Universality in Numerical Computations with Random Data. Case Studies.

Percy Deift and Thomas Trogdon
Courant Institute of Mathematical Sciences

Govind Menon
Brown University

Sheehan Olver
The University of Sydney

Eigenvalue Algorithms

In earlier work [2012][1], P.D. and G.M., together with C. Pfrang, considered the problem of computing the eigenvalues of real $n \times n$ random symmetric matrices $M = (M_{ij})$.

They considered matrices chosen from different ensembles E using a variety of different algorithms A.

Let $S_n = \{M = M^T : M \in \mathbb{R}^{n\times n}\}$

Standard algorithms utilize isospectral maps:

- $\varphi = \varphi_A : S_N \to S_N$
  $\operatorname{spec} \varphi_A(M) = \operatorname{spec} M$
- Given $M \in S_n$, $M_{k+1} \equiv \varphi_A(M_k)$, $k \geq 0$, $M_0 = M$
- As $k \to \infty$, $M_k \to \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$
- Necessarily, the λ_i's are the desired eigenvalues of M.

[1] Pfrang, CW, P Deift, and G Menon. How long does it take to compute the eigenvalues of a random symmetric matrix? *arXiv Prepr. arXiv1203.4635*, 2012

Universal Fluctuations

In their paper, the authors discovered the following phenomenon:

For a given accuracy ϵ, a given matrix size n (ϵ small, n large) and a given algorithm A, the **fluctuations** in the time to compute the eigenvalues to accuracy ϵ with the given algorithm A, were **universal**, independent of the choice of ensemble E. More precisely, they considered fluctuations in the **deflation time** T.

Recall that if an $n \times n$ matrix has block form

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

where M_{11} is $k \times k$ and M_{22} is $(n-k) \times (n-k)$ for some $1 \leq k \leq n-1$ then one says that the block diagonal matrix

$$\hat{M} = \begin{pmatrix} M_{11} & 0 \\ 0 & M_{22} \end{pmatrix}$$

is **obtained from M by deflation**.

If $\|M_{12}\| = \|M_{21}\| \leq \epsilon$, then the eigenvalues $\{\lambda_i\}$ of M differ from the eigenvalues $\{\hat{\lambda}_i\}$ of $\hat{M}$ by $\mathcal{O}(\epsilon)$.

Universal Fluctuations

Let $T = T_{\epsilon,n,A,E}(M)$ be the time (= # of steps = # iterations of φ_A) it takes to deflate a random matrix M, chosen from an ensemble E, to order ϵ, using algorithm A, **i.e.** T is the smallest time such that for some k, $1 \leq k \leq n-1$,

$$\|(\varphi_A^{(T)}(M))_{12}\| = \|(\varphi_A^{(T)}(M))_{21}\| \leq \epsilon.$$

The fluctuations $\tau_{\epsilon,n,A,E}(M)$ of T are defined by

$$\tau_{\epsilon,n,A,E}(M) = \frac{T_{\epsilon,n,A,E}(M) - \langle T_{\epsilon,n,A,E} \rangle}{\sigma_{\epsilon,n,A,E}},$$

where $\langle T_{\epsilon,n,A,E} \rangle$ is the sample average of $T_{\epsilon,n,A,E}(M)$ taken over matrices M from E, and $\sigma^2_{\epsilon,n,A,E}$ is the sample variance.

For a given E, a typical sample size in [2012] was of order 5,000 to 10,000 matrices M, and the output of the calculations in [2012] was recorded in the form of a histogram for $\tau_{\epsilon,n,A,E}$.

Universal Fluctuations

A sample of the histograms from [2012] when $A = \mathrm{QR}$ is given below: For $M = QR$, Q orthogonal, R upper triangular, $R_{ii} > 0$,

$$\varphi_A(M) = \varphi_{\mathrm{QR}}(M) \equiv RQ = Q^T M Q.$$

Two-component universality is evident.

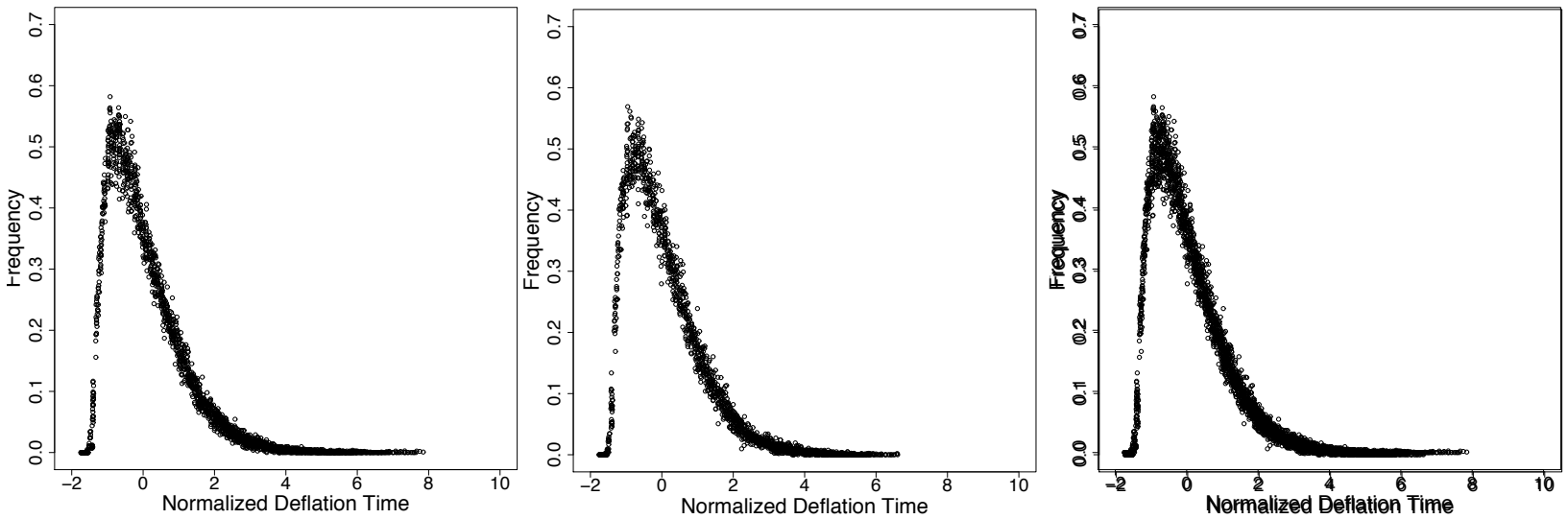


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A = \mathrm{QR}$. This figure is taken from [2012]. Overlayed histograms demonstrate the collapse of the histogram of $\tau_{\epsilon,n,A,E}$ to single curve. In the left figure, $E = \mathrm{GOE}$, and 40 histograms for $\tau_{\epsilon,\epsilon,A,E}$, are plotted one on top of the other for $\epsilon = 10^{-k}$, $k = 2,4,6,8$ and $n = 10,30,\ldots,190$. The histograms are created with $\approx 10{,}000$ samples. The middle figure displays the same information as that in the left position, but now for $E = BE$. In the right figure, all $40 + 40$ histograms are overlayed and universality is evident: the data appears to follow a universal law for the fluctuations.

The Gaussian Orthogonal Ensemble (GOE) is given by $(X + X^T)/\sqrt{4n}$ where X is an $n \times n$ matrix of iid Gaussian variables with mean zero and variance one.

The Bernoulli Ensemble (BE) is given by an $n \times n$ matrix X consisting of iid random variables that take the values $\pm 1/\sqrt{n}$ with equal probability subject only to the constraint $X^T = X$.

Universal Fluctuations

A sample of the histograms from [2012] when $A = \mathrm{QR}$ is given below: For $M = QR$, Q orthogonal, R upper triangular, $R_{ii} > 0$,

$$\varphi_A(M) = \varphi_{\mathrm{QR}}(M) \equiv RQ = Q^T M Q.$$

Two-component universality is evident.

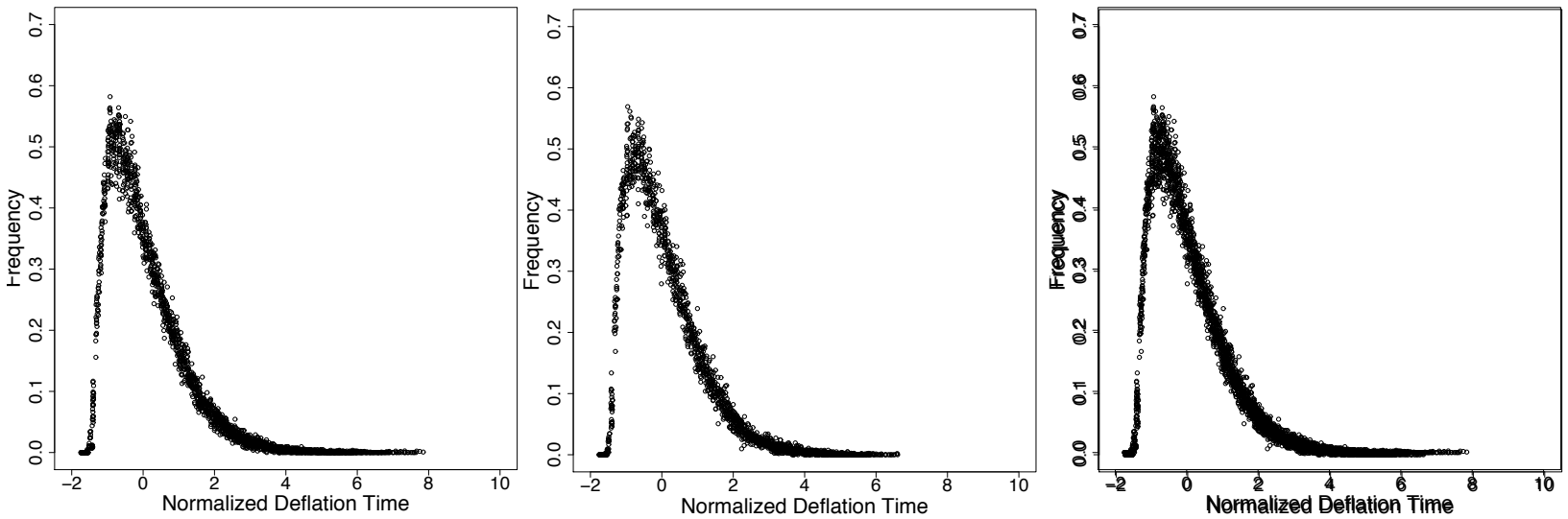


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A = \mathrm{QR}$. This figure is taken from [2012]. Overlayed histograms demonstrate the collapse of the histogram of $\tau_{\epsilon,n,A,E}$ to single curve. In the left figure, $E = \mathrm{GOE}$, and 40 histograms for $\tau_{\epsilon,\epsilon,A,E}$, are plotted one on top of the other for $\epsilon = 10^{-k}$, $k = 2,4,6,8$ and $n = 10,30,\ldots,190$. The histograms are created with $\approx 10{,}000$ samples. The middle figure displays the same information as that in the left position, but now for $E = BE$. In the right figure, all $40 + 40$ histograms are overlayed and universality is evident: the data appears to follow a universal law for the fluctuations.

In recent work, the authors have extended the results of [2012] in various ways as follows:

1. Jacobi algorithm
2. Ensembles with dependent entries
3. Conjugate gradient algorithm
4. GMRES (generalized minimal residual) algorithm
5. Discretization of a random PDE
6. Genetic algorithm
7. Curie–Weiss decision making model

The Jacobi Algorithm

Givens Rotations

For $M \in S_n$, choose $i < j$ such that $|M_{ij}| \geq \max_{1 \leq i' < j' \leq n} |M_{i'j'}|$, and let $G^{(ij)}$ be the *Givens rotation matrix*,

$$G^{(ij)} = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & -s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ \vdots & \cdots & s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

Here "$-s$" is in position (i,j), etc., and $c = \cos\theta$, $s = \sin\theta$. The angle θ is chosen so that $(G^{(ij)}(\theta)^T M G^{(ij)}(\theta))_{ij} = 0$ and then

$$\varphi_{\text{Jacobi}}(M) \equiv (G^{(ij)}(\theta))^T M G^{(ij)}(\theta).$$

The Jacobi algorithm is very different from QR-type algorithms...

The Jacobi Algorithm

In place of $T_{\epsilon,n,A,E}$, we record the *halting time* $k_{\epsilon,n,A,E}$: the number of iterations it takes for the Jacobi algorithm to reduce the Frobenius norm (the square-root of the sum of the squares) of the off-diagonal elements to be less than a given ϵ. This is sufficient to conclude that at least one element on the diagonal of the transformed matrix is within ϵ of an exact eigenvalue of the original matrix.

Histograms are produced for the fluctuations $\tau_{\epsilon,n,A,E}$:

$$\tau_{\epsilon,n,A,E}(M) = \frac{k_{\epsilon,n,A,E}(M) - \langle k_{\epsilon,n,A,E} \rangle}{\sigma_{\epsilon,n,A,E}}.$$

Jacobi Fluctuations

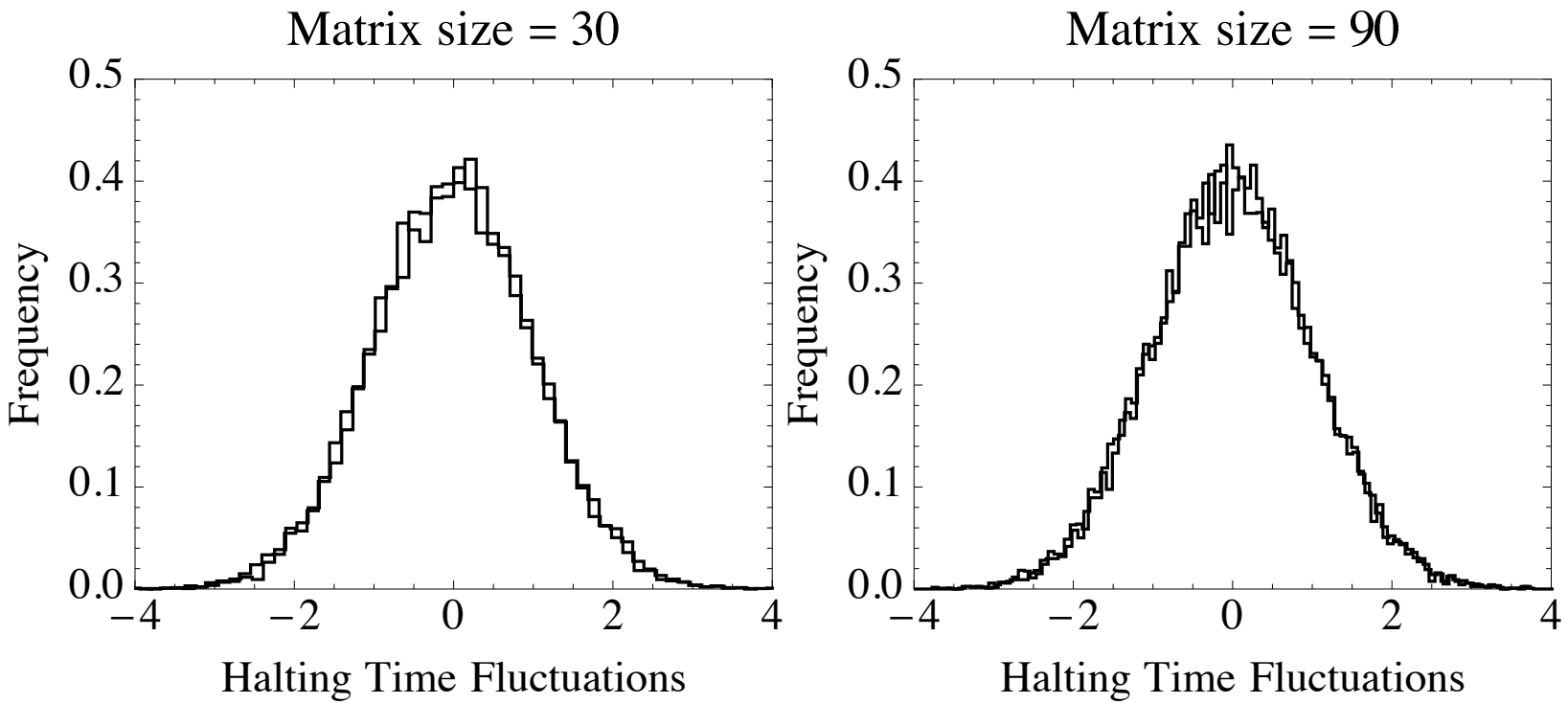


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A =$ Jacobi, $E =$ GOE, BE and $\epsilon = \sqrt{n}\,10^{-10}$. The left figure displays two histograms, one on top of the other, one for GOE and one for BE, when $n = 30$. The right figure displays the same information for $n = 90$. All histograms are produced with 16,000 samples. Again, we see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

The Gaussian Orthogonal Ensemble (GOE) is given by $(X + X^T)/\sqrt{4n}$ where X is an $n \times n$ matrix of iid Gaussian variables with mean zero and variance one.

The Bernoulli Ensemble (BE) is given by an $n \times n$ matrix X consisting of iid random variables that take the values $\pm 1/\sqrt{n}$ with equal probability subject only to the constraint $X^T = X$.

Jacobi Fluctuations

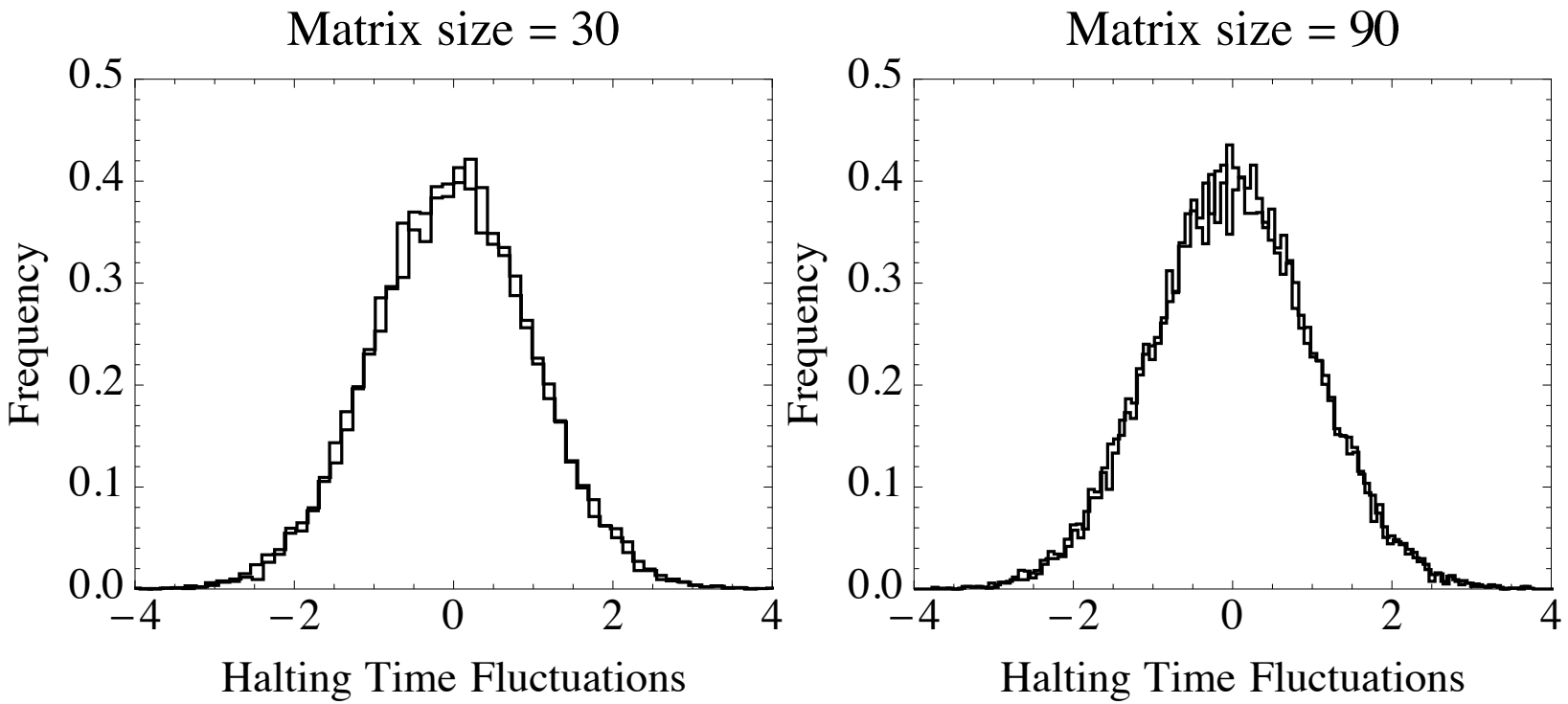


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A =$ Jacobi, $E =$ GOE, BE and $\epsilon = \sqrt{n}\,10^{-10}$. The left figure displays two histograms, one on top of the other, one for GOE and one for BE, when $n = 30$. The right figure displays the same information for $n = 90$. All histograms are produced with 16,000 samples. Again, we see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

Ensembles with Dependent Entries

In all the above calculations, M is real and symmetric with independent entries. In the following calculations we consider $n \times n$ Hermitian $M = M^*$ from various unitary invariant ensembles with distributions proportional to

$$e^{-n\mathrm{tr}V(M)}dM$$

where $V(x) : \mathbb{R} \to \mathbb{R}$ grows sufficiently rapidly. The entries are independent iff V is proportional to x^2: non-trivial matter to sample ensembles for general V (see Olver, Nadakuditi and Trogdon (2014)).

Histograms for the deflation time fluctuations are given below.

Dependent QR Fluctuations

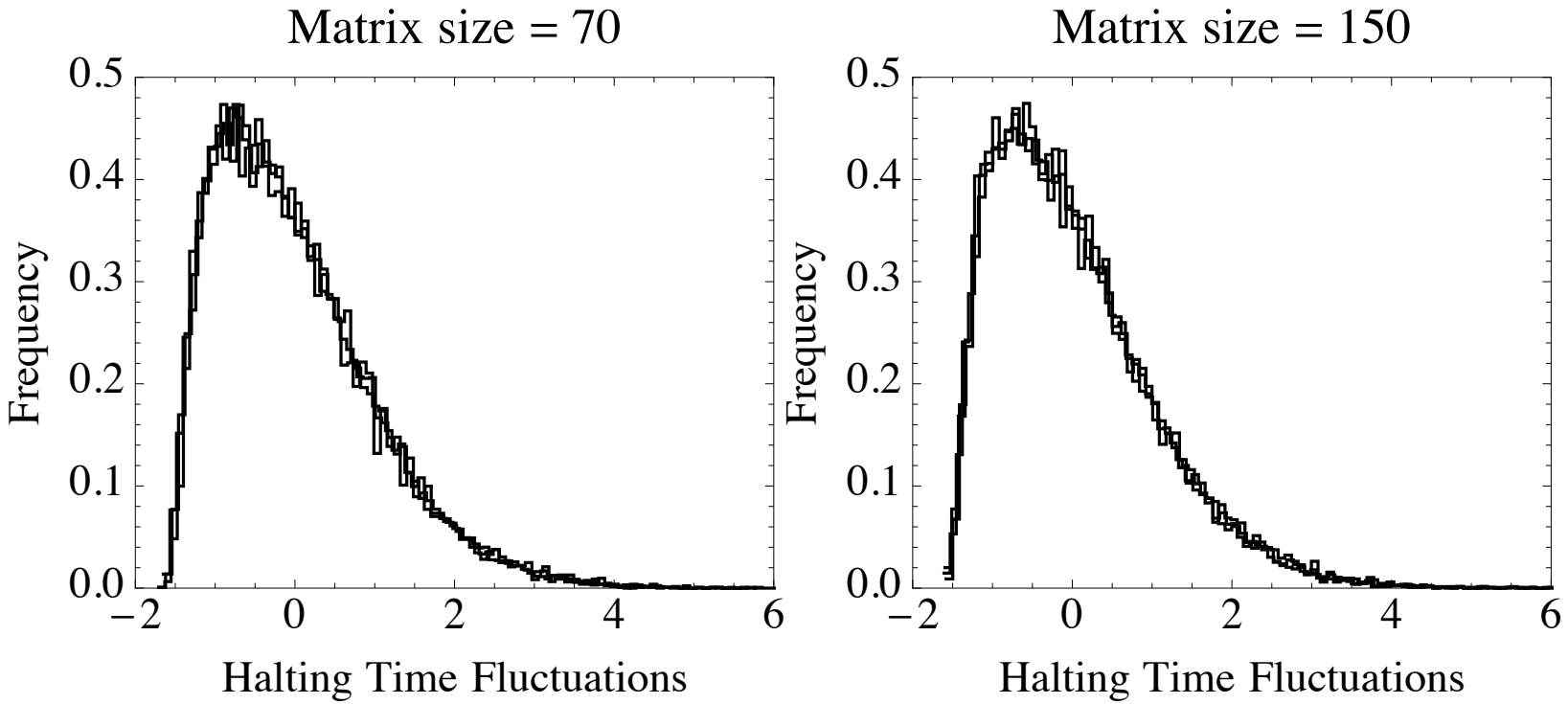


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A = \text{QR}$, $E = \text{QUE},\ \text{COSH},\ \text{GUE}$ and $\epsilon = 10^{-10}$. Here we are using deflation time (= halting time), as in [2012]. The left figure displays three histograms, one each for GUE, COSH and QUE, when $n = 70$. The right figure displays the same information for $n = 150$. All histograms are produced with 16,000 samples. Again, we see that two-component universality emerges for n sufficiently large: the histograms follow a universal (independent of E) law. This is surprising because COSH and QUE have eigenvalue distributions that differ significantly from GUE in that they do not follow the so-called *semi-circle law*.

The Gaussian Unitary Ensemble (GUE) is a complex, unitary invariant ensemble with probability distribution proportional to $e^{-n\mathrm{tr}M^2}\,dM$.

The Quartic Unitary Ensemble (QUE) is a complex, unitary invariant ensemble with probability distribution proportional to $e^{-n\mathrm{tr}M^4}\,dM$.

The Cosh Unitary Ensemble (COSH) has its distribution proportional to $e^{-\mathrm{tr}\cosh M}\,dM$.

Both QUE and COSH do not follow the semi-circle law for the global distribution of the eigenvalues.

Dependent QR Fluctuations

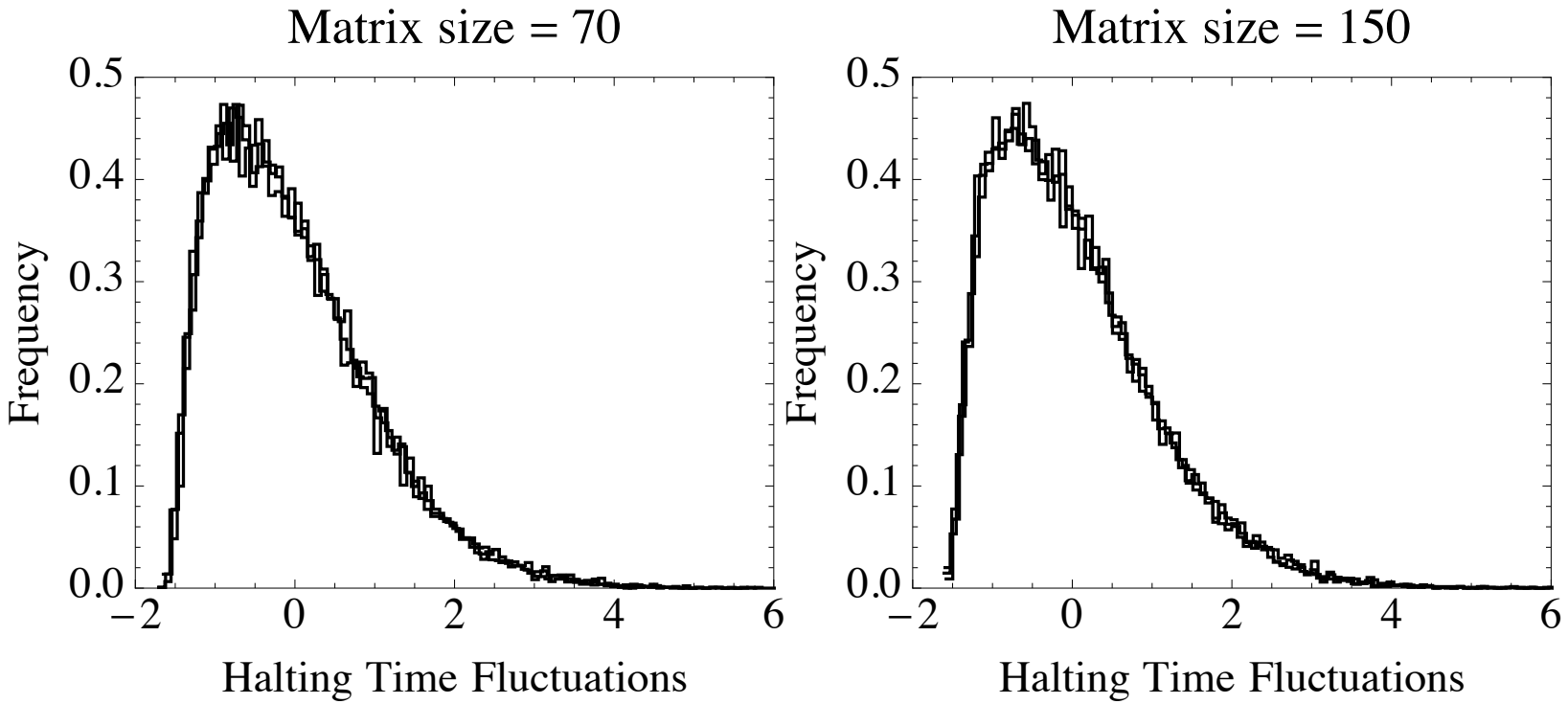


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A = \text{QR}$, $E = \text{QUE},\ \text{COSH},\ \text{GUE}$ and $\epsilon = 10^{-10}$. Here we are using deflation time (= halting time), as in [2012]. The left figure displays three histograms, one each for GUE, COSH and QUE, when $n = 70$. The right figure displays the same information for $n = 150$. All histograms are produced with 16,000 samples. Again, we see that two-component universality emerges for n sufficiently large: the histograms follow a universal (independent of E) law. This is surprising because COSH and QUE have eigenvalue distributions that differ significantly from GUE in that they do not follow the so-called *semi-circle law*.

The Conjugate Gradient Algorithm

Here the authors start to address the question of whether two-component universality is just a feature of eigenvalue computation, or is present more generally in numerical computation. In particular, the authors consider the solution of the linear system of equations $Wx = b$ where W is a real and positive definite, using the conjugate gradient (CG) method.

The method is iterative and at iteration k of the algorithm an approximate solution x_k of $Wx = b$ is found and the residual $r_k = Wx_k - b$ is computed. For any given $\epsilon > 0$, the method is halted when $\|r_k\|_2 < \epsilon$, and the halting time $k_\epsilon(W, b)$ recorded.

Here the authors consider $n \times n$ matrices A chosen from two different positive definite ensembles E and vectors $b = (b_j)$ chosen independently with iid entries $\{b_j\}$. Given ϵ (small) and n (large), and $(W, b) \in E$, the authors record the halting time $k_{\epsilon,n,A,E}$, $A = \mathrm{CG}$, and compute the fluctuations $\tau_{\epsilon,n,A,E}(W, b)$. The histograms for $\tau_{\epsilon,n,A,E}$ are given below, and again, two-component universality is evident.

Conjugate Gradient Fluctuations

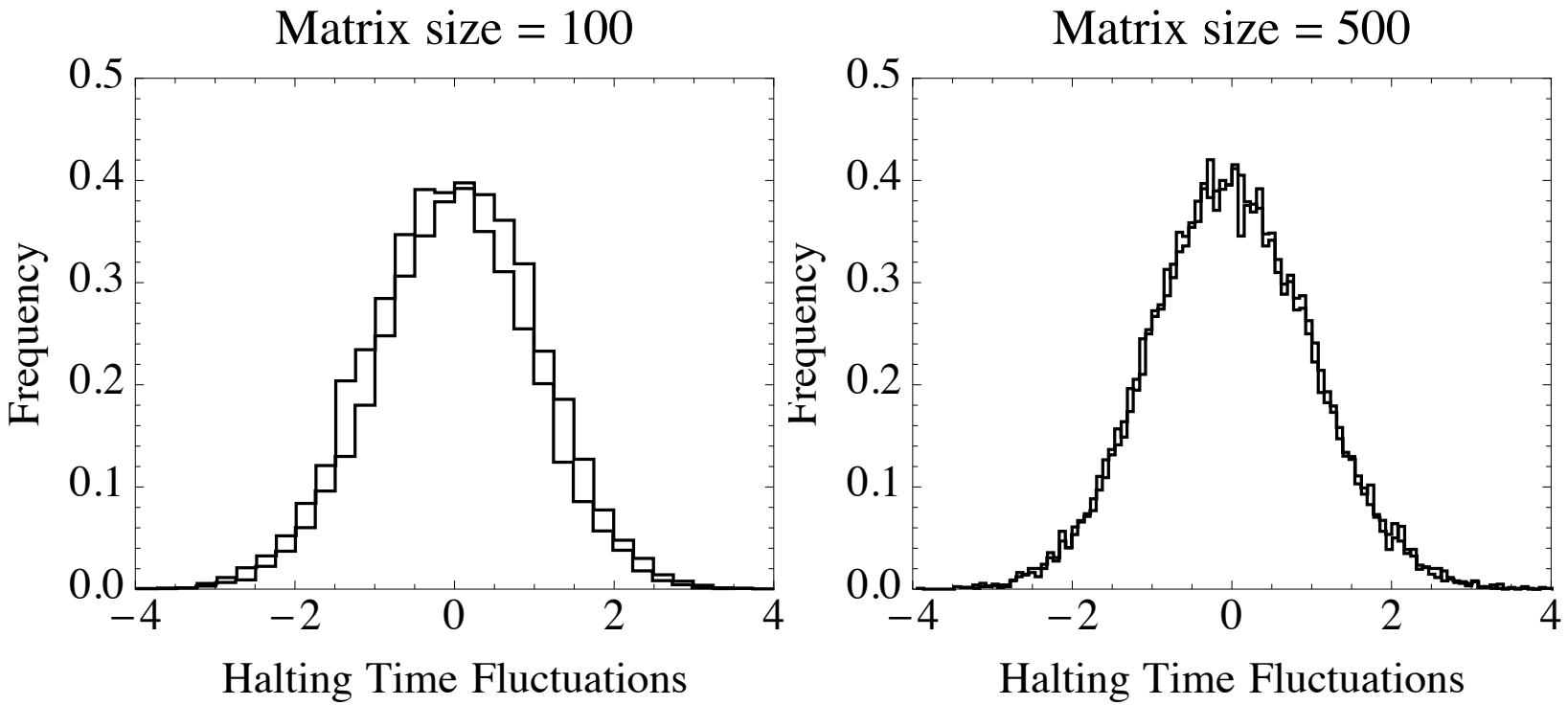


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A = \text{CG}$ and $E = \text{cLOE},\ \text{cPBE}$. The left figure displays two histograms, one for cLOE and one for cPBE, when $n = 100$. The right figure displays the same information for $n = 500$. All histograms are produced with 16,000 samples. Again, we see two-component universality emerges for n sufficiently large: the histograms follow a universal (independent of E) law. With the chosen scaling, we see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

The critically-scaled Laguerre Orthogonal Ensemble (cLOE) is given by XX^T/m where X is an $n \times m$ matrix with iid Gaussian (mean zero, variance one) entries. The critically-scaled positive definite Bernoulli ensemble (cPBE) is given by XX^T/m where X is an $n \times m$ matrix consisting of iid Bernoulli variables taking the values ± 1 with equal probability.

The critical scaling refers to the choice $m = n + 2\lfloor\sqrt{n}\rfloor$.

Conjugate Gradient Fluctuations

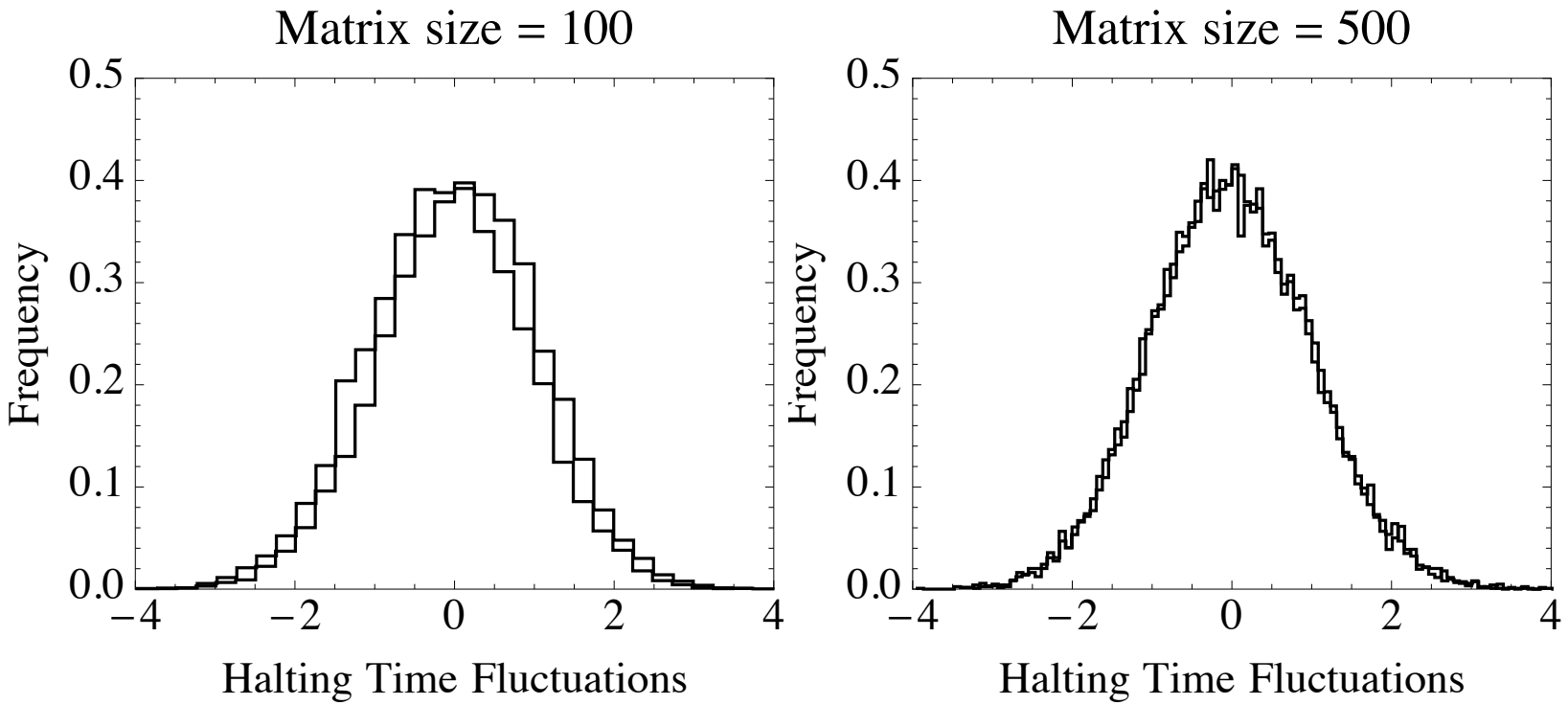


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A = \text{CG}$ and $E = \text{cLOE},\ \text{cPBE}$. The left figure displays two histograms, one for cLOE and one for cPBE, when $n = 100$. The right figure displays the same information for $n = 500$. All histograms are produced with 16,000 samples. Again, we see two-component universality emerges for n sufficiently large: the histograms follow a universal (independent of E) law. With the chosen scaling, we see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

The GMRES Algorithm

The GMRES Algorithm

The authors again consider the solution of $Wx = b$ but here W has the form $I + X$ and $X \equiv X_n$ is a random, real non-symmetric matrix and $b = (b_j)$ is independent with uniform iid entries $\{b_j\}$. As $W = I + X$ is (almost surely) no longer positive definite the conjugate gradient algorithm breaks down, and the authors solve $(I + X)x = b$ using the Generalized Minimal Residual (GMRES) algorithm.

Again, the algorithm is iterative and at iteration k of the algorithm an approximate solution x_k of $(I + X)x = b$ is found and the residual $r_k = (I + X)x_k - b$ is computed. As before, for any given $\epsilon > 0$, the method is halted when $\|r_k\|_2 < \epsilon$ and $k_{\epsilon,n,A,E}(X, b)$ is recorded. For these computations X is chosen from two distinct ensembles. As in the conjugate gradient problem, the authors compute the histograms for the fluctuations of the halting time $\tau_{\epsilon,n,A,E}$ for two ensembles E, where now $A =$ GMRES. The results are given below, where again two component universality is evident.

GMRES Fluctuations

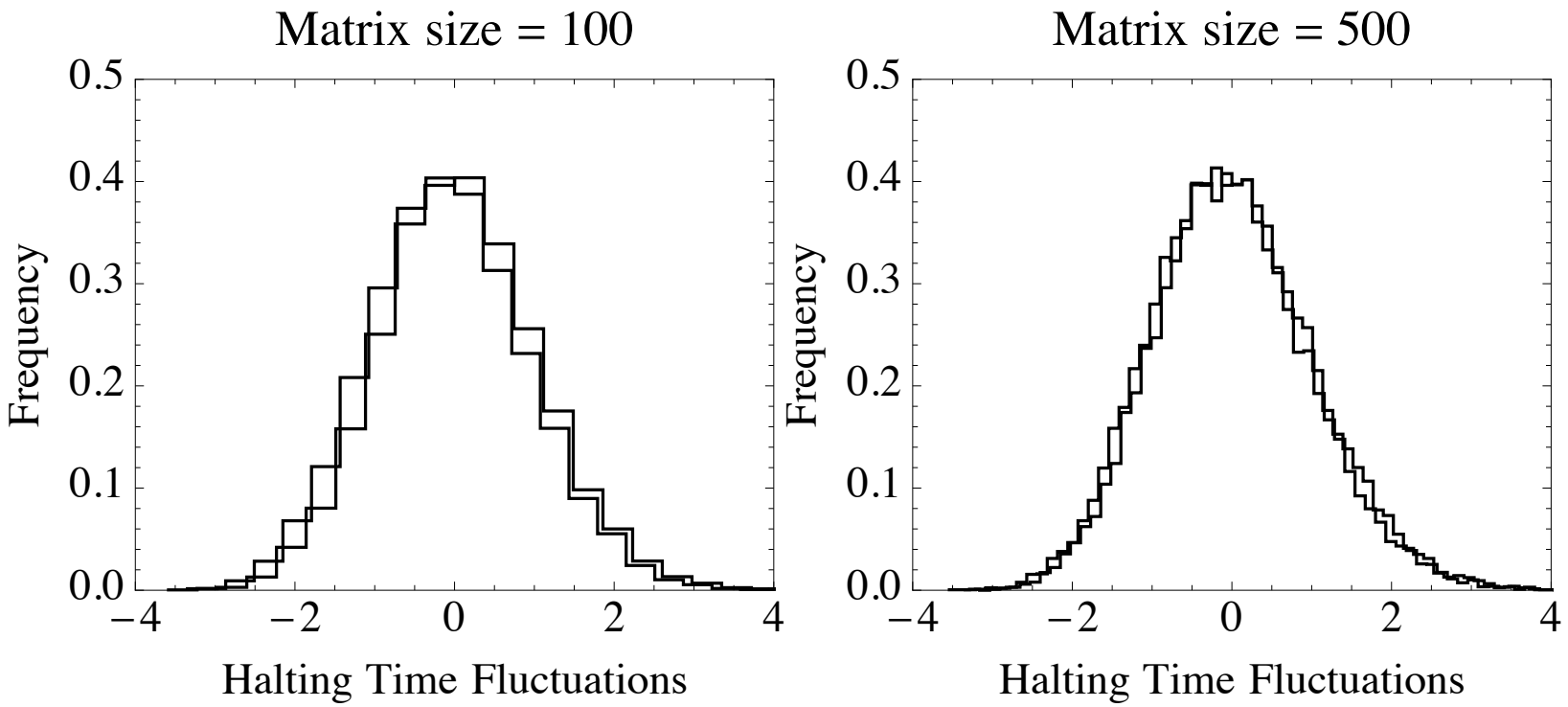


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A =$ GMRES, $E =$ cSGE, cSBE and $\epsilon = 10^{-8}$. The left figure displays two histograms, one for cSGE and one for cSBE, when $n = 100$. The right figure displays the same information for $n = 500$. All histograms are produced with 16,000 samples. We see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

The critically-scaled shifted Bernoulli Ensemble (cSBE) is given by $I + X/\sqrt{n}$ where X is an $n \times n$ matrix consisting of iid Bernoulli variables taking the values ± 1 with equal probability. The critically-scaled shifted Ginibre Ensemble (cSGE) is given by $I + X/\sqrt{n}$ where X is an $n \times n$ matrix of iid Gaussian variables with mean zero and variance one.

The scaling is chosen so that $\mathbb{P}(|\|X/\sqrt{n}\| - 2| > \epsilon)$ tends to zero as $n \to \infty$.

GMRES Fluctuations

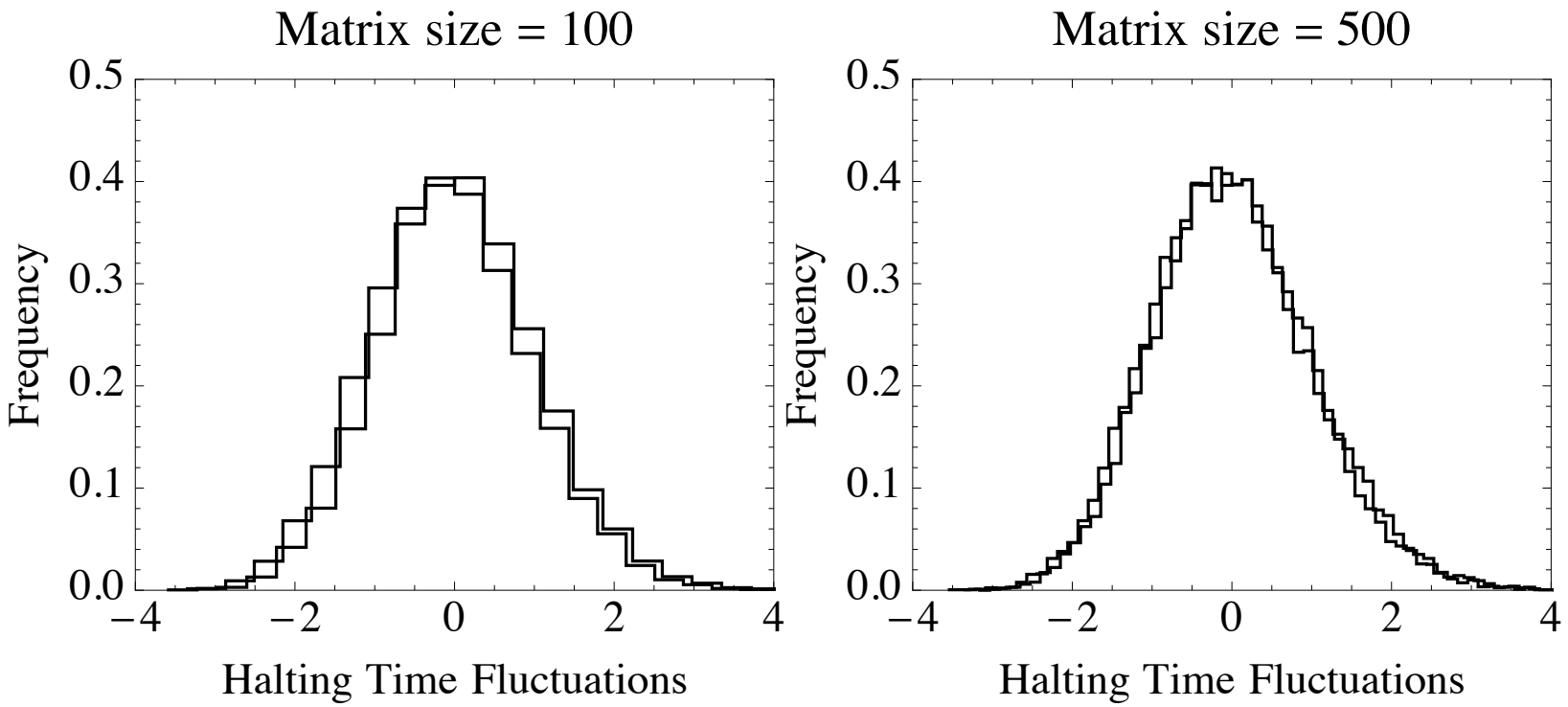


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A =$ GMRES, $E =$ cSGE, cSBE and $\epsilon = 10^{-8}$. The left figure displays two histograms, one for cSGE and one for cSBE, when $n = 100$. The right figure displays the same information for $n = 500$. All histograms are produced with 16,000 samples. We see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

Discretization of a Random PDE

Here the authors raise the issue of whether two-component universality is just a feature of finite-dimensional computation, or is also present in problems which are intrinsically infinite dimensional.

Here the authors raise the issue of whether two-component universality is just a feature of finite-dimensional computation, or is also present in problems which are intrinsically infinite dimensional.

What about PDEs?

Here the authors raise the issue of whether two-component universality is just a feature of finite-dimensional computation, or is also present in problems which are intrinsically infinite dimensional.

What about PDEs?

In particular, is the universality present in numerical computations for PDEs? As a case study, the authors consider the numerical solution of the Dirichlet problem $\Delta u = 0$ in a star-shaped region $\Omega \subset \mathbb{R}^2$ with $u = f$ on $\partial\Omega$. In this case, the boundary is described by a periodic function of the angle θ, $r = r(\theta)$, and similarly $f = f(\theta)$, $0 \leq \theta \leq 2\pi$.

Two ensembles, BDE and UDE (described below), are derived from a discretization of the problem with specific choices for r, defined by a random Fourier series. The boundary condition f is chosen randomly by letting $\{f(\frac{2\pi j}{n})\}_{j=0}^{n-1}$ be iid uniform on $[-1,1]$. Histograms for the halting time $\tau_{\epsilon,n,A,E}$ from these computations are given below and again, two-component universality is evident.

Let Ω be the star-shaped region interior to the curve $(x, y) = (r(\theta)\cos(\theta), r(\theta)\sin(\theta))$ where $r(\theta)$ is given by

$$r(\theta) = 1 + \sum_{j=1}^{m} (X_j \cos(j\theta) + Y_j \sin(j\theta)), \quad 0 \le \theta < 2\pi$$

and X_j and Y_j are iid random variables taking values in $[-1/(2m), 1/(2m)]$. Dividing by $2m$ eliminates the possibility that r vanishes. The double-layer potential formulation of the boundary integral equation

$$\pi u(P) - \int_{\partial\Omega} u(P) \frac{\partial}{\partial n_Q} \log|P - Q| dS_Q = -f(P), \quad P \in \partial\Omega,$$

is solved by discretizing in θ with n points and applying the trapezoidal rule choosing $n = 2m$.

The Bernoulli Dirichlet Ensemble (BDE) is the case where X_m and Y_m are Bernoulli variables taking values $\pm 1/(2m)$ with equal probability. The Uniform Dirichlet Ensemble (UDE) is is the case where X_m and Y_m are uniform variables on $[-1/(2m), 1/(2m)]$.

More GMRES Fluctuations

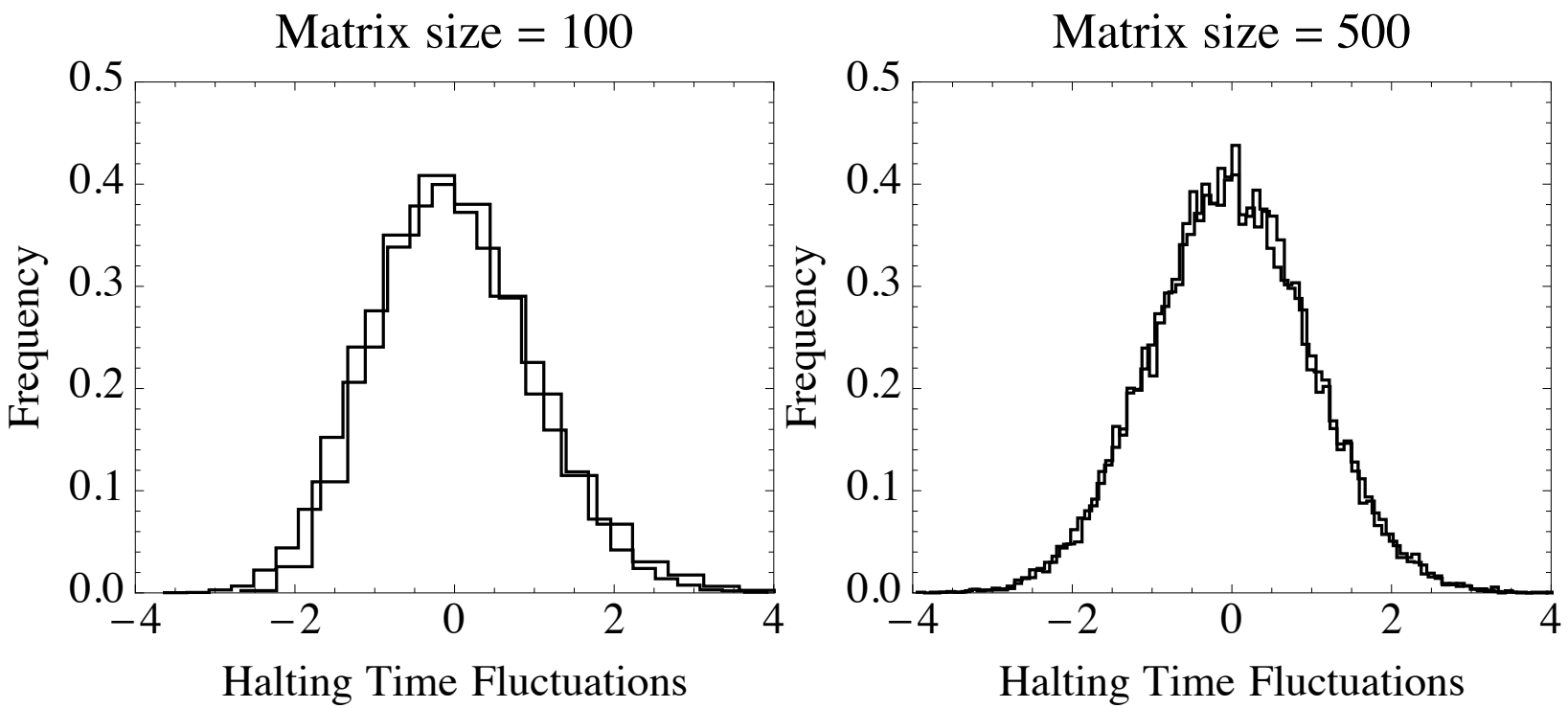


Figure : The observation of two-component universality for $\tau_{\epsilon,n,A,E}$ when $A =$ GMRES, $E =$ UDE, BDE and $\epsilon = 10^{-8}$. The left figure displays two histograms, one for UDE and BDE, when $n = 100$. The right figure displays the same information for $n = 500$. All histograms are produced with 16,000 samples. We see two-component universality emerge for n sufficiently large: the histograms follow a universal (independent of E) law.

The following figure conflates the previous computations from GMRES applied to the shifted ensembles and GMRES applied to the Dirichlet problem given above.

All GMRES Fluctuations

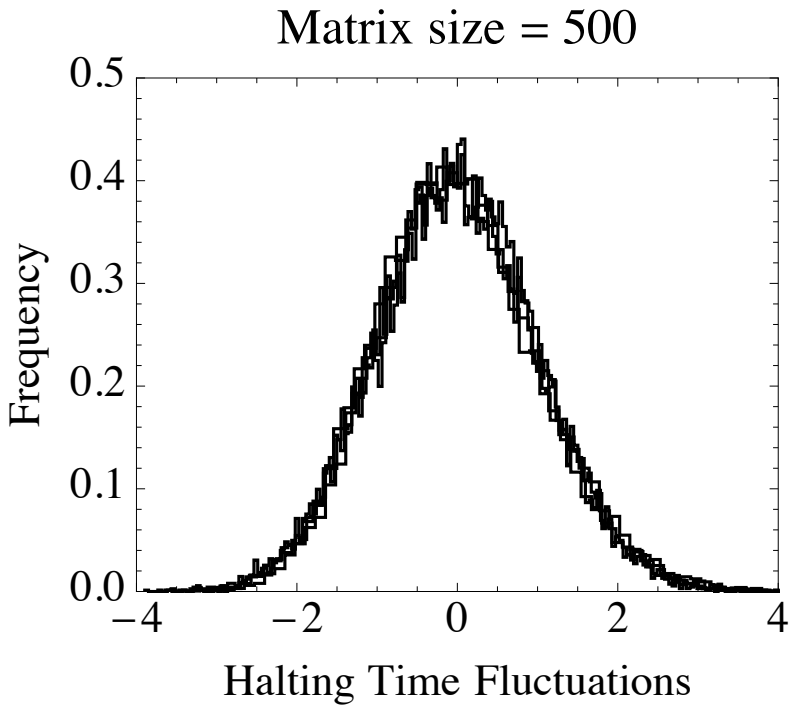
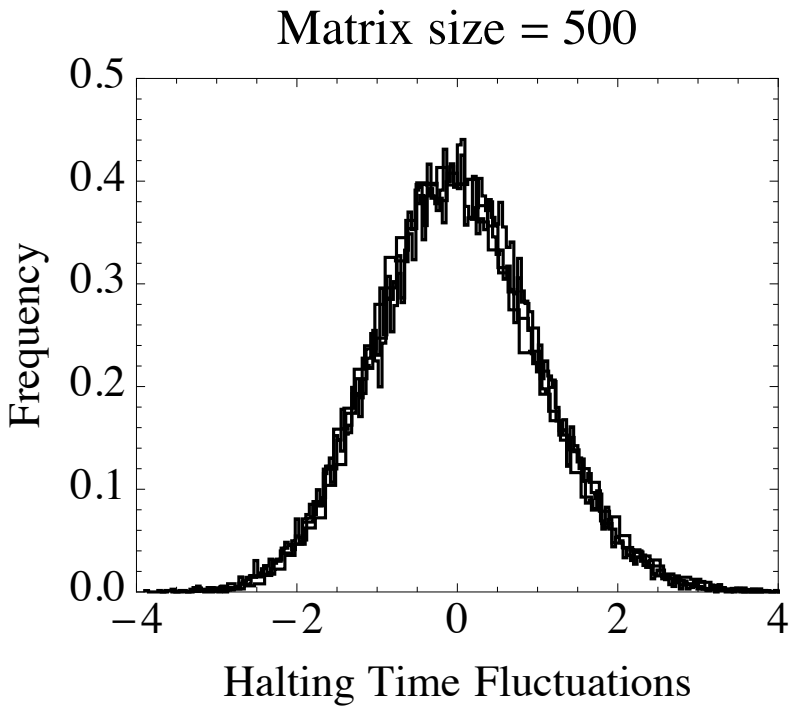
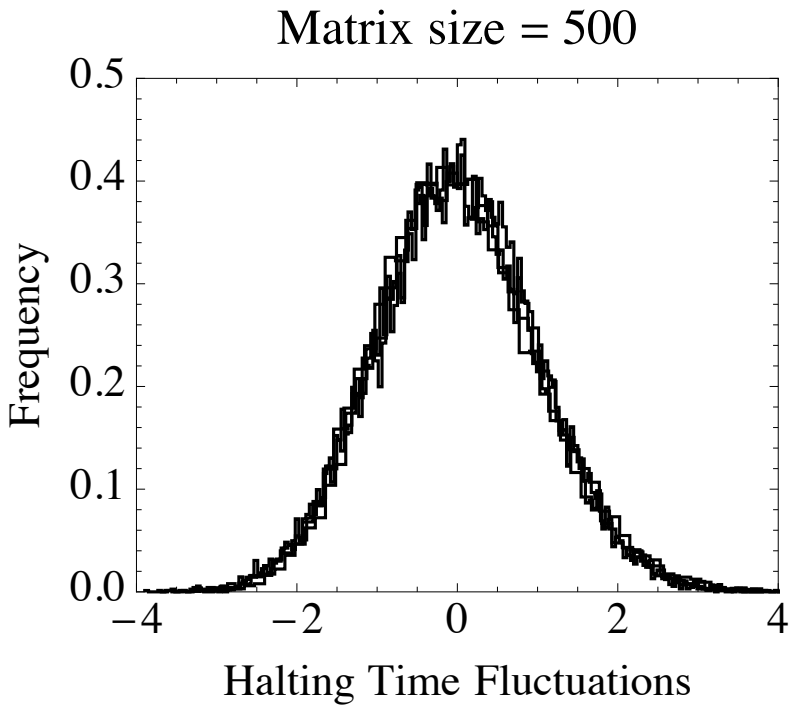


Figure : This figure consists of four histograms, two taken GMRES applied to the previous shifted ensembles and two taken from GMRES applied to the Dirichlet problem.



What is surprising, and quite remarkable, about these computations is that the histograms in the case of the Dirichlet problem are the **same** as the histograms for the shifted ensembles. In other words, UDE and BDE are structured with random components, whereas cSGE and cSBE have no structure, yet they produce the same statistics.



This brings to mind the situation in the 1950s when Wigner introduced random matrices as a model for scattering resonances of neutrons off heavy nuclei: the neutron-nucleus system has a well-defined and structured Hamiltonian, but nevertheless the resonances for neutron scattering are well-described statistically by the eigenvalues of an (unstructured) random matrix.

A Genetic Algorithm

In all the computations discussed so far, the randomness in the computations resides in the initial data. In the sixth set of computations, the authors consider an algorithm which is intrinsically stochastic. In particular, they consider a **genetic algorithm**, which they use to compute **Fekete points**. Such points $P^* = (P_1^*, P_2^*, \ldots, P_N^*) \in \mathbb{R}^N$ are the global minimizers of the objective function

$$H(P) = \frac{2}{N(N-1)} \sum_{1 \le i \ne j \le N} \log |P_i - P_j|^{-1} + \frac{1}{N} \sum_{i=1}^{N} V(P_i)$$

for real-valued functions $V = V(x)$ which grow sufficiently rapidly as $|x| \to \infty$. It is well-known that as $N \to \infty$, the counting measures $\delta_{P^*} = \frac{1}{N} \sum_{i=1}^{N} \delta_{P_i^*}$ converge to the so-called equilibrium measure μ_V which plays a key role in the asymptotic theory of the orthogonal polynomials generated by the measure $e^{-NV(x)}dx$ on $\mathbb{R}$. Genetic algorithms are particularly useful for large scale optimization problems, such as those that occur, for example, in the financial industry, and involve two basic components , "mutation" and "crossover". The authors implement the genetic algorithm in the following way.

Fix a distribution $\mathfrak{D}$ on $\mathbb{R}$. Draw an initial population $\mathcal{P}_0 = \mathcal{P} = \{P_i\}_{i=1}^n$ consisting of $n = 100$ vectors in $\mathbb{R}^N$, N large, with elements that are iid uniform on $[-4, 4]$. The random map $F_{\mathfrak{D}}(\mathcal{P}) : (\mathbb{R}^N)^n \to (\mathbb{R}^N)^n$ is defined by one of the following two procedures:

Mutation

Pick one individual $P \in \mathcal{P}$ at random (uniformly). Then pick two integers $n_1,\ n_2$ from $\{1, 2, \ldots, N\}$ at random (uniformly and independent). Three new individuals are created.

- $\tilde{P}_1$ — draw n_1 iid numbers $\{x_1, \ldots, x_{n_1}\}$ from $\mathfrak{D}$ and perturb the first n_1 elements : $(\tilde{P}_1)_i = (P)_i + x_i$, $i = 1, \ldots, n_1$, and $(\tilde{P}_1)_i = (P)_i$ for $i > n_1$.
- $\tilde{P}_2$ — draw $N - n_2$ iid numbers $\{y_{n_2+1}, \ldots, y_N\}$ from $\mathfrak{D}$ and perturb the last $N - n_2$ elements of P: $(\tilde{P}_2)_i = (P)_i + y_i$, $i = n_2 + 1, \ldots, N$, and $(\tilde{P}_2)_i = (P)_i$ for $i \leq n_2$.
- $\tilde{P}_3$ — draw $|n_1 - n_2|$ iid numbers $\{z_1, \ldots, z_{|n_1 - n_2|}\}$ from $\mathfrak{D}$ and perturb elements $n_1^* = 1 + \min(n_1, n_2)$ through $n_2^* = \max(n_1, n_2)$: $(\tilde{P}_3)_i = (P)_i + z_{i - n_1^* + 1}$, $i = n_1^*, \ldots, n_2^*$, and $(\tilde{P}_3)_i = (P)_i$ for $i \notin \{n_1^*, \ldots, n_2^*\}$.

Crossover

Pick two individuals $P,\ Q$ from $\mathcal{P}$ at random (independent and uniformly). Then pick two numbers $n_1,\ n_2$ from $\{1, 2, \ldots, N\}$ (independent and uniformly). Two new individuals are created.

- $\tilde{P}_4$ — Replace the n_1th element of P with the n_2th element of Q and perturb it (additively) with a sample of $\mathfrak{D}$.
- $\tilde{P}_5$ — Replace the n_1th element of Q with the n_2th element of T and perturb it (additively) with a sample of $\mathfrak{D}$.

At each step, the application of either crossover or mutation is chosen with equal probability.

The new individuals are appended to $\mathcal{P}$ (after mutation we have $\tilde{\mathcal{P}} = \mathcal{P} \cup \{\tilde{P}_1, \tilde{P}_2, \tilde{P}_3\}$ and after crossover we have $\tilde{\mathcal{P}} = \mathcal{P} \cup \{\tilde{P}_4, \tilde{P}_5\}$) and $\mathcal{P} \mapsto \mathcal{P}' = F_{\mathfrak{D}}(\mathcal{P}) \in (\mathbb{R}^N)^n$ is constructed by choosing the 100 P_i's in $\tilde{\mathcal{P}}$ which yield the smallest values of $H(P)$. The algorithm produces a sequence of populations $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_k, \ldots$ in $(\mathbb{R}^N)^n$, $\mathcal{P}_{k+1} = F_{\mathfrak{D}}(\mathcal{P}_k)$, $n = 100$, and halts, with halting time recorded, for a given ϵ, when $\min_{P \in \mathcal{P}_k} H(P) - \inf_{P \in \mathbb{R}^N} H(P) < \epsilon$.

The histograms for the fluctuations $\tau_{\epsilon,N,A,E}$, with $A =$ Genetic are given below, for two choices of V, $V(x) = x^2$ and $V(x) = x^4 - 3x^2$, and different choices of $E \simeq \mathfrak{D}$. Again, two-component universality is evident.

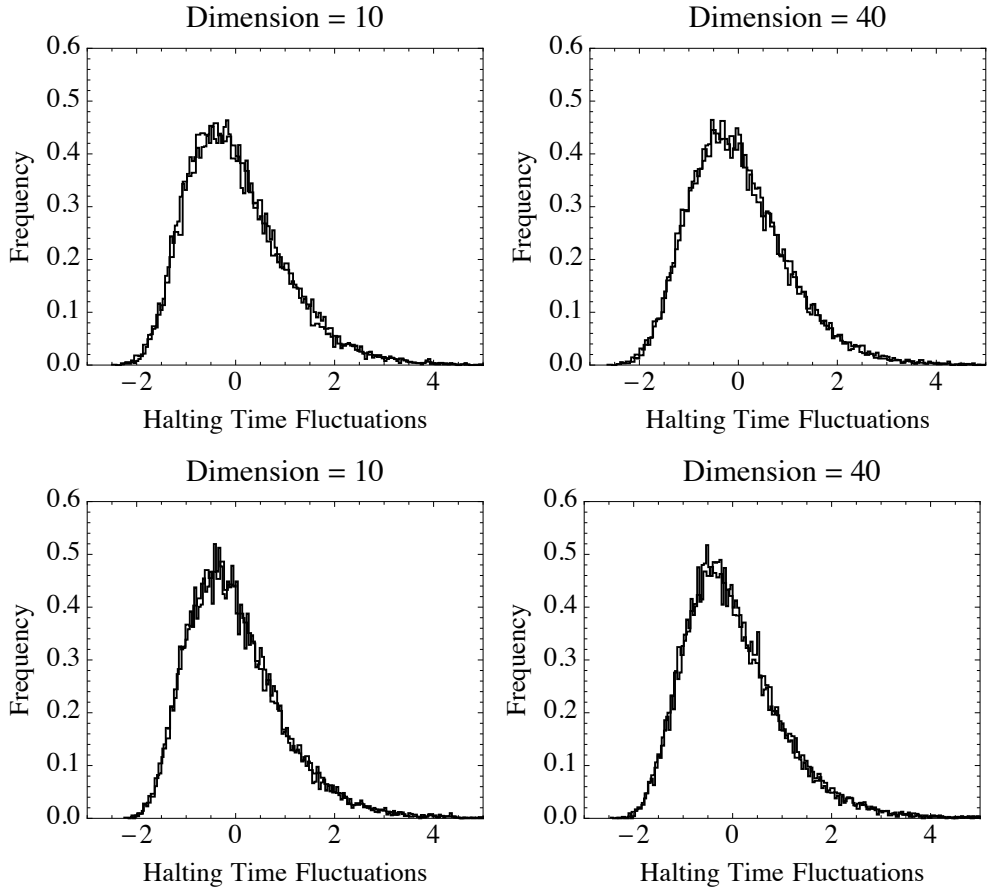


Figure : The observation of two-component universality for $\tau_{\epsilon,N,A,E}$ when $A =$ Genetic, $\epsilon = 10^{-2}$ and $E \simeq \mathfrak{D}$ where $\mathfrak{D}$ is chosen to be either uniform on $[-1/(10N), 1/(10N)]$ or taking values $\pm 1/(10N)$ with equal probability. The top row is created with the choice $V(x) = x^2$ and the bottom row with $V(x) = x^4 - 3x^2$. Each of the plots in the left column displays two histograms, one for each choice of $\mathfrak{D}$ when $N = 10$. The right column displays the same information for $N = 40$. All histograms are produced with 16,000 samples. The equilibrium measure for $V(x) = x^2$ is supported on one interval whereas the equilibrium measure for $V(x) = x^4 - 3x^2$ is supported on two intervals. It is evident that the histograms collapse onto a universal curve, one for each V.

Curie–Weiss Decision Making Model

In the final set of computations, the authors pick up on a common notion in neuroscience that the human brain is a computer with software and hardware. If this is indeed so, then one may speculate that two-component universality should be present certainly in some cognitive actions.

The authors focus on recent work of Bakhtin and Correll (2012), who have conducted and analyzed the data obtained from experiments with 45 human participants. The participants are shown 200 pairs of images. The images in each pair consist of nine black disks of variable size. The disks in the images within each pair have approximately the same area so that there is no *a priori* bias. The participants are then asked to decide which of the two images covers larger (black) area. Bakhtin and Correll then record the time T that it takes for each participant to make a decision. For each participant, the decision times for the 200 pairs are collected and the fluctuation histogram is tabulated. They then compare their experimental results with a dynamical Curie–Weiss model frequently used in describing decision processes, resulting in good agreement.

At its essence the model is Glauber dynamics on the hypercube $\{-1,1\}^N$ with a microscopic approximation of a drift-diffusion process. Consider N variables $\{X_i(t)\}_{i=1}^N$, $X_i(t) \in \{-1,1\}$. The state of the system at time t is $X(t) = (X_1(t), X_2(t), \ldots, X_N(t))$. The transition probabilities are given through the expressions

$$\mathbb{P}(X_i(t+\Delta t) \neq X_i(t) | X(t) = x) = c_i(x)\Delta t + o(\Delta t),$$

where $c_i(x)$ is the spin flip intensity. The observable considered is

$$M(X(t)) = \frac{1}{N}\sum_{i=1}^{N} X_i(t) \in [-1,1],$$

and the initial state of the system is chosen so that $M(X(0)) = 0$, a state with no *a priori* bias, as in the case of the experimental setup.

Given $\epsilon \in (0,1)$, which may not be small, the halting (or decision) time for this model is $k = \inf\{t : |M(X(t))| \geq \epsilon\}$, the time at which the system makes a decision.

Following standard procedures, this model is simulated by first sampling an exponential random variable with mean

$$\left(\sum_i c_i(X(t)) \right)^{-1}$$

to find the time increment Δt at which the system changes state. With probability one, just a single spin flipped.

One determines which spin flips by sampling a random variable Y with distribution

$$\mathbb{P}(Y = i) = \frac{c_i(X(t))}{\sum_i c_i(X(t))}, \quad i = 1, 2, \ldots, N,$$

so producing an integer j. Define

$$\begin{aligned} X_i(t+s) &\equiv X_i(t) \text{ if } s \in [0, \Delta t) \text{ for } i = 1, 2, \ldots, N, \\ X_i(t+\Delta t) &\equiv X_i(t), \text{ if } i \neq j, \\ X_j(t+\Delta t) &\equiv -X_j(t). \end{aligned}$$

This procedure is repeated with t replaced by $t + \Delta t$ to evolve the system.

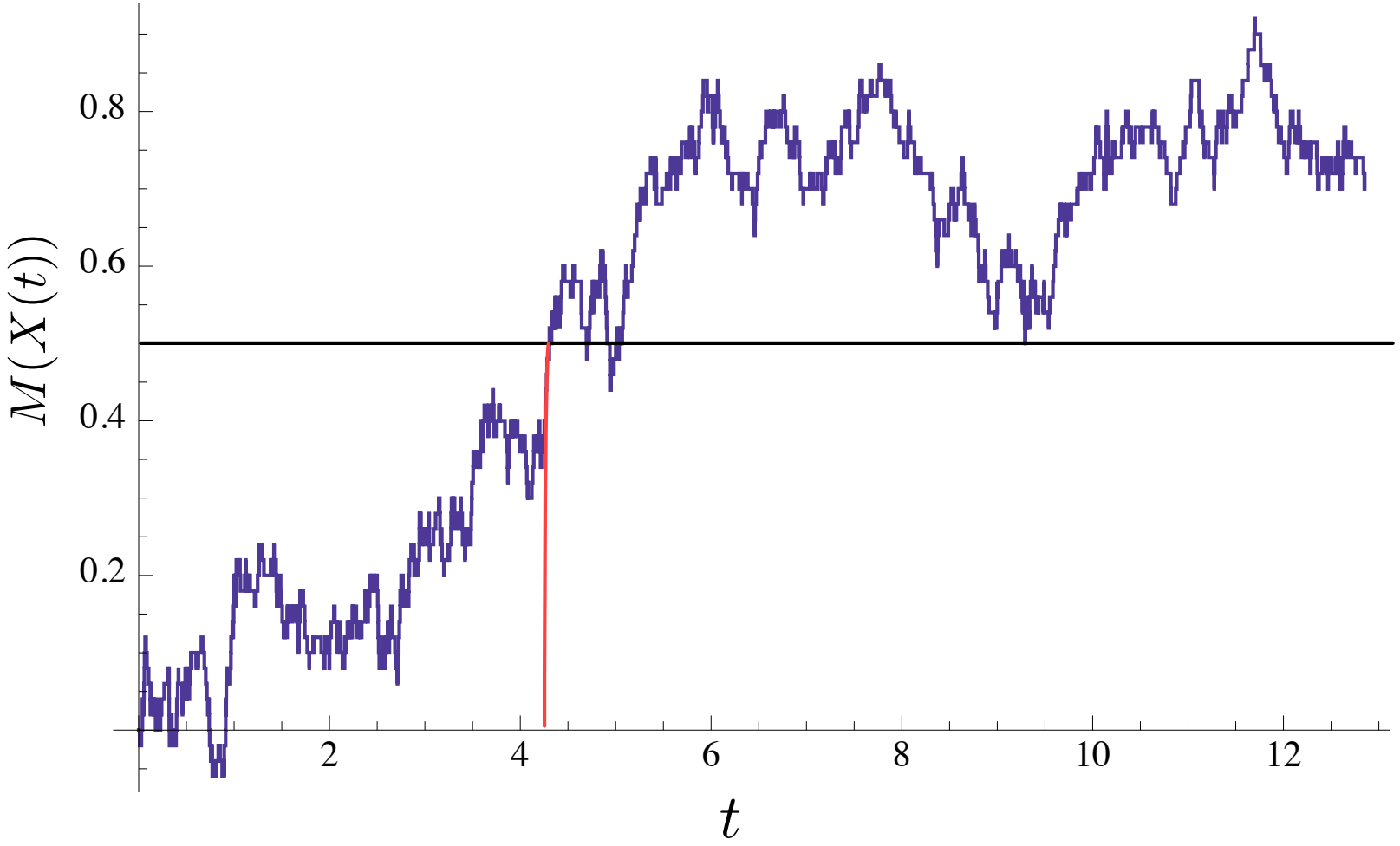
Central to the application of the model is the assumption on the statistics of the spin flip intensity $c_i(x)$. The authors in the present paper raise the following question.

If one changes the basic statistics of the c_i's, will the limiting histograms for the fluctuations of k be affected as N becomes large?

In response to this question the authors consider the following choices for $E \simeq c_i(x)$ ($\beta = 1.3$):

1. $c_i(x) = o_i(x) = e^{-\beta x_i M(x)}$ (the case studied by Bakhtin and Correll (2012)),
2. $c_i(x) = u_i(x) = e^{-\beta x_i (M(x) - M^3(x)/5)}$,
3. $c_i(x) = v_i(x) = e^{-\beta x_i (M(x) + M^8(x))}$.

A generic sample path of the observable $M(X(t))$ with $c_i = o_i$ is displayed below. Here for $\epsilon = 0.5$, $k \approx 4$.



The histograms for the fluctuations $\tau_{\epsilon,N,A,E}$ of k are given below for all three choices of c_i. Once again, two-component universality is evident. Thus these computations demonstrate two-component universality for a range of decision process models.

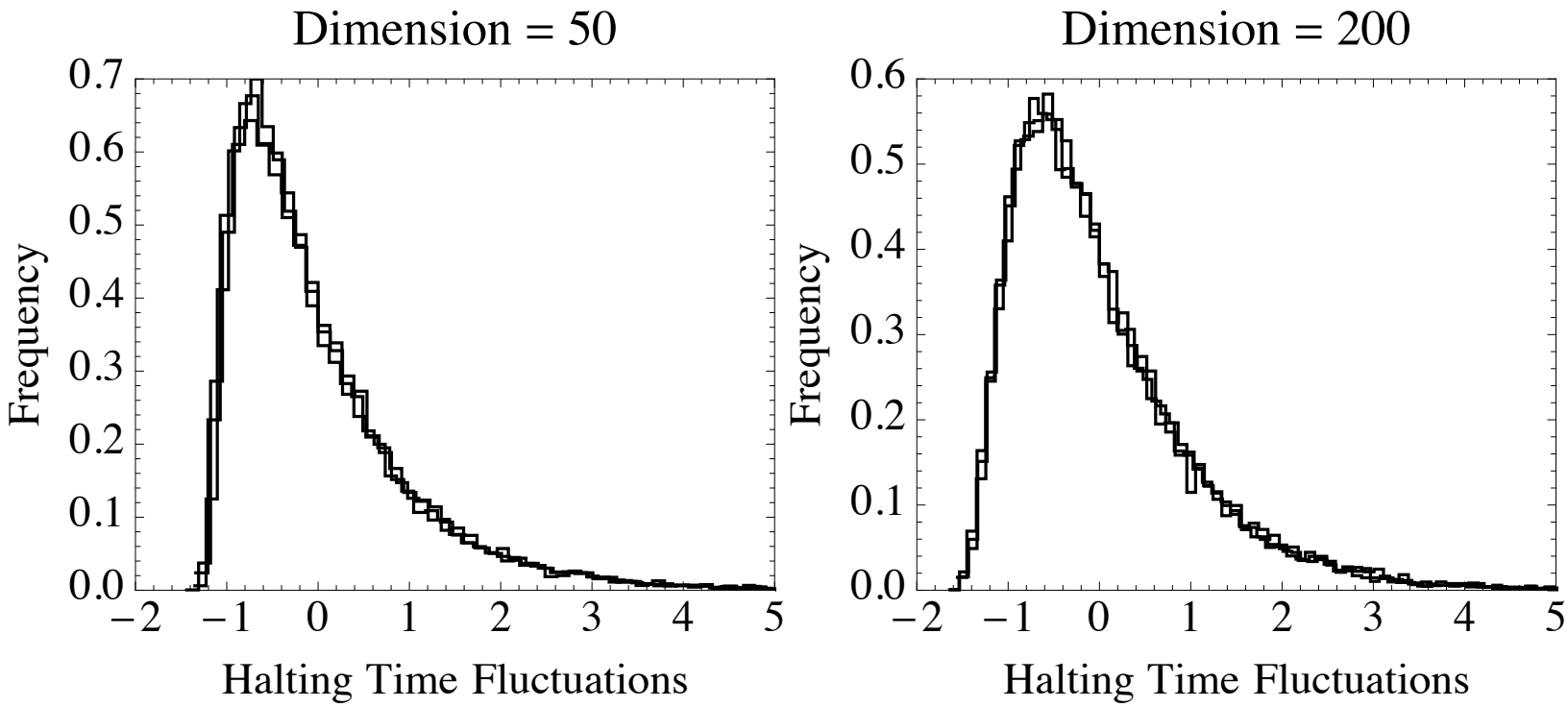


Figure : The observation of two-component universality for $\tau_{\epsilon,N,A,E}$ when $A =$ Curie–Weiss, $E \simeq o_i,\ u_i,\ v_i$, $\epsilon = .5$ and $\beta = 1.3$. The left figure displays three histograms, one for each choice of E when $N = 50$. The right figure displays the same information for $N = 200$. All histograms are produced with 16,000 samples. The histogram for $E = o_i$ corresponds to the case studied by Bakhtin and Correll (2012). It is clear from these computations that the fluctuations collapse on to the universal curve for $E = o_i$. Thus, reasonable changes in the spin flip intensity do not appear to change the limiting histogram. This indicates why the specific choice made in Bakhtin and Correll (2012) of $E = o_i$ is perhaps enough to capture the behavior of many individuals.

Bakhtin, Y, and J Correll. A neural computation model for decision-making times. *J. Math. Psychol.*, 56(5):333–340, 2012.

Olver, S, RR Nadakuditi, and T Trogdon. Sampling unitary invariant ensembles. *arXiv Prepr. arXiv1404.0071*, 2014.

Pfrang, CW, P Deift, and G Menon. How long does it take to compute the eigenvalues of a random symmetric matrix? *arXiv Prepr. arXiv1203.4635*, 2012.