18.338 Final Project Report
Largest-Eigenvalue of Jacobi Ensembles and the Tracy-Widom Distribution

Qing He

May 17, 2013

1 Introduction

In this report, we will demonstrate through Matlab simulation that the empirical distribution of the largest eigenvalue of Jacobi ensembles tends towards the Tracy-Widom distribution.

2 Statement of Results

Here, we restate the result described in [1]. Let X be an $m \times n$ normal data matrix: each row is an independent observation from $\mathbf{N}_n(0, \sigma)$. A $n \times n$ matrix $A = X'X$ is then said to have a Wishard distribution $A \sim W_n(\sigma, m)$.

Let $A \sim W_n(I, m_1)$ be independent of $B \sim W_n(I, m_2)$, where $m_1 \geq n$. Assume n is even and that n, m_1, and m_2 satisfy

$$\lim_{n\to\infty} \frac{\min(n, m_2)}{m_1 + m_2} > 0, \tag{1}$$

and,

$$\lim_{n\to\infty} \frac{n}{m_1} < 1. \tag{2}$$

Then, the largest eigenvalue of $B/(A+B)$ is a random variate having distribution θ_n. Let W_n be the logit transform of θ_n such that $W_n = \text{logit}\theta_n = \log\frac{\theta_n}{1-\theta_n}$. Let us define Z as a shifted and rescaled version of W_n such that $Z = \frac{W_n - \mu_n}{\sigma_n}$, then Z is approximately Tracy-Widom distributed. The the centering and scaling parameters are given by,

$$\mu_n = 2\text{logtan}\left(\frac{\phi+\gamma}{2}\right)$$

$$\sigma_n^3 = \frac{16}{(m_1+m_2-1)^2}\frac{1}{\sin^2(\phi+\gamma)\sin\phi\sin\gamma}$$

and the angle parameters ϕ, γ are defined by,

$$\sin^2\left(\frac{\gamma}{2}\right) = \frac{\min(n,m_2)-1/2}{m_1+m_2-1}$$

$$\sin^2\left(\frac{\phi}{2}\right) = \frac{\max(n,m_2)-1/2}{m_1+m_2-1}$$

3 Simulation Script

First, we generate the empirical distribution of θ, by taking samples of the largest eigenvalue of Jacobi ensembles, generating the histogram and making appropriate normalization. The script for generating the empirical distribution is shown in Figure 1.

Since W_n is the logit transform of θ such that $W_n = \log\frac{\theta_n}{1-\theta_n}$ (i.e. $\theta = \frac{e^{W_n}}{1+e^{W_n}}$). Hence, the derived distribution of W_n is given by:

$$f(W_n) = f_\theta\left(\frac{e^{W_n}}{1+e^{W_n}}\right)\left|\frac{e^{W_n}}{1+e^{W_n}} - \left(\frac{e^{W_n}}{1+e^{W_n}}\right)^2\right| \quad (3)$$

$$= f_\theta(\theta)\left|\theta-\theta^2\right| \quad (4)$$

The script for generating the derived distribution of f_{W_n} is given in Figure 2.

```matlab

%Simulate the empirical distribution of the largest eigenvalue
%of Jacobi ensemble

for ii = 1:nsamples

    G1 = randn(m1,n); G2=randn(m2,n);
    A = G1'*G1; B = G2'*G2;
    J = B/(A+B);
    lambda = eigs(J);
    lambda_sample(ii) = max(lambda);

end

[count x_val] = hist(lambda_sample,numBin);

theta = x_val; %theta: r.v. denoting the largest eigenvalue of Jacobi ensemble
f_theta = count/sum(count)/((x_val(end)-x_val(1))/numBin); %pdf of theta

```

Figure 1: MATLAB script: generate the empirical distribution of θ

Lastly, the derived distribution of Z (shifted and rescaled version of W_n) is given by:

$$Z = \frac{W_n - \mu_n}{\sigma_n}$$

and,

$$f_Z = f_{W_n}|\sigma_n|$$

The script for generating the derived distribution of Z is given in Figure 3.

```
Wn = log(theta./(1-theta));
f_Wn = f_theta.*abs(theta-theta.^2);
```

Figure 2: MATLAB script: generate the derived distribution of f_{W_n}

```
Z = (Wn-mu)/sigma;
f_Z = f_Wn*abs(sigma);
```

Figure 3: MATLAB script: generate the derived distribution of f_Z

4 Simulation Results

In this section we show the simulations plots using various simulation parameters.

First, we fix all other parameters and increase n, the dimension of the Jacobi ensembles. The parameters are given by,

- number of samples: 30,000
- nbin = 50
- $m_1 = 1.5n$ and $m_2 = 2n$

Figure 4 correspond to the case where, $n = 30$, $m_1 = 1.5n$ and $m_2 = 2n$. The red curve corresponds to the Tracy-Widom distribution with $\beta = 1$ and the blue curve corresponds to the simulated distribution for Z.

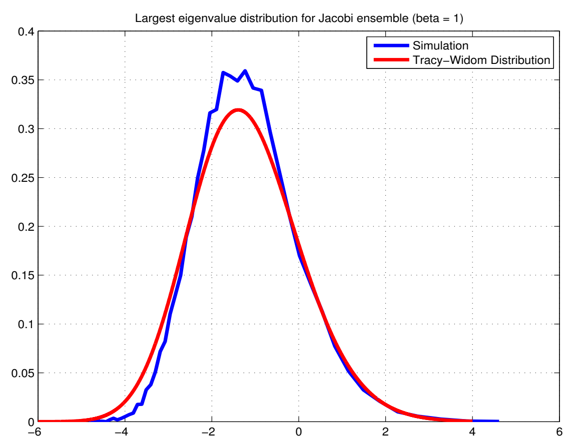

Figure 4: $n = 30$; $m_1 = 1.5n$ and $m_2 = 2n$

Figures 5, 6, 7 and 8 correspond to the case where $n = 80$, $n = 150$, $n = 500$ and $n = 1000$ respectively. As we can see, with $n = 50$, the simulated distribution has converged to the Tracy-Widom distribution.

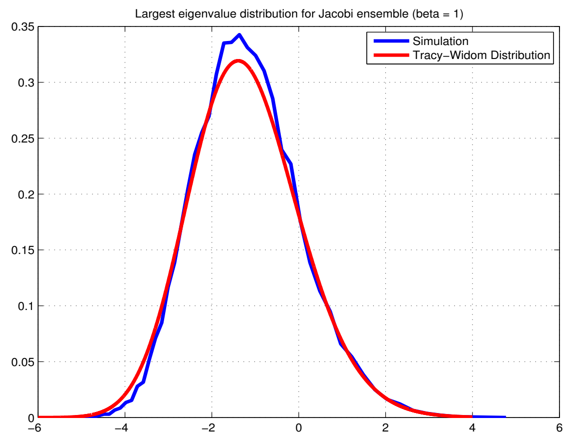

Figure 5: $n = 80$; $m_1 = 1.5n$ and $m_2 = 2n$

Next, we vary the parameters m_1 and m_2, with respect to a fixed n. Figure 9 corresponds to n = 500; $m_1 = 10n$ and $m_2 = n$.

Figure 10 corresponds to n = 500; $m_1 = 10n$ and $m_2 = 10n$.

Figure 11 corresponds to n = 300; $m_1 = 1.1n$ and $m_2 = 5n$.

We can see that the simulated distribution and the theoretical Tracy-Widom distribution coincide for all different m_1 and m_2 values.

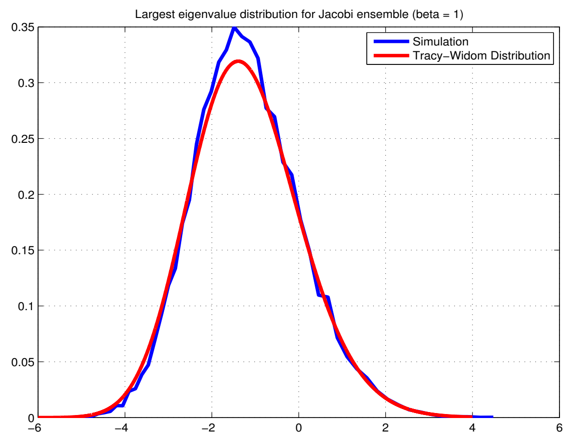

Figure 6: $n = 150$; $m_1 = 1.5n$ and $m_2 = 2n$

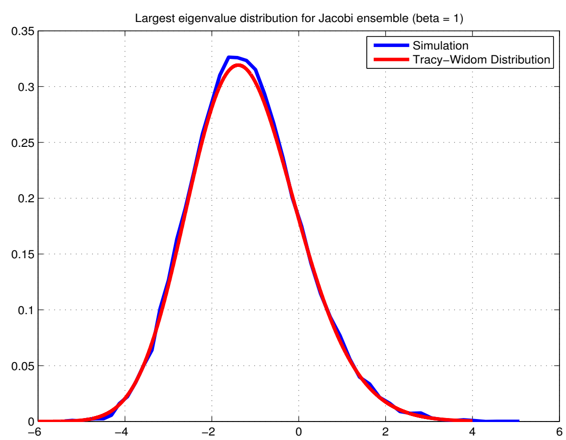

Figure 7: $n = 500$; $m_1 = 1.5n$ and $m_2 = 2n$

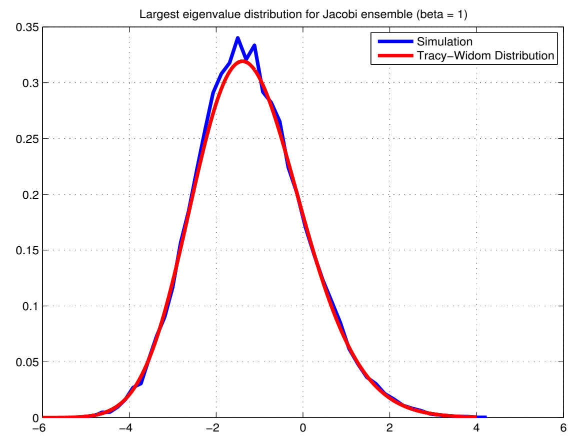

Figure 8: $n = 1000$; $m_1 = 1.5n$ and $m_2 = 2n$

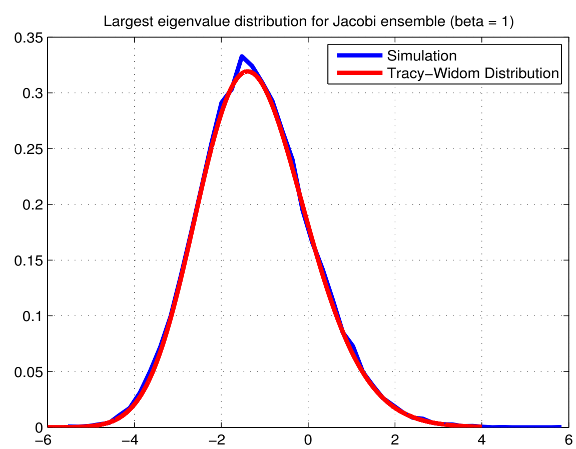

Figure 9: n = 500; $m_1 = 10n$ and $m_2 = n$

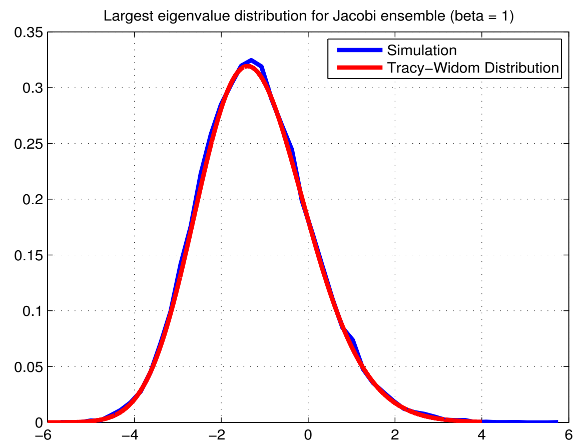

Figure 10: n = 500; $m_1 = 10n$ and $m_2 = 10n$

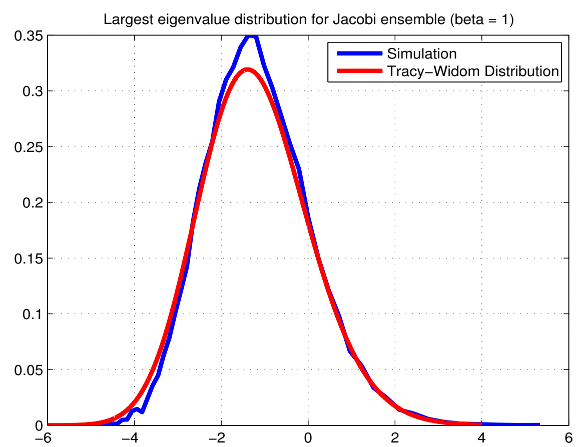

Figure 11: n = 300; $m_1 = 1.1n$ and $m_2 = 5n$

5 Conclusions

In this report, we have demonstrated using simulations that the simulated empirical distribution of the largest eigenvalue of Jacobi ensembles converges to the Tracy-Widom distribution at n increases. We have also showed that the convergence holds for various m_1 and m_2 values.

References

[1] Iain M. Johnstone Multivariate Analysis And Jacobi Ensembles: Largest Eigenvalue, Tracy-Widom Limits and Rates of Convergence NIH-PA Author Manuscript, February 12, 2010

```
% MATLAB code to demonstrate convergence of the distribution
% of the logit transform of the largest eigenvalues of
% Jacobi ensembles to the Tracy-Widom distribution

clear all;
close all;

n = 300;
nsamples  = 3000;
numBin = 50;

m1 = 2*n;
m2 = 5*n;

xaxis = -6:0.1:4;
beta = 1;

%Compute parameters
gamma = 2*asin(sqrt((min(n,m2)-1/2)/(m1+m2-1)));
phi = 2*asin(sqrt((max(n,m2)-1/2)/(m1+m2-1)));

mu = 2*log(tan((gamma+phi)/2));
alpha1= 16/((m1+m2-1)^2);
alpha2 = 1/((sin(gamma+phi)^2)*sin(gamma)*sin(phi));
sigma3 = alpha1*alpha2;
sigma = sigma3^(1/3);

%Simulate the empirical distribution of the largest eigenvalue
%of Jacobi ensemble

for ii = 1:nsamples

    G1 = randn(m1,n); G2=randn(m2,n);
    A = G1'*G1; B = G2'*G2;
    J = B/(A+B);
    lambda = eigs(J);
    lambda_sample(ii) = max(lambda);

end

[count x_val] = hist(lambda_sample,numBin);

theta = x_val; %theta: r.v. denoting the largest eigenvalue of Jacobi ensemble
f_theta = count/sum(count)/((x_val(end)-x_val(1))/numBin); %pdf of theta


Wn = log(theta./(1-theta));
f_Wn = f_theta.*abs(theta-theta.^2);

Z = (Wn-mu)/sigma;
f_Z = f_Wn*abs(sigma);
```

```
%Simulation plot
final_plot = figure;
grid on;
plot(Z, f_Z, 'b', 'LineWidth', 3);
hold on;

%Plot Tracy_Widom
load TW_beta1.mat
plot(x, TW_s_tag, 'r', 'LineWidth', 3);
legend('Simulation', 'Tracy-Widom Distribution')
title('Largest eigenvalue distribution for Jacobi ensemble (beta = 1)');
grid on;

saveas(final_plot,'final_plot','fig');
```

Published with MATLAB® 8.0