The Gaussian orthogonal ensemble with missing and spurious levels: A model for experimental level-spacing distributions

Kevin K. Lehmann
Frick Chemical Laboratories, Princeton University, Washington Road, Princeton, New Jersey 08544

Stephen L. Coy
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

(Received 27 January 1987; accepted 7 July 1987)

This paper reports analytical approximations to the level-spacing distributions for two types of eigenspectra: a Gaussian orthogonal ensemble (GOE) eigenspectrum, a fraction of whose levels are omitted, and the superposition of a fractional GOE distribution with levels from a random sequence. The combined distribution would be expected when a spectrum with a GOE spacing distribution is observed, but with insufficient dynamic range to observe all the lines, and with the possibility of contamination from other sequences or from noise identified as spectral lines. Berry recently demonstrated that GOE type spacing statistics are expected in the semiclassical limit of a classically chaotic Hamiltonian. Therefore, the present formula may prove useful in testing for quantum signatures of classical chaos in quantum molecular spectra.

The study of the statistical properties of eigenvalue spectra has a long history in nuclear physics, but has recently become of interest in molecular physics as well (see Ref. 1 for an extensive review). This has been motivated by the possible close association of the statistical properties of an eigenvalue spectrum and the type of classical dynamics on the same Hamiltonian. In the semiclassical limit, the generic integrable classical Hamiltonian is expected to have a quantum energy level spectrum with uncorrelated eigenvalues. Such a spectrum has Poisson level spacings and no spectral rigidity. The limit of a strongly chaotic classical Hamiltonian is expected to have an eigenvalue spectrum with level spacing statistics very similar to that of the Gaussian orthogonal ensemble (GOE) of random matrices. The GOE spectrum displays level repulsion and long-range eigenvalue correlation or spectral rigidity. Many of the statistical properties of GOE eigenvalues have been determined analytically. Numerical evidence suggests that the fluctuation properties of the GOE spectrum may be characteristic of a large class of Hamiltonians, and thus the lack of physical basis for the GOE does not prevent it from being a useful model. Most classical Hamiltonians are neither strongly chaotic nor integrable. There is little definite information on the spacing distributions expected for intermediate situations, although Berry and Robnik have derived the form which corresponds to the superposition of several independent distributions, each corresponding to either classically chaotic or classically regular motion.

A number of experimental tests of spectral statistical properties have been developed that are matched to the very high quality data available on nuclear energy levels from neutron scattering experiments. The molecular spectra that are available for testing statistical properties are inferior, at least at present. As a result, tests have to be reexamined to see how they are corrupted when observed spectra have significant fractions of missing and spurious levels. We consider here the level spacing distribution function, which is a basic measure of short-range order, or level repulsion. The measure of long-range order or spectral rigidity developed by Levandier et al. may be less sensitive to flawed spectra since it operates on the Fourier transform of the entire recorded spectrum, but will still be sensitive to the quality of the spectra. Long-range and short-range order are only weakly connected. The next neighbor distribution is a measure of eigenstate repulsion. For chaotic spectra, this repulsion is assumed to come from strong coupling of all the basis states. This is at least superficially related to the concept of avoided crossings. In fact, an early attempt to define "quantum chaos" described it as many overlapping avoided crossings. The origin of long-range order (spectral rigidity) is less clear. We conclude that determining the degree of eigenvalue repulsion is an independent and useful complement to determining spectral rigidity.

This paper reports analytical approximations to the level spacing distributions for a Gaussian orthogonal ensemble (GOE) eigenspectrum, a fraction of whose levels are omitted, and for the superposition of the fractional GOE distribution with levels from a random sequence. The combined distribution is expected when a spectrum with a GOE spacing distribution is observed, but with insufficient dynamic range to observe all the lines, and with the possibility of contamination from other sequences, or from noise identified as spectral lines. We have used this distribution to compare with double resonance data on the NO₂ optical spectrum. The observed spectrum has many missing levels and a significant number of spurious levels, yet GOE statistics can be ruled out for the underlying spectrum.

APPROXIMATION TO THE GOE DISTRIBUTION WITH MISSING LEVELS

The spacing distribution function of a distribution whose statistical properties are uniform over the range being considered can be derived in terms of the n-th neighbor spac-
ing distribution function. The $n$th neighbor spacing distribution function $P(n,s)$ gives the probability $P(n,s)ds$ that the energy interval between two levels separated by $n$ intervening levels is between $s$ and $s + ds$ energy units. $P(0,s)$ is the next neighbor distribution function. When a fraction $f$ of the levels is observed at random, the expected next neighbor distribution function $p(s,f)$ is

$$p(s,f) = f \sum_{n=0}^{\infty} (1-f)^n P(n,s).$$  \hspace{1cm} (1)

In order to produce a useful approximation to $p(s,f)$ for the GOE distribution, we will use the known means and variances as well as tabulated values of the $n$th neighbor functions. If the original density of levels is $\rho$, then the mean value of $n + 1$ spacings is $(n + 1)/\rho$. For a GOE spectrum, the variance of $P(n,s)$, $\sigma^2(n)$, is given to excellent approximation by (Brody et al.)$^1$

$$\rho^2 \sigma^2(n) = \frac{2}{\pi^2} \{ \ln[2\pi(n + 1)] + \gamma + 1 - \pi^2/8 \} - 1/6. \hspace{1cm} (2)$$

$P(n,s)$ is known to become Gaussian in shape for large $n$. Mehta and des Cloizeaux$^8$ have derived a procedure for evaluation of $P(n,s)$ for any $n$, but the formulas are difficult to use, so we use a simpler level of approximation. Though $P(0,s)$ is known to be close to the Wigner surmise $P(0,s) = (rs/2\pi) \exp[-r^2/4]$, to be more accurate, we used the exact values tabulated by Gaudin,$^9$ using a cubic spline to interpolate between points in Table I. Initially, we estimated $P(1,s)$ by interpolation from the table of points calculated by Kahn.$^{10}$ A plot of $P(2,s)$ is given in Fig. 9 of the review article by Brody et al.$^1$ and already appears fairly Gaussian in shape. We have estimated $P(n,s)$ for $n \geq 2$ by a Gaussian with mean $(n + 1)/\rho$ and variance $\sigma^2(n)$. As a test of these approximations, we note that

$$\sum_{n=0}^{\infty} P(n,s) = \rho^2 \{ 1 - Y_{2,1}(s) \}. \hspace{1cm} (3)$$

$Y_{2,1}(s)$ is the two level cluster function and an expression to evaluate it is given by Brody et al. [Eq. (1.1)]. The calculated $\Sigma P(n,s)$ showed deviations of a few percent from the right-hand side of Eq. (3), and these deviations were only significant in the range of $s$ where $P(2,s)$ had significant intensity. Errors of this size in $P(2,s)$ are perhaps not unexpected given the rounding problems in computing $P(2,s)$, which involves computing a numerical second derivative, and given the small number of points in Kahn's table. We chose to use Eq. (3), and our approximations for the $P(0,s)$ and $P(n,s)$, $n \geq 2$ to compute $P(1,s)$ for $s = 0.01*n$, and a cubic spline to interpolate between these points. The $P(1,s)$ thus computed has very close to the correct width as well as correct mean. A similar method of calculation of $p(s,f)$ was used by Watson et al.$^{11}$ in the analysis of imperfect sequences of nuclear levels.

While the above procedure generates a $p(s,f)$ function, it is rather cumbersome and slow to use in a fitting of experimental data. Further, the numerical form of the results make it difficult to include the effects of other statistically independent sequences (see the next section). What is needed is an accurate, analytical approximation to the computed $p(s,f)$ functions. We have fitted the calculated $p(s,f)$ distributions to a generalization of the Brody distribution function$^2$ with two adjustable exponents: $D(s,u,v) = A(s+\alpha)^\nu \exp[ - (s/\alpha)^\lambda ]$. Since the distributions to be fitted, $p(s,f)$, are normalized and have known mean $1/(f \rho s)$, the parameters $A$ and $\alpha$ may be expressed in terms of $u$ and $v$:

$$\alpha = \frac{\Gamma[(u + 1)/v]}{(\rho s \Gamma[(u + 2)/v])};$$

$$A = \frac{\nu}{(\alpha \Gamma[(u + 1)/v])}. \hspace{1cm} (4)$$

If we constrain $v = u + 1$ we have the Brody distribution,$^{13}$ but we have found that the two parameter form is much more accurate in fitting model and experimental distributions.$^{12}$ Table I gives $u$ and $v$ exponents which result from fits of $p(s,f)$ to $D(s,u,v)$ over the range of 0 to 5/f mean level spacings ($1/f$ is the observed mean spacing) for 21 values of $f$ from 0 to 1.0. The rms deviations of the fits are typically less then 0.01, and in the worst case ($f = 0.10$) was 0.016. Plots of the calculated and best fit curves show that this form does quite well at reproducing the proper shapes. Figure 1 is a plot of the calculated and fitted curves for $f = 0.50$. Figure 2 is a plot of the two parameters as a function of the observed fraction of levels $f$. A polynomial expansion of the best fit values of $u$ and $v$ as a function of $f$ converges slowly. The fitted form $D(s,u,v)$ is a convenient way to estimate $p(s,f)$ for any fraction of missing levels.

**SPACING DISTRIBUTION WITH RANDOM SPURIOUS LEVELS**

Since the GOE eigensystem allows coupling between all unperturbed levels, comparisons between observed spectra and the GOE distribution require that levels with the same good quantum numbers be selected, that is levels in a pure sequence. In addition to the problem of missing levels in the

<table>
<thead>
<tr>
<th>$f$</th>
<th>$U$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.01</td>
<td>0.2653</td>
<td>0.7831</td>
</tr>
<tr>
<td>0.05</td>
<td>0.5142</td>
<td>0.6749</td>
</tr>
<tr>
<td>0.10</td>
<td>0.7599</td>
<td>0.6101</td>
</tr>
<tr>
<td>0.20</td>
<td>1.0035</td>
<td>0.5679</td>
</tr>
<tr>
<td>0.30</td>
<td>1.2343</td>
<td>0.5419</td>
</tr>
<tr>
<td>0.35</td>
<td>1.4391</td>
<td>0.5294</td>
</tr>
<tr>
<td>0.40</td>
<td>1.6045</td>
<td>0.5294</td>
</tr>
<tr>
<td>0.45</td>
<td>1.7222</td>
<td>0.5413</td>
</tr>
<tr>
<td>0.50</td>
<td>1.7938</td>
<td>0.5639</td>
</tr>
<tr>
<td>0.55</td>
<td>1.8173</td>
<td>0.5983</td>
</tr>
<tr>
<td>0.60</td>
<td>1.8012</td>
<td>0.6443</td>
</tr>
<tr>
<td>0.65</td>
<td>1.7528</td>
<td>0.7027</td>
</tr>
<tr>
<td>0.70</td>
<td>1.6805</td>
<td>0.7747</td>
</tr>
<tr>
<td>0.75</td>
<td>1.5920</td>
<td>0.8616</td>
</tr>
<tr>
<td>0.80</td>
<td>1.4944</td>
<td>0.9653</td>
</tr>
<tr>
<td>0.85</td>
<td>1.3922</td>
<td>1.0886</td>
</tr>
<tr>
<td>0.90</td>
<td>1.2895</td>
<td>1.2346</td>
</tr>
<tr>
<td>0.95</td>
<td>1.1886</td>
<td>1.4079</td>
</tr>
<tr>
<td>1.00</td>
<td>1.0004</td>
<td>1.6132</td>
</tr>
</tbody>
</table>

---

J. Chem. Phys., Vol. 87, No. 9, 1 November 1987
pure sequence that we just considered, levels with other quantum numbers may be incorrectly included in the sequence or some of the identified states may just be noise. Since the spacing distribution of any sequence becomes Poisson if only a small fraction of the elements are selected at random, the few incorrectly included levels should have a Poisson distribution. Noise identified as spectral lines may also be assumed to have a Poisson distribution. Berry and Robnik\textsuperscript{4} have derived the effect of mixing an independent sequence of Poisson distributed levels with a pure GOE spectrum. Their result cannot be used if a significant fraction of the GOE levels are missing, but their method can be used to derive the result of combining two known independent distributions. Using the generalized Brody distribution approximation for the spacing distribution of a fractional GOE spectrum, we can follow the methods of Berry and Robnik and arrive at an analytical result for the spacing distribution of that sequence with a superimposed Poisson distribution.

We start with a set of observed levels with density $\rho_1$ and spacing distribution given by the generalized Brody form above and a set of levels from a random sequence with density $\rho_2$. Because these levels are uncorrelated their spacing distribution is a negative exponential distribution $[P_2(s) = \rho_2 \exp(-\rho_2 s)]$. Following Berry and Robnik we arrive at the spacing distribution function for the overlap of the two spectra as

$$P_{\text{tot}}(s) = \frac{P_1}{\rho_1 + \rho_2} e^{-\rho_1 s} \left[ \frac{\rho_2^2 A \alpha}{v} \Gamma \left( \frac{u + 2}{v} x \right) - s \Gamma \left( \frac{u + 1}{v} x \right) \right] + 2 \rho_2 \frac{A \alpha}{v} \Gamma \left( \frac{u + 1}{v} x \right) + A (s/\alpha)^n e^{-s} ,$$  \hspace{1cm} (5)

where $\Gamma(a,x)$ is the incomplete Gamma function\textsuperscript{14}:

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig1.png}
\caption{Parameters $u$ and $v$ in the generalized Brody approximation to the fractional GOE distributions plotted against the fraction of the GOE distribution observed.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig2.png}
\caption{Comparison of calculated and fitted next neighbor spacing function for GOE fraction $f = 0.5$. The solid line is the calculated curve, the dashed line is the fitted one.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig3.png}
\caption{The calculated next neighbor spacing functions for GOE fraction $f = 0.5$ with the fraction of levels from a random sequence being 0, 0.2, 0.4, 0.6, and 0.8 in order of increasing intercept.}
\end{figure}
\[ \Gamma(a,x) = \int_{x}^{\infty} e^{-t} t^{a-1} \, dt. \]

In Fig. 3 we plot the calculated distributions for several fractions from a random sequence \([\rho_2/(\rho_1 + \rho_2)]\), all for the cases with 50% of the GOE levels in the initial distribution omitted.

The present results make it possible to calculate the expected next neighbor distribution for a GOE spectrum with missing and/or spurious levels. This extends the earlier work of Berry and Robnik that was restricted to the overlap of complete GOE spectra. Given the limited statistics of present experiments, it is unlikely that fits that allow \(f, \rho_1\), and \(\rho_2\) to vary will determine all three parameters with reasonable accuracy. The value of this work lies in cases where it is possible to critically examine the experiment to estimate \(f\) and \(\rho_2/(\rho_1 + \rho_2)\). In our experimental work on NO\(_2\), we estimate that our data set contains at least 60% of a complete sequence and no more than 20% spurious levels. If the full pure sequence was a GOE spectrum, we predict an observed spacing distribution function as in Fig. 4. The experimentally observed distribution is shown in Fig. 5. We see that despite the limitations with the data set we can confidently state that the underlying complete pure sequence is definitely not a GOE spectrum. We hope our present results will be of use in other attempts to look for level repulsion in atomic and molecular spectra, such as the recent experiments of Wunner et al.\(^{15}\)

**ACKNOWLEDGMENTS**

This work was supported in part by grants from the Donors of the Petroleum Research Fund, administered by the American Chemical Society, and by the National Science Foundation.


\(^{10}\) P. B. Kahn, Nucl. Phys. 41, 159 (1963).


\(^{14}\) M. Abramowitz and I. A. Stegun, U.S. Dept. of Commerce Applied Math. Series 55 (1964), Sec. 6.5. The IMSL library (IMSL Inc., Houston, TX) contains routines to compute \(\Gamma(a)\) and \(\gamma(a,x) = \Gamma(a) - \Gamma(a,x)/\Gamma(a)\) (called the gamma distribution function) from which \(\Gamma(a,x)\) can be evaluated.