{7}{8} \tag{1}$$

so that the numbers

$$x_m \equiv \mathcal{N}(E_m) \tag{2}$$

form a sequence with mean spacing unity. Such a sequence can be regarded as a singular density

$$\mathrm{d}(x) = \sum_{m=1}^{\infty} \delta(x - x_m) \tag{3}$$

concentrated at the points x_m on the x axis. We will employ the notion of asymptotic averaging, that is averaging over a range Δx satisfying $1 \ll \Delta x \ll x$, and denote the operation by $\langle\ \rangle$. Obviously $\langle d \rangle = 1$. The fluctuating part of the level density, defined as

$$\tilde{\mathrm{d}}(x) \equiv \mathrm{d}(x) - 1 \tag{4}$$

provides the entry point for the semiclassical theory to be described in §3.

In the range of $x - L/2$ to $x + L/2$ the number of zeros is

$$n(L; x) = \int_{x-L/2}^{x+L/2} \mathrm{d}x_1\, \mathrm{d}(x_1). \tag{5}$$

Obviously $\langle n(L;x) \rangle = L$. The number variance $V(L;x)$ is thus

$$V(L;x) = \langle [n(L;x) - L]^2 \rangle = \left\langle \int_{x-L/2}^{x+L/2} \mathrm{d}x_1 \int_{x-L/2}^{x+L/2} \mathrm{d}x_2\, \tilde{\mathrm{d}}(x_1)\, \tilde{\mathrm{d}}(x_2) \right\rangle. \tag{6}$$

This is conveniently expressed in terms of the form factor (Fourier transform of the pair correlation of the density fluctuations)

$$K(\tau; x) \equiv \int_{-\infty}^{\infty} \mathrm{d}\xi\, \exp(2\pi\mathrm{i}\xi\tau) \langle \tilde{\mathrm{d}}(x - \xi/2)\, \tilde{\mathrm{d}}(x + \xi/2) \rangle \tag{7}$$

because

$$\langle \tilde{d}(x_1)\,\tilde{d}(x_2)\rangle = \int_{-\infty}^{\infty} d\tau K(\tau;(x_1+x_2)/2)\exp[2\pi i\tau(x_1-x_2)]. \tag{8}$$

Substituting into (6) and setting $(x_1+x_2)/2$ equal to x for $x \gg 1$ and $L \ll x$ gives

$$V(L;x) = \frac{2}{\pi^2}\int_0^{\infty} d\tau \frac{K(\tau;x)}{\tau^2}\sin^2(\pi L\tau). \tag{9}$$

3. Long and short orbits

Now pretend that E_m are eigenvalues of a quantum Hamiltonian with a classical limit that is chaotic in the sense that all closed orbits are isolated and unstable. From any such set of E_m we can construct the scaled energies x_m from (2) using the appropriate counting function $\mathcal{N}(E)$. Thence we obtain the level density fluctuations $\tilde{d}(x)$ from (3) and (4). Gutzwiller [8, 9] and Balian and Bloch [10] have shown that in the semiclassical limit $\tilde{d}(x)$ can be expressed as a sum over all closed orbits at that energy E which corresponds to x according to (2). The sum is over all primitive orbits (labelled p) and their repetitions (labelled r where $1 \leqslant r \leqslant \infty$). Each orbit gives an oscillatory contribution to $\tilde{d}(x)$, with a phase $rS_p/\hbar$ where S_p is the classical action of the primitive orbit, and an amplitude A_{rp} that depends on the instability exponents of the orbit [8].

An important role is played by the periods of the orbits, given by

$$T_{rp} = r\,\partial S_p/\partial E. \tag{10}$$

These enter the form factor $K(\tau;x)$, which can be obtained [5] from (7) as

$$K(\tau;x) \simeq \frac{2\pi}{\hbar\rho}\left\langle \sum_{r_1p_1}\sum_{r_2p_2} A_{r_1p_1}A_{r_2p_2}\cos\{(r_1S_{p_1}-r_2S_{p_2})/\hbar\}\,\delta\{T-\tfrac{1}{2}(T_{r_1p_1}+T_{r_2p_2})\}\right\rangle \tag{11}$$

In this formula, ρ is the mean level density $d\mathcal{N}(E)/dE$, and T is a time variable related to τ by the scaling

$$T = \tau\hbar\rho. \tag{12}$$

For understanding the double sum (11) it is important to note that $\hbar\rho$ increases as $\hbar$ decreases for systems capable of displaying chaos (for example, in a classical billiard with N freedoms, $\hbar\rho \sim \hbar^{-(N-1)}$). This means that the period $T_{\min}$ of the shortest closed orbit corresponds to $\tau \ll 1$, while $\tau = 1$ corresponds to an extremely long orbit.

Choose an intermediate value τ^* satisfying

$$T_{\min}/2\pi\hbar\rho \ll \tau^* \ll 1. \tag{13}$$

For $\tau < \tau^*$, asymptotic averaging will remove the non-diagonal terms $r_1 \neq r_2$, $p_1 \neq p_2$ because of incoherence in the trigonometric factors (S_p depends on x), leaving

$$K(\tau;x) \simeq \frac{2\pi}{\hbar\rho}\sum_p \sum_{r=1}^{\infty} A_{rp}^2\delta(T-T_{rp}) \qquad (\tau < \tau^*). \tag{14}$$

Obviously the positions and strengths of the δ spikes in the sum depend on the details of the classical dynamics of the particular system being studied: because of this, $K(\tau;x)$ is described as non-universal when $\tau<\tau^*$.

On the other hand, for $\tau>\tau^*$ only very long orbits are involved; these proliferate exponentially with increasing T for chaotic systems and so contribute as though distributed continuously. Their contribution can be calculated [5] using a remarkable sum rule of Hannay and Ozorio de Almeida [11] relating the orbit amplitudes A_{rp} to their increasing number, and some heuristic semiclassical arguments. The result is that $K(\tau;x)$ becomes universal, that is independent of the detailed dynamics (provided these are chaotic and without time-reversed symmetry). Moreover, the expression obtained is precisely the form factor of the GUE, namely [12]

$$K(\tau;x)\simeq K_{\mathrm{GUE}}(\tau)=\tau\theta(1-\tau)+\theta(\tau-1)\qquad(\tau>\tau^*)\tag{15}$$

where θ denotes the unit step.

Substituting the non-universal and universal formulae (14) and (15) for $K(\tau;x)$ into (9) gives the number variance as

$$V(L;x)=8\sum_p\sum_{r=1}^{T_{rp}<2\pi\hbar\rho\tau^*}\frac{A_{rp}^2}{T_{rp}^2}\sin^2(LT_{rp}/2\hbar\rho)+\frac{2}{\pi^2}\int_{\tau^*}^{\infty}\mathrm{d}\tau\frac{K_{\mathrm{GUE}}(\tau)}{\tau^2}\sin^2(\pi L\tau).\tag{16}$$

This applies to any classically chaotic system, and arguments similar to those given elsewhere [5] for a related statistic (the spectral rigidity) show that $V(L;x)$ grows logarithmically according to the universal GUE formula until $L\sim L_{\max}\equiv\hbar\rho/T_{\min}$ and then oscillates non-universally whilst remaining bounded.

To apply (16) to the Riemann zeros we should of course know the underlying classical system. Without this knowledge we can only proceed by analogy, identifying T_{rp} and A_{rp} from the Von Mangoldt formula as explained in [5], ignoring difficulties [6] with the analogy. The results associate primitive closed orbits with primes p:

$$T_{rp}=r\ln p\qquad A_{rp}=-\ln p\exp(-r\ln p/2)/2\pi.\tag{17}$$

For this system $\hbar$ does not appear explicitly but can be regarded as concealed in E (as with quantum billiards); the semiclassical limit is $E\to\infty$. We also require the mean density of zeros $\rho(E)$, which from (1) is

$$\rho(E)=\frac{1}{2\pi}\ln(E/2\pi).\tag{18}$$

Now we can substitute into (16) and evaluate the integral, to get the number variance of the zeros:

$$\begin{aligned}V(L;x)=&\frac{1}{\pi^2}\{\ln(2\pi L)-\mathrm{Ci}(2\pi L)-2\pi L\mathrm{Si}(2\pi L)+\pi^2L-\cos(2\pi L)+1+\gamma\}\\&+\frac{1}{\pi^2}\left\{2\sum_p\sum_{r=1}^{p^r<(E/2\pi)^{\tau^*}}\frac{\sin^2(\pi Lr\ln p/\ln(E/2\pi))}{r^2p^r}+\mathrm{Ci}(2\pi L\tau^*)-\ln(2\pi L\tau^*)-\gamma\right\}.\end{aligned}\tag{19}$$

Here Si and Ci are the sine and cosine integrals [13], γ is Euler's constant

0.577 215 . . . and x is related to E by (2). This formula is our main result. As $x \to \infty$, the dependence on τ^* disappears provided τ^* satisfies (13), i.e. $\tau^* \gg \ln 2/\ln(E/2\pi)$.

The terms in the first set of braces give the GUE number variance, with limiting behaviour

$$V_{\mathrm{GUE}}(L) \simeq \begin{cases} L & \text{if } L \ll 1 \\ \dfrac{1}{\pi^2}[\ln(2\pi L) + 1 + \gamma] & \text{if } L \gg 1. \end{cases} \tag{20}$$

The second set of braces involves the sum over prime powers. Because of the upper limit, the largest value of the argument of the $\sin^2$ function is $\pi L \tau^*$. If $L \ll 1/\tau^*$ the sum is negligible and so are the remaining terms in these braces. Thus the first set of braces dominates, and $V \simeq V_{\mathrm{GUE}}$, if $L \ll 1/\tau^*$; this is the universal regime.

If $L \gg 1/\tau^*$, asymptotics of Si and Ci give

$$V(L;x) \simeq \frac{1}{\pi^2}\left(2 \sum_p^{p^r<(E/2\pi)^{\tau^*}} \sum_{r=1} \frac{\sin^2(\pi L r \ln p/\ln(E/2\pi))}{r^2 p^r} - \ln \tau^* + 1\right) \qquad \text{if } L \gg 1/\tau^*. \tag{21}$$

This describes asymptotic oscillations with amplitudes $(2\pi r^2 p^r)^{-1}$ and wavelengths

$$\Delta L = \ln(E/2\pi)/r \ln p \tag{22}$$

Because these wavelengths are incommensurable we expect the oscillations to have a quasirandom character. Obviously the oscillations depend on the detailed 'dynamics' (prime-period closed orbits), so that $L \gg 1/\tau^*$ is the non-universal regime.

The mean of the asymptotic oscillations is

$$\begin{aligned} \bar{V} &= \frac{1}{\pi^2}\left(\sum_{p=2}^{(E/2\pi)^{\tau^*}} p^{-1} + \sum_{r=2}^{\infty}\sum_{p=2}^{\infty} (r^2 p^r)^{-1} - \ln \tau^* + 1\right) \\ &= \frac{1}{\pi^2}[\ln\ln(E/2\pi) + 1.4009]. \end{aligned} \tag{23}$$

To leading order this is $\ln\ln E/\pi^2$, in exact agreement with the estimate by Selberg [14, 15] that the mean square part of the fluctuating part of the counting function is

$$\left\langle \left[\int_0^{x(E)} \mathrm{d}x\, \tilde{\mathrm{d}}(x)\right]^2 \right\rangle \sim \ln\ln E/2\pi^2. \tag{24}$$

This follows from (6) when written as

$$\begin{aligned} V(L;x) &= \left\langle \left[\int_0^{x+L/2} \mathrm{d}x_1\, \tilde{\mathrm{d}}(x_1) - \int_0^{x-L/2} \mathrm{d}x_2\, \tilde{\mathrm{d}}(x_2)\right]^2 \right\rangle \\ &\simeq 2\left\langle \left[\int_0^{x(E)} \mathrm{d}x\, \tilde{\mathrm{d}}(x)\right]^2 \right\rangle - 2\left\langle \int_0^{x+L/2} \mathrm{d}x_1\, \tilde{\mathrm{d}}(x_1) \int_0^{x-L/2} \mathrm{d}x_2\, \tilde{\mathrm{d}}(x_2) \right\rangle \\ &\to 2\left\langle \left[\int_0^{x(E)} \mathrm{d}x\, \tilde{\mathrm{d}}(x)\right]^2 \right\rangle \qquad \text{as } L \to \infty,\ L/x \ll 1 \end{aligned} \tag{25}$$

if it is assumed that the counting fluctuations at $x \pm L/2$ become uncorrelated as $L \to \infty$, $L/x \ll 1$.

4. Comparison with computation

Using techniques explained elsewhere [4], Odlyzko computed 10^5 consecutive zeros E_m, starting with $m = 10^{12} + 1$. Thus $x \simeq 10^{12}$ and, from (1) and (2), $E = 2.677 \times 10^{11}$. From these zeros he computed the number variance $V(L; x)$ for $0 \leqslant L \leqslant 1000$ in steps of 0.1.

Odlyzko's data are to be compared with the semiclassical values of $V(L; x)$ computed using (19). This involves τ^* which from (13), (17) and (18) must satisfy

$$\ln 2/\ln(E/2\pi) \ll \tau^* \ll 1 \qquad \text{i.e.} \qquad 0.028 \ll \tau^* \ll 1. \tag{26}$$

I chose $\tau^* = \frac{1}{4}$, so that the sum in (19) included prime powers $p^r \leqslant 449$, but checked that the curves of $V(L; x)$ against L were unaffected by reducing τ^* to 0.15 (i.e. $p^r \leqslant 37$). GUE universality should break down near $L_{\max} = \hbar\rho/T_{\min} = \ln(E/2\pi)/2\pi \ln 2 = 5.62$, and be replaced by the asymptotic oscillations, whose longest-wavelength component has $\Delta L = \ln(E/2\pi)/\ln 2 = 35.31$ (equation 22) and amplitude $1/2\pi^2 = 0.051$. The asymptotic mean (23) is $\bar{V} = 0.4659$; this differs slightly from the value $\bar{V} = 0.4663$ obtained by substituting $\sin^2 = \frac{1}{2}$ in (21), indicating residual dependence on τ^* which, however, does not affect the resolution of the graphs presented here.

Figures 1–4 show the results. In figure 1 the range is $0 \leqslant L \leqslant 2$. The exact and semiclassical V are indistinguishable, and GUE is a good approximation except near $L = 2$. Note how poor a fit to the data is the variance of the Gaussian *orthogonal* ensemble of real symmetric matrices (which would correspond to a dynamical system with time-reversal symmetry).

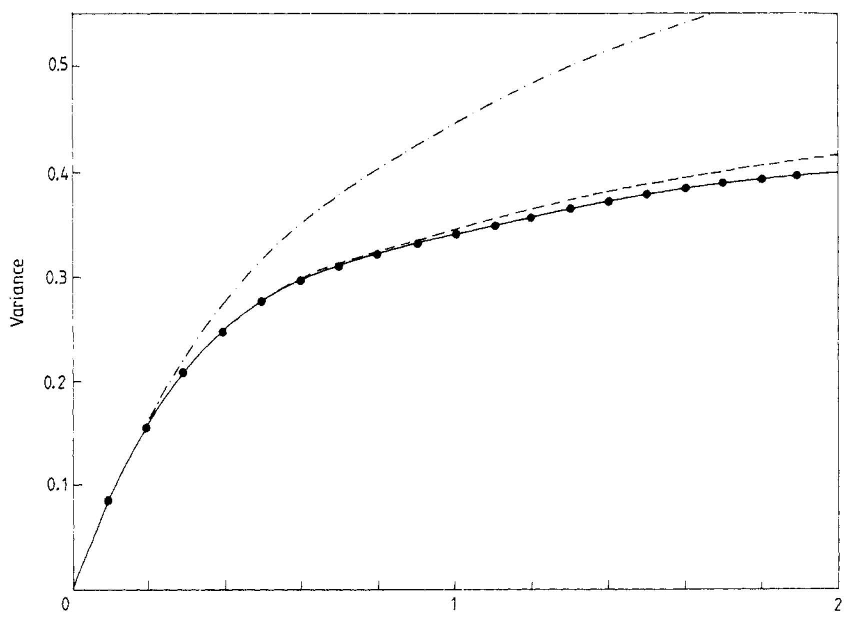

Figure 1. Number variance $V(L; x)$ of the Riemann zeros, for $0 \leqslant L \leqslant 2$ and $x = 10^{12}$. Dots: computed from the zeros by Odlyzko; full curve: semiclassical formula (19) with $\tau^* = \frac{1}{4}$; broken curve: number variance of the GUE; chain curve: number variance of the Gaussian orthogonal ensemble (GOE) of real symmetric random matrices.

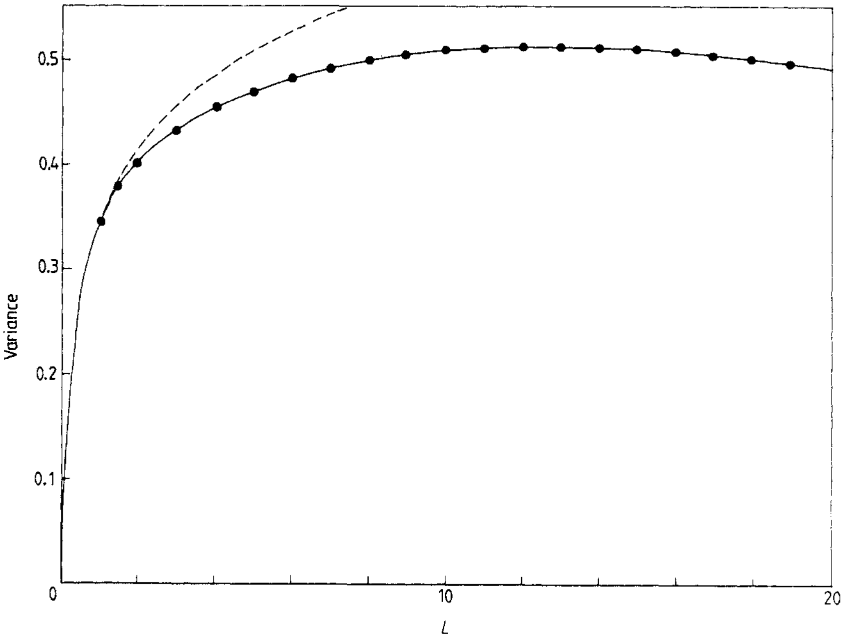

Figure 2. As figure 1 but for $0 \leqslant L \leqslant 20$ (GOE not shown).

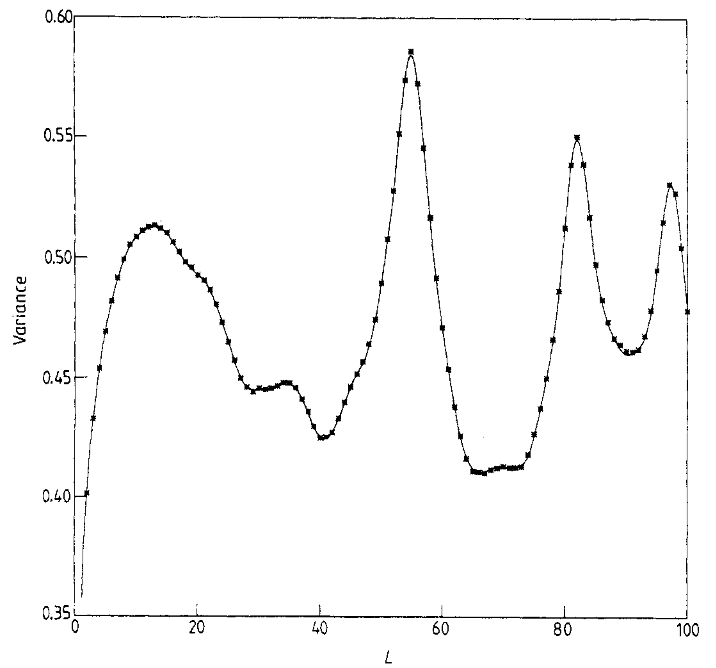

Figure 3. As figure 2 but for $0 \leqslant L \leqslant 100$ and with stars rather than dots for the variance computed from the zeros (GUE not shown).

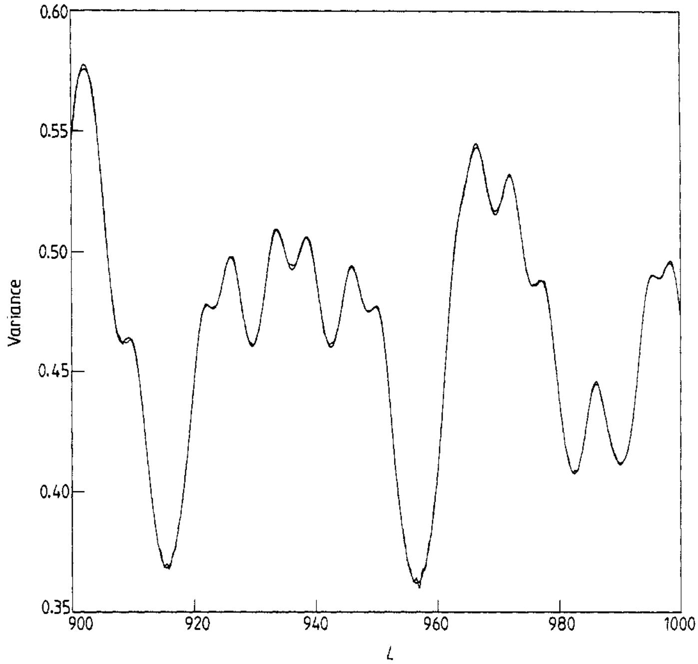

Figure 4. As figure 2 but for $900 \leqslant L \leqslant 1000$, with the 'exact' and semiclassical variances drawn as smooth curves (the 'exact' curve is more jagged) (GUE not shown).

In figure 2 the range is $0 \leqslant L \leqslant 20$. Again 'theory' and 'experiment' agree. The GUE variance ceases to be even a rough approximation when $L \sim 5$, as expected, and the maximum near $L = 13$ signals the onset of asymptotic oscillation.

In figure 3 the range is $0 \leqslant L \leqslant 100$. The asymptotic oscillations are now obvious, and quasirandom as expected with the predicted amplitudes, wavelengths and mean values. (The oscillations come from the double sum in (19), and not from the Ci and Si functions.) Slight discrepancies are visible: the 'theoretical' peaks are noticeably lower than the 'experimental' ones.

In figure 4 the range is $900 \leqslant L \leqslant 1000$. Again the quasirandom oscillations agree very well with semiclassical theory. There are, however, some rapid oscillations with $\Delta L \simeq 1$ which the semiclassical formula cannot reproduce (from (22) this would require $p^r \sim E/2\pi \sim 4 \times 10^{10}$ which is excluded from (19) because it would imply $\tau^* = 1$ in violation of (13)). The rapid oscillations are probably an artefact of averaging, because when $L = 1000$ there are only 100 independent samples.

Over the whole range $0 \leqslant L \leqslant 1000$ the largest difference between the 'exact' and semiclassical variances is 0.003 (over the range $0 \leqslant L \leqslant 100$ it is 0.002). It therefore appears that semiclassical theory gives a uniformly accurate description of the correlations in the heights of the zeros in both the universal and non-universal ranges—at least for this statistic. It is worth remarking that the transition from universal to non-universal spectral statistics has not yet been seen in any honestly quantum Hamiltonian with a chaotic classical limit, because not enough eigenvalues have been computed (the transition has been seen for an integrable system [5]).

To forestall premature optimism it must be explained why zeros near $r = 10^{12}$ might not be in the fully asymptotic regime. The whole analysis was based on pretending that $\zeta(\frac{1}{2} + \mathrm{i}E)$ is described for real E by the Euler product over primes, whereas of course this does not converge unless $\operatorname{Im} E < -\frac{1}{2}$. This could mean that pairs of zeros with $E_{m+1} - E_m < \frac{1}{2}$ will not be separated by the Euler formula. At height E the number $C(E)$ of zeros thus confused would be $C(E) \simeq (\frac{1}{2})\rho(E) = \ln(E/2\pi)/4\pi$. For these computations, $C \sim 2$ which is not large. To reach the fully asymptotic regime of large C it is necessary to go to much greater heights: thus $C = 10$ when $E = 2 \times 10^{55}$ (level number 5×10^{56}) and $C = 100$ when $E = 4 \times 10^{546}$ (level number 8×10^{548}). Note however that C increases in the same way as the limit $L_{\max}$ of GUE universality, so the asymptotic oscillations in $V(L; x)$ should survive as $x \to \infty$. The best hope is that in spite of being based on the Euler product the semiclassical formula (19) nevertheless gives the variance correctly, by being the analytic continuation to real x of some complex-x generalisation of the definition (6). (Titchmarsh's theorem 14.21 in [15] might provide the starting point for such a justification.)

Acknowledgment

It is a pleasure to thank Dr Andrew Odlyzko of the AT and T Bell Laboratories for his kindness in making available his unpublished computations of the Riemann zeros and their number variance.

References

[1] Edwards H M 1974 *Riemann's zeta function* (New York: Academic)
[2] Montgomery H L 1973 The pair correlation of the zeros of the Riemann zeta function *Proc. Symp. Pure Math.* **24** (Providence, RI: AMS) 181–93
[3] Porter C E (ed) 1965 *Statistical theories of spectra: fluctuations* (New York; Academic)
[4] Odlyzko A M 1987 On the distribution of spacings between zeros of the zeta function *Math. Comp.* **48** 273–308
[5] Berry M V 1985 Semiclassical theory of spectral rigidity *Proc. R. Soc.* A **400** 229–51
[6] Berry M V 1986 Riemann's zeta function: a model for quantum chaos? *Quantum chaos and statistical nuclear physics* (*Springer Lecture Notes in Physics No.* 263) eds T H Seligman and H Nishioka (Berlin: Springer) pp 1–17
[7] Berry M V 1987 Quantum Chaology *Proc. R. Soc.* A **413** 183–98
[8] Gutzwiller M C 1971 Periodic orbits and classical quantization conditions *J. Math. Phys.* **12** 343–58
[9] Gutzwiller M C 1978 Path integrals and the relation between classical and quantum mechanics *Path integrals and their applications in quantum, statistical and solid-state physics* ed G J Papadopoulos and J T Devreese (New York: Plenum) pp 163–200
[10] R Balian and C Bloch 1972 Distribution of eigenfrequencies for the wave equation in a finite domain: III. Eigenfrequency density oscillations *Ann. Phys., NY* **69** 76–160
[11] Hannay J H and Ozorio de Almeida A M 1984 Periodic orbits and a correlation function for the semiclassical density of states *J. Phys. A: Math. Gen.* **17** 3429–40
[12] Mehta M L 1967 *Random matrices and the statistical theory of energy levels* (New York: Academic)
[13] Abramowitz M and Stegun I A 1964 *Handbook of mathematical functions* (Washington, DC: US Govt Printing Office)
[14] Selberg A 1946 Contributions to the theory of the Riemann zeta-function *Arch. Math. Naturvid.* **B 48** 89–155
[15] Titchmarsh E C 1951 *The theory of the Riemann zeta-function* (Oxford: Oxford University Press)