Largest-Eigenvalue of Jacobi Ensembles and the Tracy-Widom Distribution

Qing He

MIT

May 13, 2013

Statement of Results

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

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

and,

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

Statement of Results — continued

- 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 (i.e. $W_n = \text{logit}\theta_n = \log\frac{\theta_n}{1-\theta_n}$)
- Let $Z = \frac{W_n - \mu_n}{\sigma_n}$, then Z is approximately Tracy-Widom distributed. where the centering and scaling parameters are given by,

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

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

Statement of Results — continued

where the angle parameters ϕ, γ are defined by,

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

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

Simulation — Script

Generate the empirical distribution of θ (random variable denoting the largest eigenvalue of Jacobi ensemble)

```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

```

Simulation — Script

Derived distribution of W_n, the logit transform of θ

- $W_n = \log \frac{\theta_n}{1-\theta_n}$ (i.e. $\theta = \frac{e^{W_n}}{1+e^{W_n}}$)
- $$f(W_n) = f_\theta\left(\frac{e^{W_n}}{1+e^{W_n}}\right)\left|\frac{e^{W_n}}{1+e^{W_n}} - \left(\frac{e^{W_n}}{1+e^{W_n}}\right)^2\right|$$
  $$= f_\theta(\theta)\left|\theta - \theta^2\right|$$

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

Simulation — Script

Derived distribution of Z (shifted and rescaled version of W_n).

- $Z = \frac{W_n - \mu_n}{\sigma_n}$
- $f_Z = f_{W_n} |\sigma_n|$

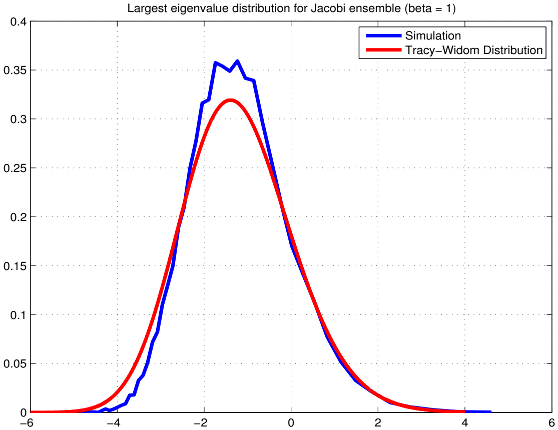
```
Z = (Wn-mu)/sigma;
f_Z = f_Wn*abs(sigma);
```

Simulation — Parameters

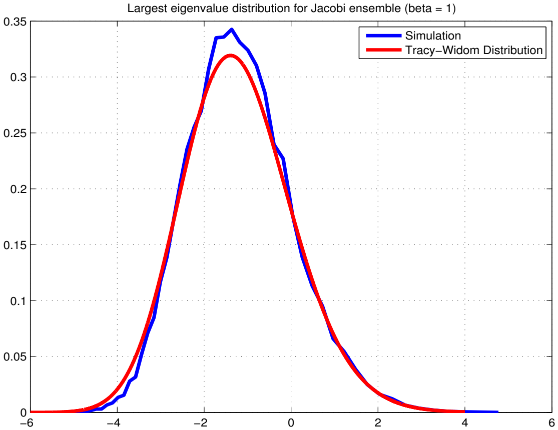
Parameters:

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

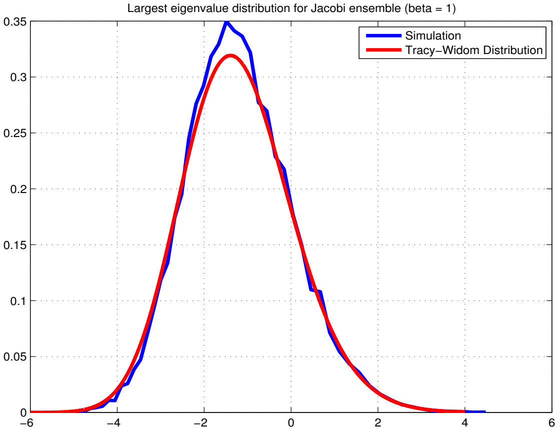
Simulation: n = 30; $m_1 = 1.5n$ and $m_2 = 2n$



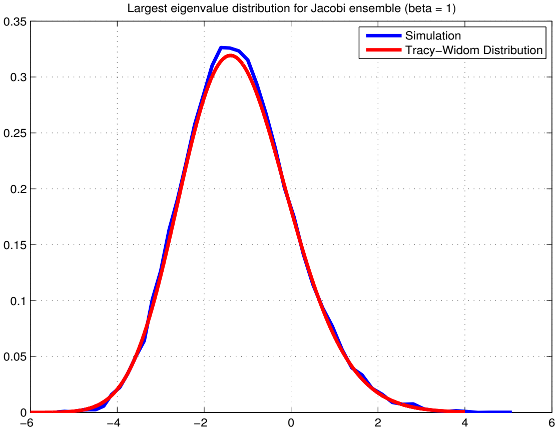
Simulation: n = 80; $m_1 = 1.5n$ and $m_2 = 2n$



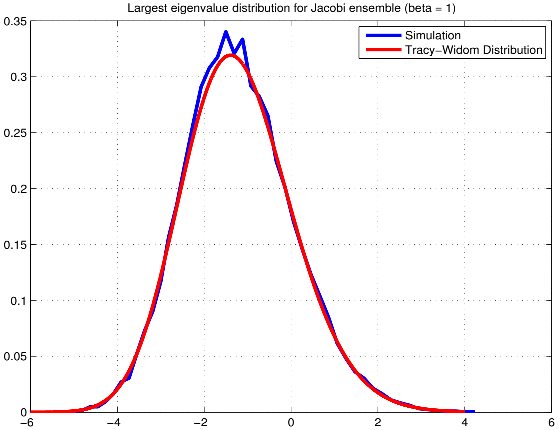
Simulation: n = 150; $m_1 = 1.5n$ and $m_2 = 2n$



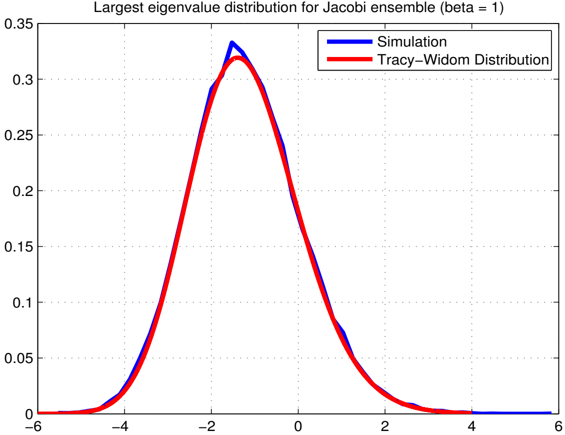
Simulation: n = 500; $m_1 = 1.5n$ and $m_2 = 2n$



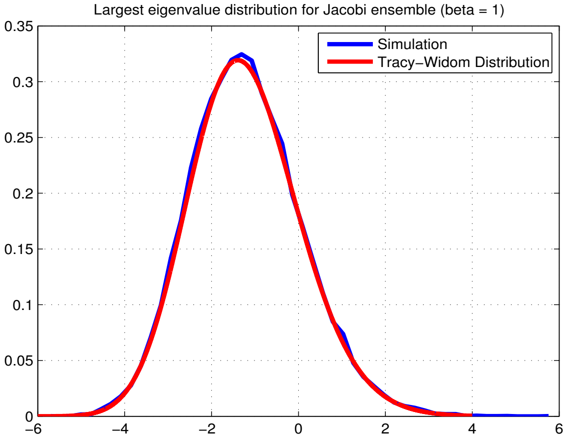
Simulation: n = 1000; $m_1 = 1.5n$ and $m_2 = 2n$



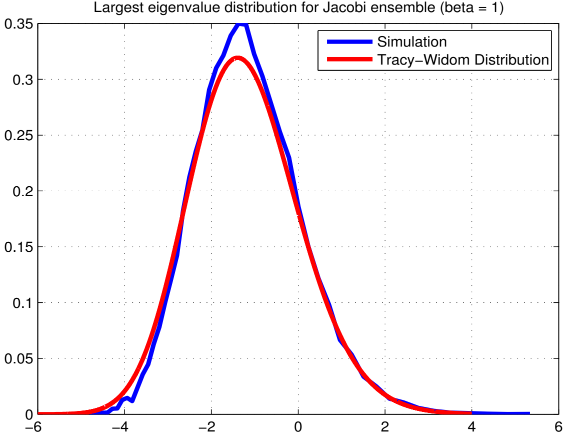
Simulation: n = 500; $m_1 = 10n$ and $m_2 = n$



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



Simulation: n = 300; $m_1 = 1.1n$ and $m_2 = 5n$



Thank you !