

Assume nothing

Eugene Wigner wrote famously about the unreasonable effectiveness of mathematics in science. His own ideas illustrate the point as well as any others. In 1956, at a conference on neutron physics in Gatlinburg, Tennessee, Wigner spoke on the energy levels of large, complex nuclei such as uranium, for which data were just becoming available. He expressed the view that a good deal might be learned by making a virtue of theoretical ignorance, and simply assuming random values for the elements of the Hamiltonian matrix, which in quantum theory determines the nuclear energy levels through its eigenvalues.

Wigner showed that this 'simple minded' approach could establish baseline expectations for the spacing of nuclear levels in the absence of any other knowledge. "The question", he noted, "is simply, what are the distances of the characteristic values of a symmetric matrix with random coefficients?" Wigner's result, worked out in a few lines of algebra, gave a probability distribution of the form $p(x) \sim x \exp(-ax^2)$, with x being the energy spacing and $a = \pi/4$, thereby pointing to a dearth of levels of similar energy. The result contrasted sharply with what might have been the expected Poisson form, $p(x) \sim \exp(-x)$, for which $x = 0$ would be a maximum rather than a minimum.

Wigner's random-matrix approach was indeed effective, as experiments over the next few years — especially those probing resonances in neutron scattering from uranium — showed a remarkably close fit to his predicted curve. Much more surprising, perhaps, has been the enormous influence of random-matrix theory since then in areas ranging from pure mathematics to the study of financial risk. All, one might say, by assuming almost nothing and then working out the consequences.

In the early 1970s, Freeman Dyson and Hugh Montgomery stumbled over a connection between Wigner's idea and pure mathematics. Montgomery had been studying the famous Riemann zeta function, defined as $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$, which has some zeros located along the 'critical line' defined by $s = 1/2$. It is unknown whether all zeros lie on this line — the assertion that they are is known as the Riemann hypothesis — or how the zeros are distributed. But it is known that the distribution of such zeros can be linked to the distribution of prime numbers, and hence holds fundamental importance for number theory.



Wigner expressed the view that a good deal might be learned by making a virtue of theoretical ignorance.

Montgomery told Dyson that he'd been studying the pair-correlation function for the zeros of $\zeta(s)$, arriving at an estimate for its asymptotic form of $1 - (\sin(\pi x)/\pi x)^2$, which Dyson recognized immediately as the pair-correlation function for the eigenvalues of random matrices of the form studied by Wigner. To this day, it is not known why the same equation should turn up in the distributions of both nuclear energy levels and the roots of $\zeta(s)$, and by implication in the distribution of the prime numbers, but computations suggest that the connection is remarkably precise. Mathematician Andrew Odlyzko has computed the locations of zeros of $\zeta(s)$ along the critical line, identifying as many as 10^{23} zeros and finding a near-perfect agreement between the predicted and measured correlations. The zeroes of $\zeta(s)$ effectively repel one another much as do nuclear energy levels (see, for example, B. Hayes, *American Scientist* July–August 2003).

One intriguing possibility is that this connection points to some mysterious relationship between $\zeta(s)$ and the mathematics of quantum theory. Even before the advent of quantum theory, mathematicians David Hilbert and George Pólya independently suggested — apparently inspired by some loose analogy to the discrete energy spectra recently discovered in atoms — that the zeroes of $\zeta(s)$ might be given by the eigenvalues of some unknown Hermitian matrix. If so, then some Hermitian operator of the kind familiar to all physicists may determine the positions of the Riemann zeroes and, with them, the distribution of the primes — certainly a strange link between physics and mathematics.

Random-matrix theory has now been applied widely in statistics, condensed-matter physics and elsewhere. But its practically most important applications may be still to come, especially in extracting meaningful information from huge quantities of data.

The most obvious way to look for cause-and-effect relationships in data, of course, is to identify correlations among variables, which may point to mechanisms of causal

influence, or be useful for making predictions. With modern technology, the automated screening of enormous volumes of high-dimensional data for interesting correlations has become a key tool of science, especially in molecular biology, environmental science, economics and finance.

But this practice faces a fundamental problem. As the number of variables being studied grows, the number of pairs of variables among which correlations might be found grows even faster. In this case, straightforward calculation is almost certain to detect correlations that look significant but really aren't. Suppose, for example, you have M input variables and N output variables — you might think of inflation, employment, a stock market index and any other number of economic quantities. Suppose you have time series for all of these quantities over time T , and you look for correlations between inputs and outputs. Then, even if these variables were all independent with Gaussian fluctuations, one would expect that the largest observed correlation — if you calculate them all — will be of the order $(\ln(MN))/T$ (see, for example, J.-P. Bouchaud *et al. Eur. Phys. J. B* 55, 201–207; 2007), which gets large for any fixed T given enough variables to study.

This has become known as the 'curse of dimensionality'. Automation makes it easy to study everything, and too easy to find meaningless patterns in doing so. Fortunately, however, here too Wigner's random-matrix approach has proven useful, in this case for separating what is meaningful from what is nonsense. Correlations calculated between truly independent variables should produce a random $M \times N$ matrix. Studying the typical spectral features of such random matrices helps to establish baseline expectations for the correlations likely to be produced by statistical fluctuations alone. In particular, studies have shown that the eigenvalues of truly random matrices tend to be confined within an interval with sharp edges. Hence, eigenvalues in empirical data found to stand out and away from these edges should signify real and meaningful correlations.

Undoubtedly, further uses will be found for Wigner's random matrices, which in truth were invented long before Wigner by people interested in correlations in empirical data. But Wigner's 1956 work stimulated the deeper mathematical analysis of such matrices, with repercussions that almost anyone would call surprising — even, perhaps, unreasonable. □

MARK BUCHANAN