Eigenvalues of large sample covariance matrices; Lecture 2

Jinho Baik
University of Michigan, Ann Arbor

February 2006

Complex Gaussian (normal) samples

n= sample size
p= population size (dimension of vectors)

Assume $n \geq p$

T_p=population covariance matrix

Complex Gaussian sample vectors $\vec{x}_j$ **with co-variance** T_p:

Sample matrix $X = T_p^{1/2} Z_p \qquad (p \times p)(p \times n)$

Z_{ij} are independent normalized Gaussian $\mathbb{E}(Z_{11}) = 0$, $\mathbb{E}|Z_{11}|^2 = 1$

Sample covariance matrix

$$B_p := \frac{1}{n} X X^* = \frac{1}{n} T_p^{1/2} Z_p Z_p^* T_p^{1/2}$$

- Density function of B_p
- Density function of the eigenvalues of B_p
- Distribution function of the largest eigenvalue

Later: asymptotic analysis of the distribution function

Density function of B_p

Density function of Z_p:

$$\frac{1}{\pi^{pn}} \prod_{i=1}^{p} \prod_{j=1}^{n} e^{-|z_{ij}|^2} (dZ) \propto e^{-\operatorname{Tr}(ZZ^*)} (dZ)$$
$$= e^{-n \operatorname{Tr}(T^{-1}B)} (dZ)$$

Change of variables $Z \mapsto B$: $B = \frac{1}{n} T^{1/2} Z Z^* T^{1/2}$

Z: $2np$ independent variables
B: p^2 independent variables

(Example) $p = 1$, $n = 2$; $[b] = \frac{1}{2}[t^{1/2}][z_1, z_2]\left[\begin{smallmatrix}\overline{z_1} \\ \overline{z_2}\end{smallmatrix}\right][t^{1/2}]$

$b = \frac{t}{2}(|z_1|^2 + |z_2|^2) = \frac{t}{2}(x_1^2 + y_1^2 + x_2^2 + y_2^2)$

Set $c_1 = y_2$, $c_2 = x_1, c_3 = y_2$

$(x_1, y_1, x_2, y_2) \mapsto (b, c_1, c_2, c_3)$

$\frac{\partial(b,c_1,c_2,c_3)}{\partial(x_1,y_1,x_2,y_2)} = tx_1 = t\sqrt{\frac{2b}{t} - |c|^2}$

$(dZ) = \frac{1}{t\sqrt{\frac{2b}{t}-|c|^2}} db\, dc_1 dc_2 dc_3$

Integrate out c_1, c_2, c_3:

$$\left\{ \iiint_{\{|c|<\sqrt{\frac{2b}{t}}\}} \frac{1}{t\sqrt{\frac{2b}{t} - |c|^2}} dc_1 dc_2 dc_3 \right\} db$$

$$= \left\{ 4\pi \int_0^1 \frac{(\frac{2b}{t})^{\frac{3}{2}}\rho^2}{t\sqrt{\frac{2b}{t}}\sqrt{1-\rho^2}} d\rho \right\} db$$

$$\propto \frac{2b}{t^2} db$$

$$\propto (\det B)^{n-p} (dB)$$

In general, [Wishart 1928] (dZ)‘ $='$ $c{\cdot}(\det B)^{n-p}(dB)$;

$$p(B)(dB) = c \cdot e^{-n\,\mathrm{Tr}(T^{-1}B)} (\det B)^{n-p} (dB)$$

see [Forrester] ‘Log-gases and Random matrces’ Section 2.6.2

`http://www.ms.unimelb.edu.au/~matpjf/matpjf.html`

Density function of (sample) eigenvalues

$B = QSQ^{-1}$

- Q= unitary matrix
- $S = diag(s_1, s_2, \ldots, s_p)$ $\qquad (s_1 \geq \cdots \geq s_p \geq 0)$

Change of variables $B \mapsto (S, Q)$:

$\frac{\partial(B)}{\partial(S,Q)} = \prod_{1 \leq i < j \leq p} |s_i - s_j|^2 =: \Delta(S)^2$ (Weyl's formula)

Hence

$$p(B)(dB) = p(QSQ^{-1})\Delta(S)^2(dQ)(dS)$$
$$= c \cdot e^{-n \operatorname{Tr}(T^{-1}QSQ^{-1})}(\det S)^{n-p}\Delta(S)^2(dS)(dQ)$$

Integrate out Q [James 1964]

$$p(S)(dS) = c \cdot (\det S)^{n-p} \Delta(S)^2 (dS) \times \int_{Q \in U(p)} e^{-n \cdot \mathrm{Tr}(T^{-1} Q S Q^{-1})} (dQ)$$

When $T = I$, integral$= e^{-n\,\mathrm{Tr}(S)}$;

$$p(s)ds = c \cdot (\det S)^{n-p} \Delta(S)^2 e^{-n\,\mathrm{Tr}(S)} ds = c \cdot \prod_{j<k} (s_j - s_k)^2 \prod_j s_j^{n-p} e^{-n s_j} ds_j$$

In general, [Harish-Chandra], [Itzykson+Zuber]:

integral$= \frac{\det(e^{-n t_j^{-1} s_k})}{\Delta(T^{-1}) \Delta(S)}$. Hence

$$p(s)ds = c \cdot \frac{\det(e^{-n t_j^{-1} s_k})}{\Delta(T^{-1})} \Delta(S) \prod_j s_j^{n-p} ds_j$$

over the space $s_1 \geq \cdots \geq s_p \geq 0$

Note that $p(s)$ is **symmetric** in $(s_1, \ldots, s_p)$: remove the ordering of $s_1, \ldots, s_p$ and think of $p(s)$ as **a density function on $\mathbb{R}^p_+$**

Distribution function of the largest eigenvalue

$$\mathbb{P}(s_{\text{max}} \leq x) = \int_{\mathbb{R}^p_+} p(s) \prod_{j=1}^{p} (1 - \chi_{(x,\infty)}) ds_j$$

We want limit as $n, p \to \infty$. Difficulty: number p of integrals increases

• Transform to a simpler formula involving only finitely many integrals which is more suitable for asymptotic analysis

Distribution function of the largest eigenvalue

Vandermonde determinant

$$\Delta(a) = \prod_{j<k} (a_j - a_k)$$
$$= \pm \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ a_1 & a_2 & a_3 & \cdots & a_p \\ a_1^2 & a_2^2 & a_3^2 & \cdots & a_p^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1^{p-1} & a_2^{p-1} & a_3^{p-1} & \cdots & a_p^{p-1} \end{pmatrix}$$
$$= \pm \det(a_j^{k-1})$$

Hence

$$p(s)ds = c \cdot \det(f_j(s_k)) \det(g_j(s_k))ds$$

- $T = I$: $f_j(u) = g_j(u) = u^{j-1}(u^{n-p}e^{-nu})^{1/2}$
- All t_j's distinct: $f_j(u) = u^{j-1+n-p}$, $g_j(u) = e^{-nt_j^{-1}u}$

∘ **General theory** [Mehta] [Tracy,Widom: JSP 1998]

$h(u)$=a test function

Andréief's identity (1883)

$$\int \cdots \int p(s) \prod_{j=1}^{p} (1+h(s_j)) ds_j$$
$$= c \cdot \int \cdots \int \det(f_j(s_k)) \det(g_j(s_k)) \prod_j (1+h(s_j)) ds_j$$
$$= c \cdot p! \det \left(\int f_j(u) g_k(u) (1+h(u)) du \right) =: (*)$$

(proof) 2×2 case:

$$\det \left(\int f_j(u) g_k(u) d\nu(u) \right)$$
$$= \int \int \det \begin{pmatrix} f_1(u_1)g_1(u_1) & f_1(u_2)g_2(u_2) \\ f_2(u_1)g_1(u_1) & f_2(u_2)g_2(u_2) \end{pmatrix} d\vec{\nu}(u)$$
$$= \int \int \det \begin{pmatrix} f_1(u_1) & f_1(u_2) \\ f_2(u_1) & f_2(u_2) \end{pmatrix} \cdot g_1(u_1) g_2(u_2) d\vec{\nu}(u)$$
$$= \int \int \det \begin{pmatrix} f_1(u_2) & f_1(u_1) \\ f_2(u_2) & f_2(u_1) \end{pmatrix} \cdot g_1(u_2) g_2(u_1) d\vec{\nu}(u)$$

It is ok to replace f_j by $\underline{\phi_j \in span\{f_1, \ldots, f_j\}}$.

Hence

$$(*) = c \cdot \det\left(\int \phi_j(u)\psi_k(u)(1+h(u))du\right)$$
$$= c \cdot \det(\Phi\Psi + \Phi H \Psi)$$

where

$$\Phi(j,u) = \phi_j(u), \quad \Psi(u,k) = \psi_k(u)$$

and H is the multiplication by $h(u)$.

For example, Φ is an operator from $L^2(\mathbb{R}_+, du)$ to $\ell^2(\{1,2,\ldots,p\})$

If $\det(\Phi\Psi) \neq 0$**,**

$$(*) = c \cdot \det(\Phi\Psi)\det(1 + \Phi H \Psi \frac{1}{\Phi\Psi}) = \det(1 + H\,\mathcal{K})$$

where

$$\mathcal{K} = \Psi \frac{1}{\Phi\Psi} \Phi$$

is an operator on $L^2(\mathbb{R}_+)$

Note: the constant factor c is gone as when $h=0$, both sides agree.

For example, $h(u) = -\chi_{(x,\infty)}$:

$$\mathbb{P}(s_{\mathsf{max}} \leq x) = \mathrm{det}(1 - \chi_{(x,\infty)}\, \mathcal{K})$$

Asymptotic analysis: if $A \to A_\infty$ in trace norm, then $\mathrm{det}(1 + A)t \to \mathrm{det}(1 + A_\infty)$.

Advantageous? In many cases, the kernel of $\mathcal{K}$ is 'simple' enough to be suitable for asymptotic analysis.

◦ **Our case**

(1) $T_p = I$:

$$p(s)ds = c \cdot \Delta(s)^2 \prod_j w(s_j)ds_j, \qquad w(u) = e^{-nu}$$

Then $f_j(u) = g_j(u) = u^{j-1}w(u)^{1/2}$.

Take

$$\phi_j(u) = \psi_j(u) = q_{j-1}(u)w^{1/2}(u)$$

where $q_j(u)=$ jth orthonormal polynomial with respect to $w(u)du$.

Then $(\Phi\Psi)(j,k) = \int q_{j-1}(u)q_{k-1}w(u)du = \delta_{jk}$.

Hence (Christoffel-Darboux kernel)

$$\mathcal{K}(u,v) = \sum_{j=1}^{p} \Psi(u,j)\Phi(j,v)$$

$$= (w(u)w(v))^{1/2} \sum_{j=1}^{p} q_{j-1}(u)q_{k-1}(v)$$

$$= (w(u)w(v))^{1/2} \frac{\gamma_{p-1}}{\gamma_p} \cdot \frac{q_p(u)q_{p-1}(v) - q_{p-1}(u)q_p(v)}{u-v}$$

(2) $T_p \neq I$, all t_j's distinct: [Okounkov] [Johansson] (see [BBP, Ann. Prob. 2005])

Here

$$f_j(u) = u^{j-1+n-p} = \phi_j(u), \quad g_j(u) = e^{-nt_j^{-1}u} = \psi_j(u)$$

Hence

$$(\Phi\Psi)(j,k) = \int_0^\infty u^{j-1+n-p} e^{-nt_k^{-1}u} du = \frac{\Gamma(j+n-p)}{(nt_k^{-1})^{j+n-p}}$$

and

$$\det(\Phi\Psi) = \Delta(n^{-1}t_k) \cdot \prod_{j=1}^{p} \frac{\Gamma(j+n-p)}{(nt_j^{-1})^{n-p+1}} \neq 0$$

By Cramer's rule,

$$\begin{aligned} \mathcal{K}(u,v) &= \sum_{m=1}^{p} \Psi(u,m)((\Phi\Psi)^{-1}\Phi)(m,v) \\ &= \sum_{m=1}^{p} \Psi(u,m) \frac{\det((\Phi\Psi)^{(m)}(v))}{\det(\Phi\Psi)} \end{aligned}$$

where $(\Phi\Psi)^{(m)}(v)$ is the matrix $\Phi\Psi$ with the mth column replaced by the vector $(\phi_1(v), \cdots, \phi_p(v))^T$.

Trick: write

$$\phi_a(v) = v^{a-1+n-p} = \frac{n\Gamma(a+n-p)}{2\pi i}\int_\Sigma \frac{e^{nvw}}{(nw)^{a+n-p}}dw$$

where $w = 0$ is inside Σ

Then

$$\det((\Phi\Psi)^{(m)}(v)) = \frac{n}{2\pi i}\int_\Sigma e^{nvw}\cdot \det(A)dw$$

where A has entries (Vendermonde again!)

$$A(j,k) = \frac{\Gamma(j+n-p)}{(nz_k)^{j+n-p}}$$

where

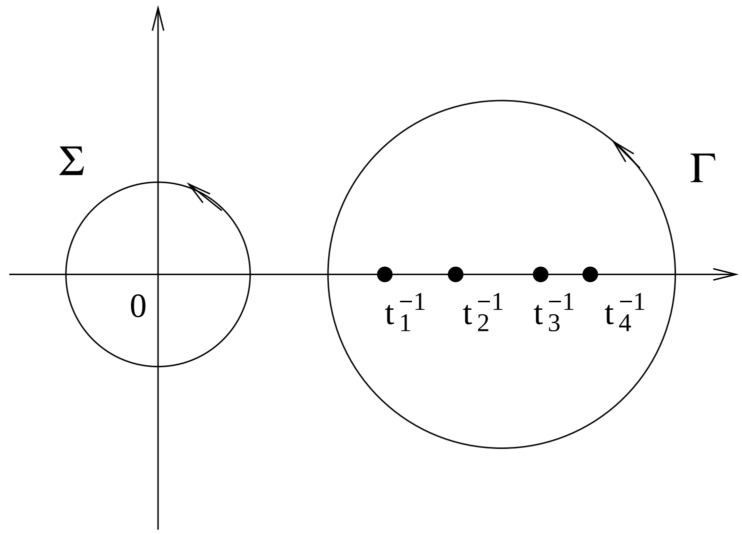
$$z_r = \begin{cases} t_r^{-1} & r \neq m \\ w & r = m. \end{cases}$$

Hence

$$\frac{\det((\Phi\Psi)^{(m)}(v))}{\det(\Phi\Psi)}$$

$$= \frac{n}{2\pi i}\int_\Sigma e^{nvw}\left[\prod_{r\neq m}\frac{w^{-1}-t_r}{t_m - t_r}\right](t_m w)^{-n+p-1}dw$$

$$= \frac{n}{2\pi i}\int_\Sigma e^{nvw}\left[\prod_{r\neq m}\frac{w-t_r^{-1}}{t_m^{-1}-t_r^{-1}}\right](t_m w)^{-n}dw$$



For Γ above ($w \in \Sigma$ is outside Γ)

$$\mathcal{K}(u,v)$$

$$= \frac{n}{2\pi i}\int_{\Sigma} dw \cdot w^{-n} e^{nvw} \left\{ \sum_{m=1}^{p} t_m^{-n} e^{-nt_m^{-1}u} \prod_{r \neq m} \frac{w - t_r^{-1}}{t_m^{-1} - t_r^{-1}} \right\}$$

$$= \frac{n}{(2\pi i)^2}\int_{\Sigma} dw \cdot w^{-n} e^{nvw}$$

$$\times \left\{ \int_{\Gamma} dz \cdot z^{n} e^{-nzu} \frac{1}{w - z} \prod_{k} \frac{w - t_k^{-1}}{z - t_k^{-1}} \right\}$$

As $Re(w) < Re(z)$, $\frac{1}{w-z} = -n \int_0^\infty e^{yn(w-z)} dy$.

Conclusion:

$$\mathcal{K} = \mathcal{I}\,\mathcal{J}, \quad \mathcal{K}(u,v) = \int_0^\infty \mathcal{I}(u+y)\,\mathcal{J}(y+v)dy$$

where

$$\mathcal{I}(u+y) = \frac{n}{2\pi}\int_\Gamma dz \cdot z^n e^{-n(u+y)z} \prod_k (z - t_k^{-1})^{-1}$$

$$\mathcal{J}(y+v) = \frac{n}{2\pi}\int_\Sigma dw \cdot w^{-n} e^{n(y+v)w} \prod_k (w - t_k^{-1})$$

This formula is valid even when some of t_k's coincide.