The gap size distribution of parked cars and the Coulomb gas model

Anthony Fader
afader@umich.edu

July 25, 2006

Abstract

We have measured the distribution of distances between four different arrangements of parked cars in Ann Arbor, Michigan. We compare the results with the normalized eigenvalue spacing distribution of the Gaussian Unitary Ensemble and find that it does not agree with any of the arrangements. We then compare the measured results with the one-dimensional Coulomb gas model, using maximum likelihood estimation for the parameter β, and find agreement between the model and the measurements.

1 Introduction

Random matrix theory is the study of asymptotic properties of matrices with random variables as entries. It has been applied to many problems, including problems in pure and applied mathematics and physics [4]. In this paper we will in particular study the Gaussian Unitary Ensemble. The GUE is the set of $N \times N$ random Hermitian matrices with independent Gaussian entries. The normalized spacing between neighboring eigenvalues can be approximated by Wigner's surmise [4]:

$$w(t) \approx \frac{32}{\pi^2} t^2 \exp(-\frac{4}{\pi} t^2) \quad (1)$$

In this paper, we study the gap size between parked cars. The car parking problem is a random sequential absorption problem where fixed-sized objects are randomly allocated into a given space. We want to study the distance between neighboring objects. In [6], the authors tested the accuracy of random car parking models by measuring the gap distances between parked cars on four connected streets in London. They found that the measurements did not agree with the theoretical models and proposed two new models that improved the agreement with the measurements by adding new assumptions to the random sequential absorption problem. However, the empirical measurements still show disagreement with both models, especially for small gaps. Subsequently, [1] compares the empirical data from [6] with Equation (1) and proposes that the gap spacings of parked cars behave according to the eigenvalue spacing of the GUE. The author presents a normalized histogram of the data with the graph of Wigner's surmise and the two graphs agree quite well.

In our paper, we try to reproduce the results of [6] and [1] as well as see whether variations on parking arrangements change the results. In our experiment, we measured the distances between parked cars in four different locations in Ann Arbor, Michigan (population 114,000 as of 2006):

1. Madison Street: parallel parking, no driveways or sidewalk gaps between parking spaces (similar to the London location in [6])

2. State Street (Figure 2): parallel parking with parking meters acting as loose boundaries for drivers

3. Forest Parking Structure (Figure 3): angled parking with thick painted buffers between each pair of parking spaces

4. Fifth Street Parking Lot (Figure 4): non-angled parking with thin painted lines between each pair of parking spaces

Each location is near University of Michigan's central campus. Madison Street is in a residential neighborhood, while the other three locations are in a nonresidential areas.

Firstly, we found that the data from the Forest and Fifth Street parking lots was more centralized and symmetric than the data from Madison and State. Secondly, we found that none of the data sets agreed with Wigner's surmise (1). In random matrix theory, the GUE spacing distribution is a special case of the general Coulomb gas model, which has a parameter β (see Section 2). Using maximum likelihood estimation, we found that $\beta = 2$, which is the value of β for the GUE, is not the most likely value for any of the data sets. For the Madison data set, this disagrees with the results from [1]. The estimated values of β fit their respective data sets very well.

The structure of this paper is as follows. First we present a summary of the relevant aspects of random matrix theory and maximum likelihood estimation. Next, we present the data from each set as a normalized histogram and compare their quantile values with the theoretical values. Lastly, we summarize our findings and give ideas for further investigation. The appendix includes diagrams of the parking arrangements, graphs, and MATLAB code used to perform the calculations in the paper.

Acknowledgements. We would like to thank Professor Jinho Baik for supervising the project, Cy Chan for providing the Beta Estimator website, and Rachel Lee for assisting with the measurements.

2 Random matrices and the Coulomb gas model

The Gaussian Unitary Ensemble is a collection of $N \times N$ Hermitian matrices with a probability density giving each matrix a certain weight. We can think of the Ensemble as a matrix with random variables (satisfying certain conditions) as entries. Each matrix in the collection can be thought of as a realization of this random matrix. We can write the random matrix as $H_N = (H_{jk})$, where H_{jk} is a random variable. Since we are allowing the entries to be complex numbers, we will write H_{jk} as $H_{jk}^R + iH_{jk}^I$, where H_{jk}^R and H_{jk}^I are random variables taking real values.

Because H_N is Hermitian, H_{jj} must be real and $H_{kj} = \overline{H_{jk}} = H_{jk}^R - iH_{jk}^I$. This implies that H_N is completely specified by N^2 random variables: N along the diagonal and $N(N-1)$ above the diagonal.

Definition 1. The *Gaussian Unitary Ensemble* (or *GUE*) is defined by the $N \times N$ matrix $H_N = (H_{jk})$, where each of its N^2 random variables are independent Gaussian with zero means and variances given by $\mathrm{Var}(H_{jj}) = 1/2$ and $\mathrm{Var}(H_{jk}^R) = \mathrm{Var}(H_{jk}^I) = 1/4$ for $j \neq k$.

Since H_N is defined by N^2 random variables, we can think of its probability density function p_N as the joint probability density of its N^2 component random variables. Since these random variables are independent, p_N is the product of these densities. For an $N \times N$ Hermitian matrix $H = (h_{ij})$:

$$\begin{aligned} p_N(H) &= \prod_{j=1}^{N} p_{jj}(h_{jj}) \cdot \prod_{j<k} p_{jk}(h_{jk}^R) p_{jk}(h_{jk}^I) \\ &= C \prod_{j=1}^{N} \exp(-h_{jj}^2) \cdot \prod_{j<k} \exp(-2[(h_{jk}^R)^2 + (h_{jk}^I)^2]) \\ &= C \exp(-\sum_{j=1}^{N} h_{ii}^2 - 2\sum_{j<k} |h_{jk}|^2) \end{aligned}$$

where C is a constant term that depends on N. The exponential term in this expression is equal to the trace of H^2, so we get:

$$p_N(H) = C \exp(-\operatorname{Tr} H^2)$$

$N \times N$ Hermitian matrices will always have N real eigenvalues and therefore it makes sense to treat the eigenvalues of the GUE matrix H_N as random variables. Let H be a realization of H_N and let $A_N = \frac{1}{\sqrt{N}} H$. Let $\lambda_1 \leq \cdots \leq \lambda_N$ be the eigenvalues of A_N. Define

$$G_N(\lambda | A_N) = \frac{\#\{j \,:\, \lambda_j \leq \lambda\}}{N}$$

$G_N(\lambda | A_N)$ is a random function that depends on the realization of A_N. Figure 1 is the normalized histogram of an instance of A_N for $N = 1000$. As can be seen from the graph, the eigenvalues seem to converge around the origin. The next theorem states this result.

Theorem 1. (Wigner's Semicircle Law).

$$\lim_{N\to\infty} G_N(\lambda|A_N) = G(\lambda)$$

where $G(\lambda)$ is a non-random function with density

$$G'(\lambda) = g(\lambda) = \frac{1}{\pi}\sqrt{2-\lambda^2}$$

Theorem 2. *The joint probability density function f for the N eigenvalues of H_N is given by*

$$f(\lambda_1,\ldots,\lambda_N) = C' \cdot \prod_{j<i}(\lambda_i-\lambda_j)^2 \exp(-\operatorname{Tr} H_N^2) \quad (2)$$

$$= C' \cdot \prod_{j<i}(\lambda_i-\lambda_j)^2 \exp(-\sum_k \lambda_k^2) \quad (3)$$

for some constant C.

We present a proof of Theorem 2 for $N = 2$.

Proof. We can write any realization H of the random matrix H_2 as

$$H = \begin{pmatrix} x & z+iw \\ z-iw & y \end{pmatrix}$$

Since H is Hermitian, it can be expressed as $H = U\Lambda U^*$, where U is a 2×2 unitary matrix and Λ is the 2×2 diagonal matrix with eigenvalue components $\lambda_1 \le \lambda_2$. We want to think of this as a map $H \mapsto (\Lambda, U)$, but U is not uniquely determined by H. If we make the top row of U have positive real components, then Λ and U are unique. Using the fact that U is unitary, it can be expressed as

$$U = \begin{pmatrix} a & \sqrt{1-a^2} \\ \sqrt{1-a^2}e^{i\theta} & -ae^{i\theta} \end{pmatrix}$$

with $a \in [0,1]$ and $\theta \in [0,2\pi)$. Using the fact that $H = U\Lambda U^*$, we can express (x,y,z,w) in terms of $(\lambda_1,\lambda_2,a,\theta)$. Letting $q(\lambda_1,\lambda_2,a,\theta)$ be the probability density function of (Λ, U), we get

$$f(\lambda_1,\lambda_2) = \int_0^{2\pi}\int_0^1 q(\lambda_1,\lambda_2,a,\theta)\,da\,d\theta \quad (4)$$

$$= \int_0^{2\pi}\int_0^1 p(x,y,z,w)|J|\,da\,d\theta \quad (5)$$

where x, y, z, and w are functions of $(\lambda_1,\lambda_2,a,\theta)$, p is the probability density function of H, and J is the Jacobian given by

$$J = \frac{\partial(x,y,z,w)}{\partial(\lambda_1,\lambda_2,a,\theta)} = a(\lambda_1-\lambda_2)^2$$

Using the fact that $p(x, y, z, w) = C \exp(-\operatorname{Tr} H^2) = C \exp(-(\lambda_1^2 + \lambda_2^2))$ and by plugging this into (4), we get

$$f(\lambda_1, \lambda_2) = C'(\lambda_1 - \lambda_2)^2 \exp(-(\lambda_1^2 + \lambda_2^2))$$

for some constant C'. □

Next we consider the normalized eigenvalue spacing distribution. Let A_N be as before. By Wigner's Semicircle law, the density of eigenvalues of A_N at a point $x \in (-\sqrt{2}, \sqrt{2})$ approaches $g(x) = \frac{1}{\pi}\sqrt{2 - x^2}$ as $N \to \infty$.

Let $(x - \rho_N, x + \rho_n)$ be a neighborhood of x such that as $N \to \infty$, $\rho_N \to 0$ and $N\rho_N \to \infty$. For small enough values of ρ_N, there are approximately $2\rho_N N \frac{1}{\pi}\sqrt{2 - x^2}$ eigenvalues in the interval $(x - \rho_N, x + \rho_N)$.

Our goal is to find the limiting density of the normalized spacing between consecutive eigenvalues in the interval $(x - \rho_N, x + \rho_N)$. The normalized value of some eigenvalue $\lambda_j \in (x - \rho_N, x + \rho_N)$ is given by

$$\hat{\lambda}_j = \frac{\lambda_j}{\text{mean spacing in } (x - \rho_N, x + \rho_N)} \tag{6}$$

Using Wigner's Semicircle Law,

$$\text{mean spacing} = \frac{2\rho_N}{2\rho_N N \frac{1}{\pi}\sqrt{2 - x^2}}$$

So $\hat{\lambda}_j = \lambda_j \frac{1}{\pi}\sqrt{2 - x^2}$.

The distribution of normalized eigenvalue spacings in $(x - \rho_N, x + \rho_N)$ is given by

$$W_N(t) = \frac{\#\{j \ : \ 0 < \hat{\lambda}_{j+1} - \hat{\lambda}_j < t, \lambda_j \in (x - \rho_N, x + \rho_N)\}}{\#\{j \ : \ \lambda_j \in (x - \rho_N, x + \rho_N)\}}$$

$W_N(t)$ is a random function that depends on the realization of A_N.

Theorem 3. $\lim_{N\to\infty} W_N(t) = W(t)$*, where* $W(t)$ *is a non-random function. The density function* $W'(t) = w(t)$ *is approximated by*

$$w(t) \approx \frac{32}{\pi^2} t^2 \exp(-\frac{4}{\pi} t^2) \tag{7}$$

The formula (7) *is called* Wigner's surmise.

Wigner's surmise is the exact solution to the GUE 2×2 case of eigenvalue spacing and can be found by applying a change of variables to the equation in Theorem 2. The expected value of the eigenvalue spacings is one because of the normalization from Equation (6). The exact formula for the density function in Theorem 3 is known, but Wigner's surmise is much simpler and is a very good estimate [1].

There are other matrix ensembles in addition to the GUE, namely the Gaussian Orthogonal Ensemble and the Gaussian Symplectic Ensemble [4]. In general, the joint probability density functions of the GOE, GUE, and GSE are given by

$$f(\lambda_1,\ldots,\lambda_N) = C \cdot \prod_{j<i} |\lambda_i - \lambda_j|^\beta \exp\left(-\frac{1}{2}\beta \sum_k \lambda_k^2\right) \tag{8}$$

for $\beta = 1, \beta = 2$, and $\beta = 4$ respectively.

Equation (8) is also the probability density function for the one-dimensional position of N point charges according to Dyson's Coulomb gas model, where $\beta = 1/kT$, T is temperature, and k is the Boltzmann constant (relating temperature to energy)[1]. Because of the physical interpretation of Equation (8), it is worthwhile to consider the normalized spacing distribution for arbitrary $\beta > 0$, as was done for the case when $\beta = 2$ for the GUE. The exact expression for this is unknown, so we will present the probability density function $w_\beta(t)$ for the case when $N = 2$ as a surmise.

Letting $u = \lambda_1 - \lambda_2$ and $v = \lambda_2$ for $\lambda_1 \geq \lambda_2$ in Equation (8) and integrating out v, we find that the density function of u is

$$C\sqrt{\frac{\pi}{\beta}} u^\beta e^{-\frac{\beta}{4}u^2}$$

for $u > 0$. We would like to scale $u = at$ so that the density function of t has mean 1. Using this linear change of variables, the density of t is therefore $w_\beta(t) = kt^\beta e^{-ct^2}$ for constants c and k. Because we want $\int w_\beta(t)\,dt = 1$ and $\int t\, w_\beta(t)\,dt = 1$, we can solve for c and k to get

$$c = \left(\frac{\Gamma(\frac{\beta+2}{2})}{\Gamma(\frac{\beta+1}{2})}\right)^2$$

$$k = 2\,\frac{\Gamma(\frac{\beta+2}{2})^{\beta+1}}{\Gamma(\frac{\beta+1}{2})^{\beta+2}}$$

3 Maximum Likelihood Estimation

Given a normalized data set from the experiment, we would like to find a value β that will fit $w_\beta(t)$ with the normalized histogram. The maximum likelihood estimator of the data points $x_1,\ldots,x_N$ is the most likely value for β assuming that $x_1,\ldots,x_N$ are distributed according to $w_\beta(t)$ for some β. Thus, the method does not determine whether $x_1,\ldots,x_N$ are samples from $w_\beta(t)$, it only gives the most likely value of β under the hypothesis that they were indeed sampled from $w_\beta(t)$.

Definition 2. Let $f(x|\theta)$ be a probability density function with a parameter θ. The *likelihood* of θ given points $x_1,\ldots,x_N$ is

$$\mathrm{Lik}(\theta|x_1,\ldots,x_N) = \prod_{i=1}^{N} f(x_i|\theta)$$

The *maximum likelihood* of θ given points $x_1, \ldots, x_N$ is a value θ_M such that $\text{Lik}(\theta_M|x_1, \ldots, x_N) \geq \text{Lik}(\theta|x_1, \ldots, x_N)$ for all possible values of θ. The *log likelihood* of θ given $x_1, \ldots, x_N$ is

$$\Lambda(\theta|x_1, \ldots, x_N) = \log(\text{Lik}(\theta|x_1, \ldots, x_N))$$

Because log is a strictly increasing function, the maximum likelihood is the same as the maximum log-likelihood.

The maximum likelihood estimator of β given $x_1, \ldots, x_N$ is a solution of

$$\frac{\partial}{\partial\beta} \sum_{i=1}^{N} log(w_\beta(x_i)) = 0 \tag{9}$$

[2] is an online β estimator that uses maximum likelihood estimation from sampling. We provide our own β estimator (Figure 13) that uses maximum likelihood estimation from the surmise function $w_\beta(t)$. Both of these methods are estimates of the real spacing distribution for arbitrary β, whose exact formula is not known. Table 1 compares the two estimators. The random values were generated using the surmise function (Figure 12) and each row shows the estimated values of computed by each estimator using 5000 samples. Because the samples were generated using the surmise function, comparing the values in Table 1 will show us the agreement between the two estimation methods, but will not necessarily tell us anything about their accuracy. The two methods do not agree for $\beta \leq 1$, but agree very well for $\beta > 1$.

β	β-surmise	β-sampling [2]
0.25	0.2491	0.4
0.5	0.4944	0.6
1	0.9835	1.1
2	2.0074	2.0
5	5.1313	5.1
8	8.1961	8.2
10	9.9985	10.0

Table 1: Comparison between the two different β-estimators. β is the value that was used to generate the samples (using the code in Figure 12), β-surmise is the maximum likelihood estimator used in this paper, and β-sampling is the maximum likelihood estimator from [2].

4 Empirical results

We measured the spacings between parked cars in four different locations with various parking arrangements. The locations and their arrangements are:

1. Madison Street: parallel parking, no driveways or sidewalk gaps between parking spaces (similar to the London location in [6]).

2. State Street (see Figure 2): parallel parking with parking meters acting as loose boundaries for drivers. The distance between each parking meter is 420 inches and each parking meter designates two parking spaces.

3. Forest Parking Structure (see Figure 3): angled parking with thick, painted buffers between each pair of parking spaces. Each parking space is 80 inches wide and each buffer is 22 inches wide.

4. Fifth Street Parking Lot (see Figure 4): non-angled parking with thin painted lines between each pair of parking spaces. Each parking space is 96 inches wide and each line is 4 inches wide.

The Madison Street parking is similar to the type of parking measured in [6] in London, while the State Street data differs from the London parking due to the parking meters along the street. The Forest Street and Fifth Street parking arrangements differ greatly from the London arrangement.

Set	# Points	# Days Measured	μ (inches)	σ (inches)
Madison	127	11	49	28
State	293	10	57	28
Forest	509	8	32	7
Fifth	321	-	34	8

Table 2: Summary of measurements.

Table 4 gives the total number of points, the number of measurement days, the mean, and the standard deviation for each data set. For the Madison, State, and Forest locations, we kept track of each day's data separately. When normalizing these data sets to have unit mean, we normalized each day individually. For the Fifth street data set, the measurements were not separated by day, so the entire data set was normalized without being subdivided.

Figures 5 and 6 show histograms for the Madison and State data after being normalized to unit mean and scaling the histogram to have unit area. For the These histograms have long right-hand tails and look very similar to the London data histogram. Figures 7 and 9 show the histograms for the Forest and Fifth Street sets. They are more symmetric and centralized than the Madison, State, and London histograms and appear normal. Figures 8 and 10 compare histograms for Forest and Fifth (after being shifted and scaled to have zero mean and unit variance) with the standard normal distribution.

The curves superimposed on the histograms for Figures 5, 6, 7, and 9 are $w_\beta(t)$ for various values of β. The values of β were calculated for each data set using maximum likelihood estimation of the surmise function and the β-estimator from [2] (see Section 3 for more information).

Tables 4-7 compare the experimental quantiles to the quantile values of $w_\beta(t)$ with β chosen by maximum likelihood estimation. The Madison values (Table

4) tend to be larger than the theoretical values near 1.0, but the other data sets seem to match $w_\beta(t)$ quite well. Tables 8 and 9 compare the Forest and Fifth data (after being shifted and scaled to have zero mean and unit variance) to the standard normal distribution, but the agreement is not as good.

5 Discussion

We measured cars parked in four different arrangements: parallel street parking (Madison), parallel street parking with meters (State), angled lot parking with painted lines (Forest), and non-angled lot parking with painted lines (Fifth). Contrary to [1], we found that Wigner's surmise distribution $w_\beta(t)$ for $\beta = 2$ agrees with neither the Madison data, which was closest to the London parking arrangement in [6], nor the State data, which was similar. The Fifth and Forest data differed drastically from the Madison, State, and London data. Table 3 summarizes the values of β estimated for each data set using the methods from this paper and from [2]. Although the two methods do not agree with each other on every data set, if we arrange the estimates from smallest to largest for each method, the orders of the data sets agree: Madison $<$ State $<$ Fifth $<$ Forest.

Data set	β-surmise	β-sampling [2]
Madison	0.5515	0.7
State	1.2645	1.2
Fifth	8.7808	9.3
Forest	9.8826	9.8

Table 3: The estimated β-values for the spacing data using the method described in this paper and the method from [2].

A possible explanation for the large values of β for Forest and Fifth is that the painted boundary lines force the cars into rigid arrangements. In Dyson's Coulomb gas model, temperature T is inversely proportional to β, so decreasing T, which locks the particles in place, corresponds to an increase of the value of β. Thus if parked cars act as charged gas particles according to the Coulomb gas model, it makes sense that locking the cars into fixed positions would result in a higher measured value for β.

We found that parallel street parking without painted boundaries has a lower estimated value of β than the lot parking with painted boundaries. Within the the class of parallel street parking, we found that the arrangement without meters has a lower estimated value of β than the arrangement with meters. Within the class of lot parking, we found that the arrangement with thin lines has a lower estimated value of β than the arrangement with thick buffers. All of these observations support the idea that parked cars are arranged according to the Coulomb gas model, where increasing the parking space boundaries corresponds to decreasing the temperature of the gas. It's not clear why the Madison data differed from the London data. It might be worthwhile to investigate how different environmental parameters, e.g. painted spaces, parallel vs. angled spaces, or location affect the estimated value for β.

6 Appendix

6.1 Figures, Graphs, and Tables

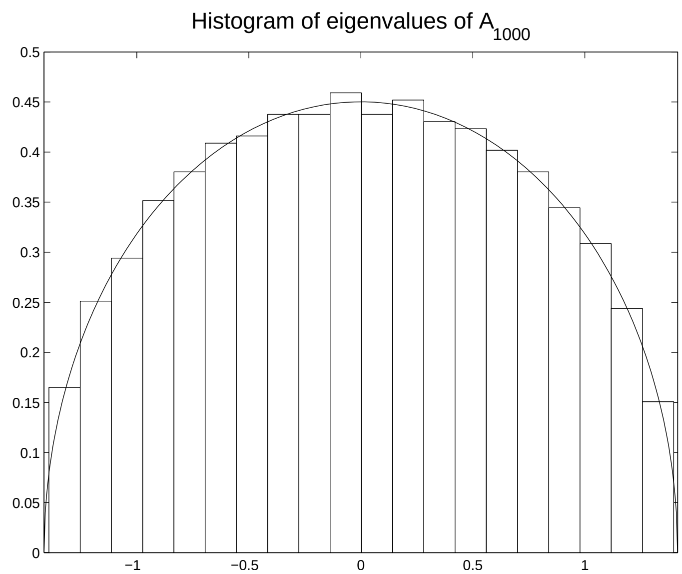

Figure 1: The normalized histogram of the eigenvalues of $A_N = \frac{1}{\sqrt{N}} H_N$ for $N = 1000$ with the curve $y = \frac{1}{\pi}\sqrt{2 - x^2}$.

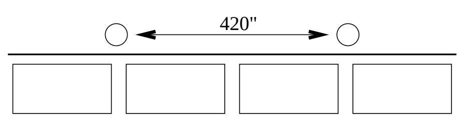

Figure 2: State parking arrangement. The distance between each parking meter is 420 inches and each parking meter designates two parking spaces.

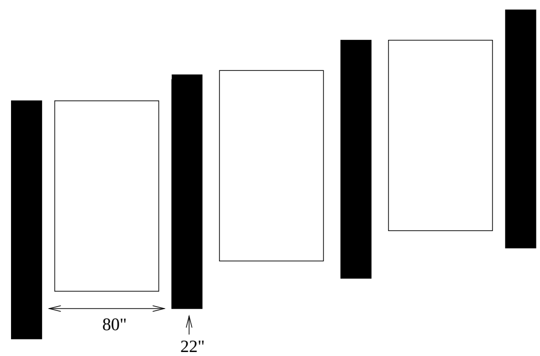

Figure 3: Forest street parking arrangement. The distance between each painted buffer is 80 inches and each painted buffer is 22 inches wide.

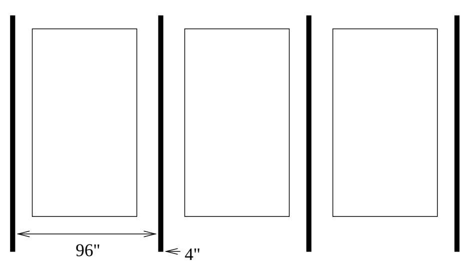

Figure 4: Fifth street parking arrangement. The distance between each line is 96 inches and the lines are 4 inches wide.

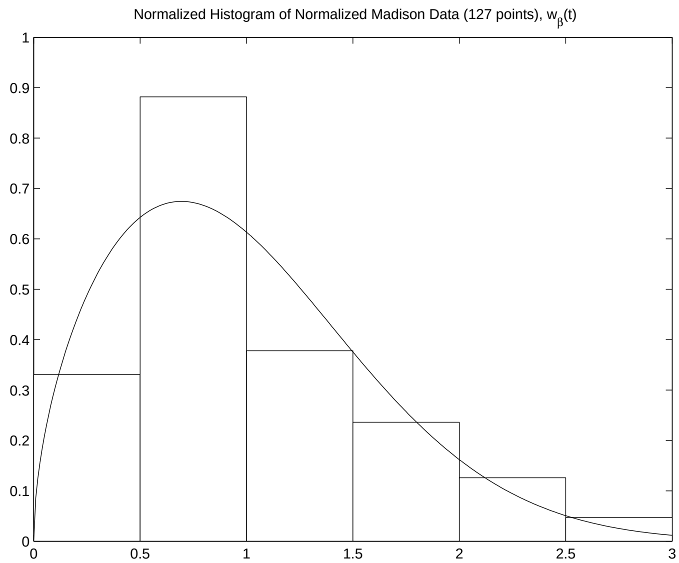

Figure 5: Madison Street data and $w_\beta(t)$ with $\beta = 0.5515$. The data has been normalized to have a unit mean and the histogram has been scaled to have unit area.

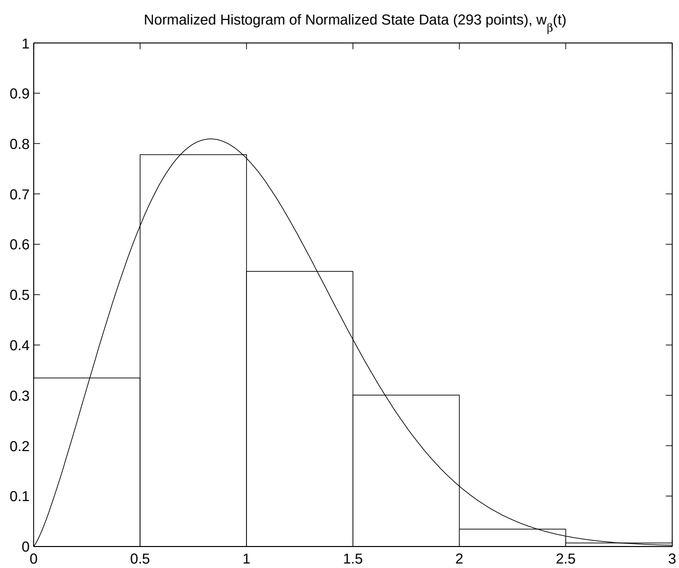

Figure 6: State Street data and $w_\beta(t)$ with $\beta = 1.2645$. The data has been normalized to have a unit mean and the histogram has been scaled to have unit area.

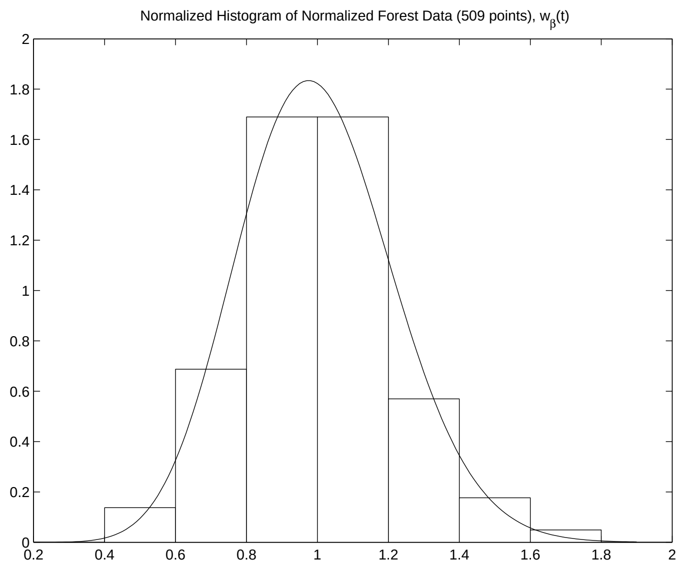

Figure 7: Forest Street data and $w_\beta(t)$ with $\beta = 9.8826$. The data has been normalized to have a unit mean and the histogram has been scaled to have unit area.

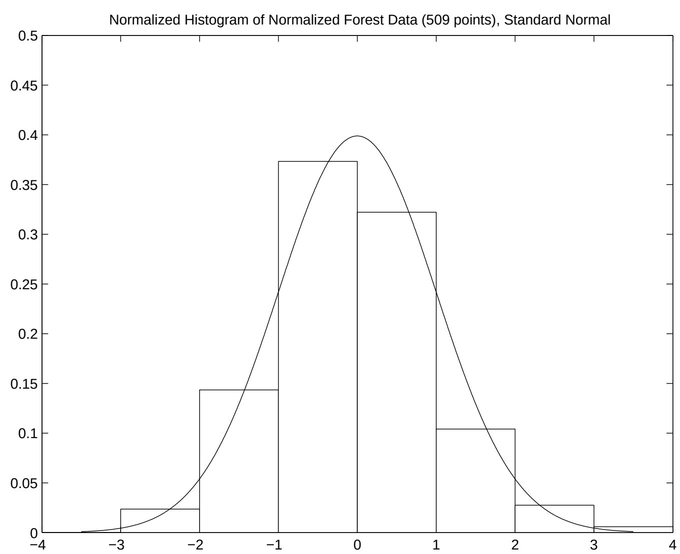

Figure 8: Forest Street data the standard normal density function. The data has been normalized to have zero mean, unit variance, and the histogram has been scaled to have unit area.

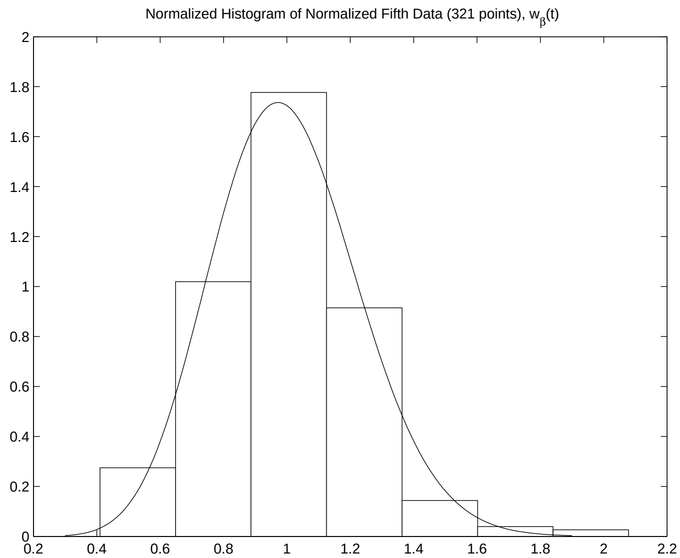

Figure 9: Fifth Street data and $w_\beta(t)$ with $\beta = 8.7870$. The data has been normalized to have a unit mean and the histogram has been scaled to have unit area.

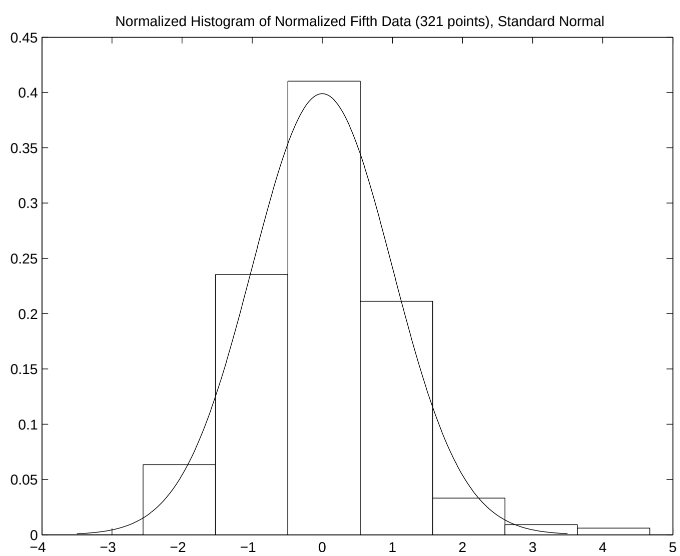

Figure 10: Fifth Street data the standard normal density function. The data has been normalized to have zero mean, unit variance, and the histogram has been scaled to have unit area.

Quantile	Experimental	Surmise	Difference
0.01	0.0843	0.0645	0.0199
0.05	0.1557	0.1816	-0.0260
0.10	0.2456	0.2900	-0.0444
0.30	0.6360	0.6152	0.0208
0.50	0.8372	0.9141	-0.0769
0.70	1.1928	1.2627	-0.0699
0.90	1.8947	1.8281	0.0665
0.95	2.3132	2.1094	0.2039
0.99	2.6025	2.6719	-0.0693

Table 4: The quantile values of the Madison data (after normalizing to unit mean) with the quantile values of $w_\beta(t)$ for $\beta = 0.5515$.

Quantile	Experimental	Surmise	Difference
0.0100	0.1465	0.1406	0.0058
0.0500	0.3111	0.2930	0.0182
0.1000	0.3908	0.4043	-0.0135
0.3000	0.7030	0.7002	0.0028
0.5000	0.9256	0.9492	-0.0236
0.7000	1.2200	1.2246	-0.0047
0.9000	1.7570	1.6641	0.0929
0.9500	1.8650	1.8867	-0.0217
0.9900	2.2365	2.2969	-0.0604

Table 5: The quantile values of the State data (after normalizing to unit mean) with the quantile values of $w_\beta(t)$ for $\beta = 1.2645$.

Quantile	Experimental	Surmise	Difference
0.0100	0.4828	0.5391	-0.0563
0.0500	0.6786	0.6562	0.0224
0.1000	0.7268	0.7266	0.0002
0.3000	0.8848	0.8796	0.0051
0.5000	0.9918	0.9917	0.0001
0.7000	1.0946	1.1089	-0.0143
0.9000	1.2603	1.2832	-0.0229
0.9500	1.3877	1.3711	0.0166
0.9900	1.5886	1.5352	0.0534

Table 6: The quantile values of the Forest data (after normalizing to unit mean) with the quantile values of $w_\beta(t)$ for $\beta = 9.8826$.

Quantile	Experimental	Surmise	Difference
0.01	0.4977	0.5156	-0.0180
0.05	0.6440	0.6387	0.0054
0.10	0.7202	0.7119	0.0083
0.30	0.8782	0.8730	0.0052
0.50	0.9953	0.9902	0.0051
0.70	1.1124	1.1140	-0.0016
0.90	1.2881	1.3008	-0.0127
0.95	1.3598	1.3916	-0.0318
0.99	1.6734	1.5703	0.1030

Table 7: The quantile values of the Fifth data (after normalizing to unit mean) with the quantile values of $w_\beta(t)$ for $\beta = 8.7870$.

Quantile	Experimental	Standard Normal	Difference
0.0100	-2.3626	-2.3263	0.0363
0.0500	-1.4514	-1.6449	-0.1934
0.1000	-1.3093	-1.2816	0.0277
0.3000	-0.4563	-0.5244	-0.0681
0.5000	-0.0299	0	0.0299
0.7000	0.3966	0.5244	0.1278
0.9000	1.1074	1.2816	0.1742
0.9500	1.6831	1.6449	-0.0382
0.9900	2.6711	2.3263	-0.3447

Table 8: The quantile values of the Forest data (after normalizing to mean zero and variance one) with the quantile values of the standard normal distribution.

Quantile	Experimental	Standard Normal	Difference
0.0100	-2.1760	-2.3263	0.1504
0.0500	-1.5419	-1.6449	0.1030
0.1000	-1.2122	-1.2816	0.0694
0.3000	-0.5274	-0.5244	-0.0030
0.5000	-0.0201	0	-0.0201
0.7000	0.4871	0.5244	-0.0373
0.9000	1.2480	1.2816	-0.0336
0.9500	1.5587	1.6449	-0.0862
0.9900	2.9168	2.3263	0.5905

Table 9: The quantile values of the Fifth Street data (after normalizing to zero mean, unit variance) with the quantile values of the Standard Normal distribution.

6.2 MATLAB Code

```
function y = wig(x, beta)
% wig(x, beta)  Returns Wigner's surmise for the given beta at x.

alpha = beta./2 + 1;
g1 = gamma(alpha);
g2 = gamma(alpha - .5);

k = 2 * (g1./g2) .^ (beta + 1) ./ g2;
c = ( g1 ./ g2 ).^2;

y = k * x.^(beta) .* exp(-c.*x.^2);
```

Figure 11: MATLAB code for computing values of $w_\beta(t)$.

```
function X = wigrnd(n, beta)

% surmisernd(n) Returns a vector of n random values distributed with the
%               generalized surmise density function for the given beta. The
%               values are generated by the rejection method using an
%               exponential random variable.

% Setting up constants
c = ( gamma(.5*(beta+2)) / gamma(.5*(beta+1)) )^2;
m = (1 + sqrt(8*c*beta + 1))/(4*c);

% Factor to multiply exp by
a = exp(m) * wig(m, beta);

for i = 1:n
    done = 0;
    while (done == 0)
        z = exprnd(1);
        u = rand;
        if a * u * exp(-z) <= wig(z, beta)
            done = 1;
        end
    end
    X(i) = z;
end
```

Figure 12: MATLAB code for sampling from $w_\beta(t)$.

```
function b = wigmle(X, min, max)
% function wigmle(X)   Finds the maximum likelihood estimation of beta given a
%                      set of points X and some expected interval.

f = @(beta) sum( log( wig(X, beta) ) );

% minimum of the negative == maximum of the positive
fneg = @(beta) -f(beta);
b = fminbnd(fneg, min, max);
```

Figure 13: MATLAB code for calculating the maximum likelihood estimator for β.

References

[1] A.Y. Abul-Magd. *Modelling gap-size distribution of parked cars using random-matrix theory.* arXiv:physics/0510136.

[2] C. Chan. *Beta Estimator.* http://people.csail.mit.edu/cychan/BetaEstimator.html

[3] Y. V. Fyodorov. *Introduction to the Random Matrix Theory: Gaussian Unitary Ensemble and Beyond.* arXiv:math-ph/0412017.

[4] M.L. Mehta. *Random Matrices, 3rd ed.* New York: Academic Press, 1991.

[5] F. Mezzadri. *Random Matrix Theory.* http://www.maths.bris.ac.uk/ mafm/lectures/RMT

[6] S. Rawal, G.J. Rodgers. *Modelling the gap size distribution of parked cars.* Physica A, Volume 346, Issue 3-4, p. 621-630.

[7] Eric W. Weisstein. "Wigner's Semicircle Law." From *MathWorld*–A Wolfram Web Resource. http://mathworld.wolfram.com/WignersSemicircleLaw.html