Eigenvalues of large sample covariance matrices; Lecture 4

Jinho Baik
University of Michigan, Ann Arbor

March 2006

Limiting distribution functions

(a) $t_1 \leq 1 + \sqrt{c}$:

$$F_d(x) = \det(1 - \mathcal{A}_x^{(d)})$$

$d = 0$: Airy kernel

$d > 0$: a generalization

(b) $t_1 > 1 + \sqrt{c}$:

$$G_d(x) = \det(1 - \mathcal{H}_x^{(d)})$$

Distribution $G_d(x)$

$d = 1$: $N(0, 1)$

$$G_1(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a^2} da$$

$d > 1$: compare with the case $T_p = T_d = t_1 I$ when $p = d$ fixed, $n \to \infty$,

Recall: eigenvalue density of $B_p = B_d$:

$$\mathbb{P}(s_{\max} \le t) = c \cdot \int_0^t \cdots \int_0^t p(s) ds$$

where

$$\begin{aligned} p(s) &= c \cdot \prod_{j<k} (s_j - s_k)^2 \prod_{j=1}^{d} s_j^{n-d} e^{-n t_1^{-1} s_j} \\ &= c \cdot \Delta(s)^2 \prod_j e^{-n f(s_j)} s_j^{-d} \end{aligned}$$

and

$$f(s) = t_1^{-1} s - \log s$$

Critical point $f'(s) = t_1^{-1} - \frac{1}{s}$.
Hence $s_c = t_1$ and $f''(t_1) = t_1^{-2}$.

For d fixed, $n \to \infty$, setting $\frac{n^{1/2}}{t_1}(s_j - t_1) = a_j$,

$$\lim_{n\to\infty} \mathbb{P}\left((s_{\text{max}} - t_1) \frac{n^{1/2}}{t_1} x \right)$$

$$= c' \cdot \int_{-\infty}^{x} \cdots \int_{-\infty}^{x} \Delta(y)^2 \prod_{j=1}^{d} e^{-\frac{1}{2}a_j^2} s_j^{-d}$$

The largest eigenvalue of $d \times d$ GUE (Gaussian unitary ensemble).

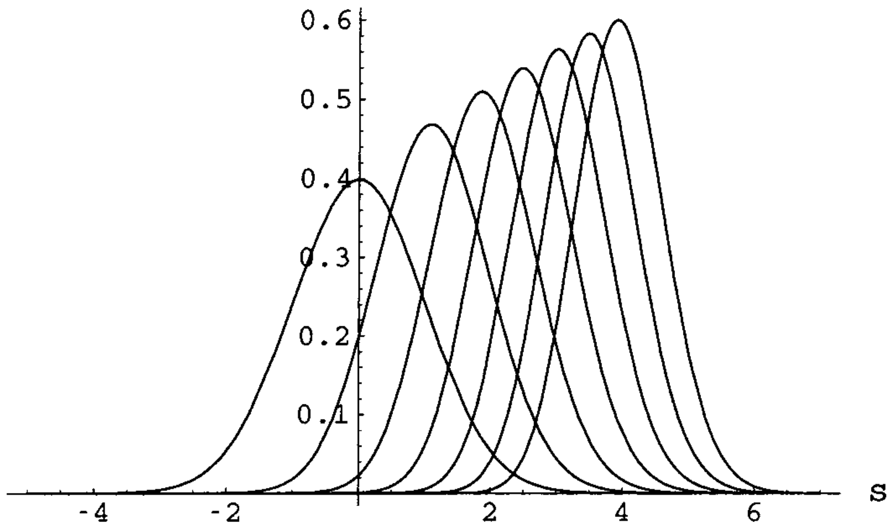
Compare with $n, p \to \infty$, $\frac{p}{n} \to \infty$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left(s_{\text{max}} - (t_1 + \frac{ct_1}{t_1 - 1}) \right) \frac{n^{1/2}}{\nu'} \le x \right) = G_d(x)$$

with $\nu' = \frac{t_1\sqrt{(t_1-(1-\sqrt{c}))(t_1-(1+\sqrt{c}))}}{t_1-1}$

Extra $\frac{ct_1}{t_1-1}$ in the mean is the effect of infinitely many unit eigenvalues of T_p

But the limiting distribution $G_d(x)$ turns out to be the same as finite d case.



	mean	variance
G_1	0	1
G_2	1.12838	0.72676
G_3	2.52811	0.61474
G_4	3.06327	0.50426

Distribution $F_d(x)$

$d = 0$: $F_0(x)$ is the limiting distribution of the null case

Universality in random matrix theory: F_0 is also the limiting distribution of GUE and other random matrix models [Soshnikov, Bleher-Its, Deift+Gioev]

Note that the Fredholm determinant formula of F_0 is not practical to plot the graph. Alternative formula [Tracy+Widom]

Let $u(x)$ be the solution of $u'' = 2u^3 + xu$ (Painlevé II equation) satisfying $u(x) \sim -\mathrm{Ai}(x)$ as $x \to +\infty$. Then

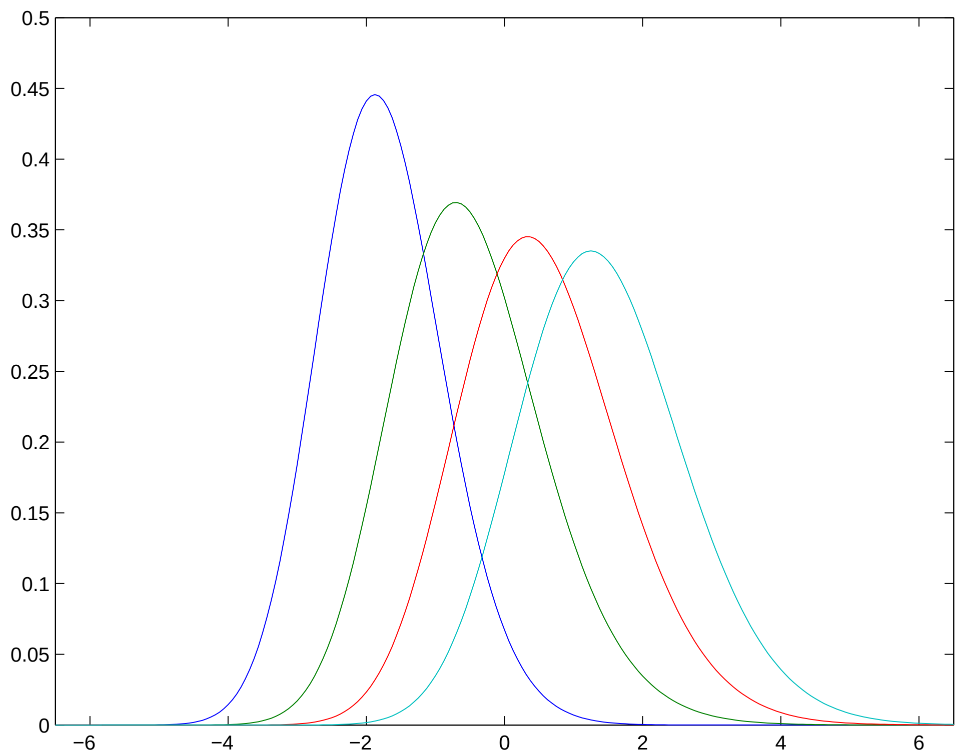
$$F_0(x) = e^{-\int_x^\infty (y-x)u^2(y)dy}$$

$F_0(x) = 1 + O(e^{-cx^{3/2}})$ as $x \to +\infty$
$F_0(x) = O(e^{-c|x|^3})$ as $x \to -\infty$

$d > 0$: Similar formula? [Baik]

Set $\mathcal{E}(x) = e^{\int_x^\infty u(s)ds}$.

$$F_1(x) = F_0(x)\,\mathcal{E}(x)$$
$$F_2(x) = F_0(x)\,\mathcal{E}(x)^2\{1 + u(x + 2u^2 + 2u')\}$$
$$F_3(x) = F_0(x)\,\mathcal{E}(x)^3\{1 + 2u(x + 2u^2 + 2u') + \frac{1}{2}(u^2 - u')(x + 2u^2 + 2u')^2\}$$



	mean	standard deviation
F_0	-1.771. . .	0.90. . .
F_1	-0.494. . .	1.11. . .
F_2	0.543. . .	1.18. . .
F_3	1.445. . .	1.21. . .

Observation: $F_1(x) = (F_{GOE}(x))^2$

$F_{GOE}(x)$ is also the limiting distribution function of the largest sample eigenvalue of *real* Gaussian samples when $T_p = I$

Note: Consider two sets of eigenvalues from GOE, and superimpose them. The largest of them converges in distribution to $F_1 = F_{GOE}^2$ and the second largest converges to F_0. [Baik+Rains]

(1) Define

$$C_w(u) = \frac{1}{2\pi} \int e^{i(\frac{1}{3}a^3 + ua)} \frac{1}{w + ia} da$$

Let

$$f(x; w) = 1 - \int_x^\infty \left(\frac{1}{1 - \mathcal{A}_x} C_w \right)(u) \, \mathrm{Ai}(u) du$$

Then $F_d(x) = F_d(x; 0, 0, \ldots, 0)$ where

$$F_d(x; w_1, \ldots, w_d) = F_0(x) \cdot \frac{\det\left((w_j + \frac{\partial}{\partial x})^{k-1} f(x, w_j) \right)_{d \times d}}{\prod_{j<k}(w_k - w_j)}$$

(2) $f(x,w)$ satisfies

$$-\frac{\partial^2 f}{\partial x^2} + \left(\frac{u'}{u} - w\right)\frac{\partial f}{\partial x} + u^2 f = 0$$

and

$$- \frac{\partial^2 f}{\partial w^2} + \left(\frac{u}{wu + u'} + w^2 - x\right)\frac{\partial f}{\partial w}$$
$$+ \left(-\frac{u^3}{wu + u'} + u^4 + xu^2 - (u')^2\right) f = 0$$

And

$$f(x,0) = \mathcal{E}(x), \quad \frac{\partial f}{\partial x}(x,0) = -u(x)\,\mathcal{E}(x),$$
$$\frac{\partial f}{\partial w}(x,0) = (u + u')\,\mathcal{E}$$

Transition around $t_1 = 1 + \sqrt{c}$

For $j = 1, \ldots, d$,

$$t_j = 1 + \sqrt{c} - \frac{c^{1/6}(1+\sqrt{c})^{2/3} w_j}{n^{2/3}},$$

then

$$\lim_{n\to\infty} \mathbb{P}\left((s_{\mathsf{max}} - \mu)\frac{n^{2/3}}{\nu} \le x \right) = F_d(x; w_1, \ldots, w_d)$$

Curious coincidence?

The distribution of s_{max} describes other probabilistic models: last passage percolation, queues in tandem, interacting particle systems

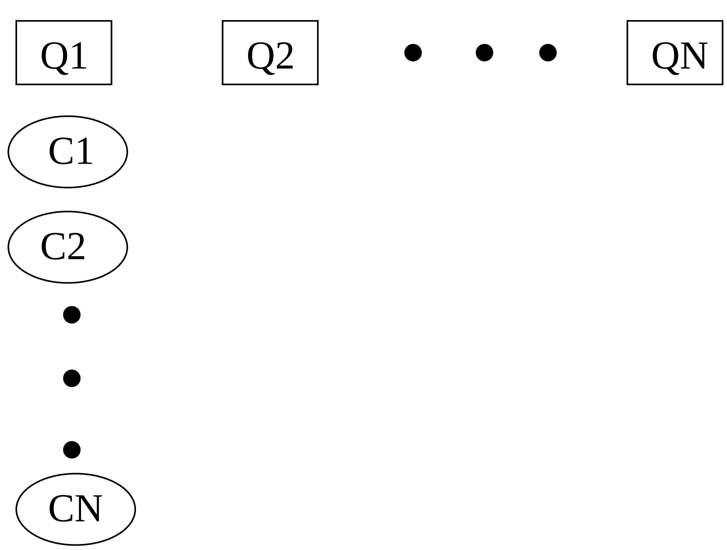
[Johansson 2000 CMP] [Okounkov] [BBP, Baik]

Queues in tandem

n customers, p servers; each customer go through all servers.

$X(i,j)$= service time for ith customer at jth server

Initial condition ($N = n$, $k = p$):



$\mathbf{D(n,p)}$:= departure time of nth customer from all p queues

Service time: $X(i,j)$ is exponential of rate t_j^{-1}

It turned out for each $n \geq p$,

$$D(n,p) =^d s_{max}(n,p)$$

Not from a mapping between models, but from independent computation of the distribution function

Hence even if a few servers are slower than others, the total departure time is not changed (to the second order) as long as the rate of the service time is not slower by a factor larger than $1 + \sqrt{c}$

(Symmetry: the results also apply to slow customers)

Interacting particle systems in 1D

- discrete space $\mathbb{Z}$, continuous time

- at most one particle at each site

- all jumps are to the right neighboring site

- when the right next site becomes vacant, the particle waits random time then jumps

1	1	1	0	1	1	0	1	0	0	0
1	1	1	0	1	0	1	1	0	0	0

- Initial configuration $(\cdots, 1, 1, 1, 0, 0, 0, \cdots)$

jth particle jumps with rate t_j^{-1}
=waiting time for a jump is exponential of rate t_j

$t_j = 1$ for all j: **totally asymmetric simple exclusion process**

$T(n,p)$:= time for the pth particle to perform total n jumps (i.e. arrives at site $N+1-k$)

It turned out for $n \geq p$,

$$T(n,p) =^d s_{\text{max}}(n,p)$$

Integrated density of particles at a given location

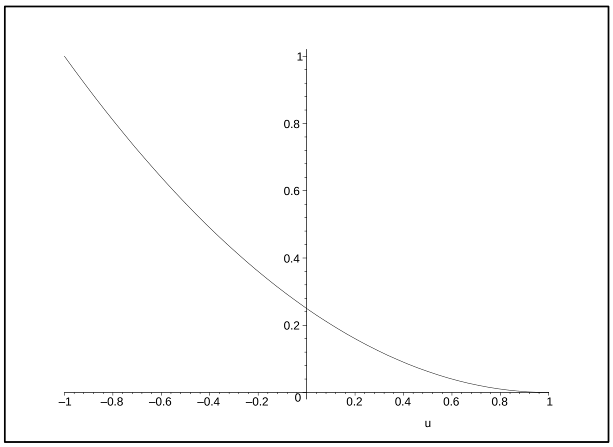
$\#(t;k)$:= number of particles to the right of the site k at time t = integrated current

$$\#(t;k) > p \Leftrightarrow x_p(t) \geq k+1 \Leftrightarrow T(k+p,p) \leq t$$

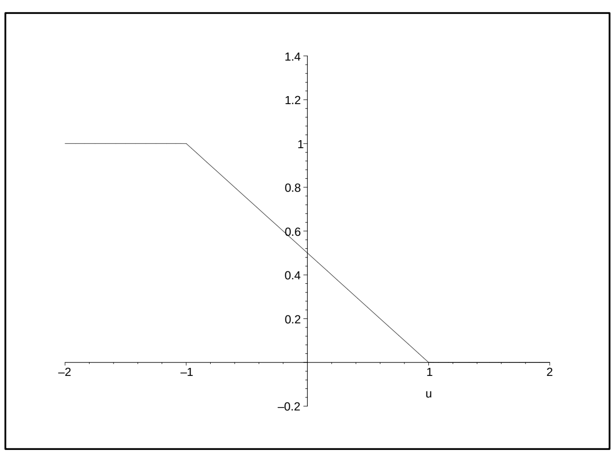
$t_j = 1$ **for all** j

Global profile [Rost 1981] $t \to \infty$

$$\frac{1}{t}\#(t;[ut]) \to \frac{1}{4}(1-u)^2 =: I(u), \quad |u| < 1.$$



Current density: $-I'(u) = \frac{1}{2}(1-u), \ -1 < u < 1.$



Fluctuation [Johansson 2000]: $1 - F_0(-x)$

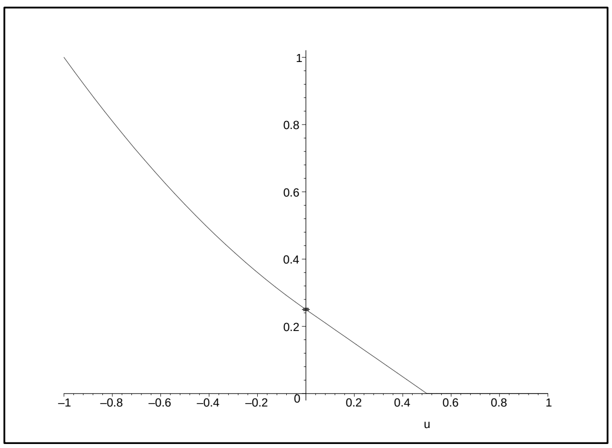
A few slow cars

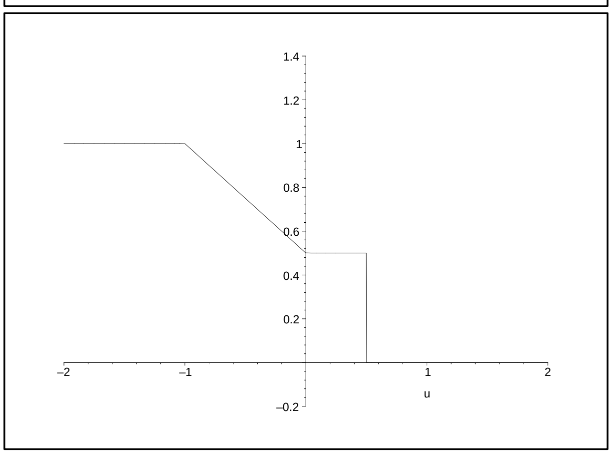
All but first a few cars jump at rate 1

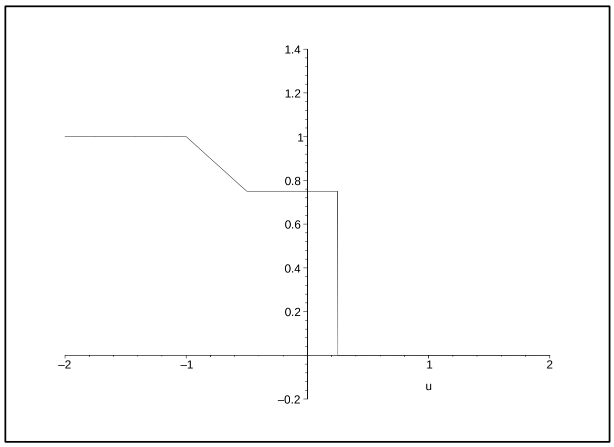
$r :=$ **smallest jump rate (slowest car)**

Global profile: when $r < 1$,

$$\frac{1}{t}\#(t;[ut]) \to \begin{cases} \frac{1}{4}(1-u)^2, & -1 < u < 2r-1 \\ (1-r)(r-u), & 2r-1 < u < r. \end{cases}$$







Fluctuation: $F_m(x)$ or $G_m(x)$

Slow start from stop

There is a symmetry of $L(i,j)$ and $L(j,i)$

Cars jump with slower rates for the first j_0 than the subsequent jumps

Global profile: when $r < 1$,

$$\frac{1}{t}\#(t;[ut]) \to \begin{cases} \frac{1}{4}(1-u)^2, & 1-2r < u < 1 \\ r - r^2 - ru, & -r < u < 1-2r. \end{cases}$$