Extreme eigenvalue distributions of β-Jacobi ensembles and an application

Ioana Dumitriu

Department of Mathematics
University of Washington (Seattle)

Banff International Research Centre
April 13, 2016

1. Extreme value distributions for β-Hermite, -Laguerre, -Jacobi
2. PDFs for special β-Jacobi ($\Sigma = I_m$)
3. Application: RURV
   - Efficiency
   - Communication
4. More β-Jacobi eigenvalue calculations
5. Numerics
6. Conclusions

β-Hermite

- General $\beta > 0$.
- λ_{max}
  - Ramirez, Rider, Virag: asymptotic fluctuations given by stochastic Airy operator (following Edelman, Sutton).
  - Explicit, exact distributions, n fixed?...
- Smallest (in absolute value) eigenvalues?...
  - Perhaps not interesting; however, GUE absolute values (Edelman and La Croix) are a union of two Laguerre ensembles. What about β?

β-Laguerre

- General $\beta > 0$.
- $\underline{\lambda_{max}}$
  - Ramirez, Rider, Virag: asymptotic scaled fluctuations given by the stochastic Airy operator; scale depends on matrix dimensions.
  - One size much larger than the other: Jiang and Li showed scaled fluctuation converges to stochastic Airy operator limit. (Also LDP.)
  - CDF, PDF for λ_{max} in terms of hypergeometric functions of matrix argument (see Koev et al survey-like paper)
- $\underline{\lambda_{min}}$
  - Ramirez, Rider: asymptotic fluctuations given by stochastic Bessel operator (following Edelman, Sutton); when dimensions differ by a constant. Also tail analysis by Ramirez, Rider, Zeitouni.
  - Asymptotics for some cases covered by Forrester through a hypergeometric function limit.
  - Finite n: CDF for λ_{min} in terms of a hypergeometric function, PDF only in certain cases (when the hypergeometric terminates)

β-Jacobi

- General $\beta > 0$.
- CDF, PDF for λ_{min} and λ_{max} (Koev and D., D., Koev et al.)
- $\underline{\lambda_{max}}$
  - RRV?
  - Jiang: in special cases, stochastic Airy operator limits.
  - Forrester: LD for asymptotic distribution for finite aspect ratio.
- $\underline{\lambda_{min}}$
  - D.: special cases, Tricomi/Bessel/hypegeometric function asymptotics.
  - ?

β-Wishart, MANOVA ($\Sigma \neq I_n$)

- β-Wishart
  - CDFs derived in Koev et al.
  - No asymptotics.
  - In special cases (spiked model); Bloemendal and Virag, Ramirez and Rider.
- β-MANOVA
  - CDF for λ_{max} derived in Dubbs and Edelman.
  - No asymptotics.

A quick demonstration for λ_{min}

Start off with the eigenvalue pdf ($\lambda_1 \geq \ldots \geq \lambda_m$):

$$\tilde{f}(\lambda_1, \ldots, \lambda_n) \propto \prod_{i=1}^{m} \lambda_i^{\frac{\beta}{2}(a+1)-1} (1-\lambda_i)^{\frac{\beta}{2}(b+1)-1} \Delta^{\beta}(\lambda_1, \ldots, \lambda_m) \, ,$$

then integrate out all but the first and get (with $\lambda = \lambda_m$ and $d\lambda = d\lambda_1 \ldots d\lambda_{m-1}$):

$$\begin{aligned} f(\lambda) \quad \propto \quad & \lambda^{\frac{\beta}{2}(a+1)-1} (1-\lambda)^{\frac{\beta}{2}(b+1)-1} \times \\ & \int_{[\lambda,1]^{m-1}} \prod_{i=1}^{m-1} \lambda_i^{\frac{\beta}{2}(a+1)-1} (1-\lambda_i)^{\frac{\beta}{2}(b+1)-1} \, (\lambda_i - \lambda)^{\beta} \, \Delta^{\beta}(\lambda_1, \ldots, \lambda_{m-1}) \, d\lambda \, . \end{aligned}$$

A quick demonstration for λ_{min}

Changing variables to $x_i = \frac{1-\lambda_i}{1-\lambda}$, mapping $[\lambda, 1]$ to $[0, 1]$, we get

$$\begin{aligned} f(\lambda) \quad \propto \quad & \lambda^{\frac{\beta}{2}(a+1)-1}(1-\lambda)^{\frac{\beta}{2}(b+1)-1} \times \\ & \int_{[0,1]^{m-1}} \prod_{i=1}^{m-1} x_i^{\frac{\beta}{2}(b+1)-1} (1-x_i)^{\beta} (1-x_i(1-\lambda))^{\frac{\beta}{2}(a+1)-1} \Delta^{\beta}(x_1, \ldots, x_{m-1})\, dx \,. \end{aligned}$$

Crucially, following Forrester,

$$\begin{aligned} & \int_{[0,1]^{m-1}} \prod_{i=1}^{m-1} x_i^{\frac{\beta}{2}(b+1)-1} (1-x_i)^{\beta} (1-x_i(1-\lambda))^{\frac{\beta}{2}(a+1)-1} \Delta^{\beta}(x_1, \ldots, x_{m-1})\, dx \quad = \\ & {}_2F_1{}^{2/\beta}\left(1-\frac{\beta}{2}(a+1), \frac{\beta}{2}(b+m-1); \frac{\beta}{2}(b+m-1)+1; (1-\lambda)I_{m-1}\right) \;, \end{aligned}$$

A quick demonstration for λ_{min}

Therefore, thanks to the hypergeometric function, the pdf of λ_{min} is

$$\begin{aligned} f(\lambda) \quad &\propto \quad \lambda^{\frac{\beta}{2}(a+1)-1}(1-\lambda)^{\frac{\beta}{2}m(b+m)-1} \times \\ & {}_2F_1{}^{2/\beta}\left(1-\frac{\beta}{2}(a+1), \frac{\beta}{2}(b+m-1); \frac{\beta}{2}(b+m-1)+1; (1-\lambda)I_{m-1}\right) . \end{aligned}$$

As a corollary we can get the distribution of λ_{max} as well.

Why care?

Application: **RURV**, a randomized, efficient, communication-optimal, and very-likely-to-work way to find the *numerical* rank of a product of matrices and inverses. Part of a similarly bells-and-whistles Divide-and-Conquer algorithm for computing non-symmetric eigenvalues.

Why care?

Application: **RURV**, a randomized, **efficient**, communication-optimal, and very-likely-to-work way to find the *numerical* rank of a product of matrices and inverses. Part of a similarly bells-and-whistles Divide-and-Conquer algorithm for computing non-symmetric eigenvalues.

Efficiency: matrix multiplication exponent

How many ops involved in multiplying two $n \times n$ matrices?

Efficiency: matrix multiplication exponent

How many ops involved in multiplying two $n \times n$ matrices?
undergraduate student answer: $O(n^3)$

Efficiency: matrix multiplication exponent

How many ops involved in multiplying two $n \times n$ matrices?
undergraduate student answer: $O(n^3)$
graduate student answer: $O(n^\omega)$, where $\omega < 2.3729$

Efficiency: matrix multiplication exponent

How many ops involved in multiplying two $n \times n$ matrices?
undergraduate student answer: $O(n^3)$
graduate student answer: $O(n^\omega)$, where $\omega < 2.3729$
But are the results of the fast algorithm accurate?

Efficiency: matrix multiplication exponent

How many ops involved in multiplying two $n \times n$ matrices?
undergraduate student answer: $O(n^3)$
graduate student answer: $O(n^\omega)$, where $\omega < 2.3729$
But are the results of the fast algorithm accurate?
Demmel, D., Holtz: if the algorithm exists, we can make it stable.

Efficiency: rank-revealing algorithms

How many ops involved in rank-revealing?

All "serious" algorithms do at least one matrix multiplication, so at least $O(n^\omega)$.

Demmel, D. Holtz: **RURV** runs stably in $O(n^{\omega+\epsilon})$ for any ϵ.

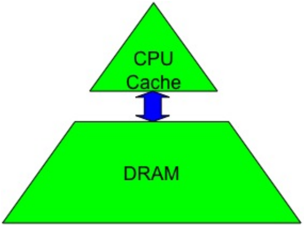
Why care?

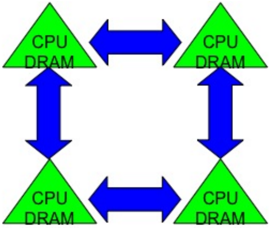
Application: **RURV**, a randomized, efficient, **communication-optimal**, and very-likely-to-work way to find the *numerical* rank of a product of matrices and inverses. Part of a similarly bells-and-whistles Divide-and-Conquer algorithm for computing non-symmetric eigenvalues.

Communication Cost Model

Algorithms have two costs:

1. arithmetic (flops)
2. communication: moving data between
   - levels of a memory hierarchy (sequential case)
   - processors over a network (parallel case)





Communication Cost Model

- Running time of an algorithm is sum of 3 terms:
  - # flops * time per flop
  - # words moved / bandwidth
  - # messages * latency

Communication Cost Model

- Exponentially growing gaps between
  - Sequentially:
    time per flop $\ll$ 1 / network BW $\ll$ network latency
    improving 59% per year vs. 26% per year vs. 15% per year
  - In parallel:
    time per flop $\ll$ 1 / memory BW $\ll$ memory latency
    improving 59% per year vs. 23% per year vs. 5.5% per year
- Need to reorganize linear algebra to *avoid* communication (# words and # messages moved)

Limits and optimality

There is such a thing as minimal cost for algorithms (Ballard, Demmel, Holtz, Schwartz), and **RURV** is nearly cost-optimal (and worth it for large matrices).

RURV

A rank-revealing decomposition ($A = URV$ with U, V orthogonal/unitary and R upper triangular) that works on products of matrices and inverses, e.g. AB^{-1}, *without forming the inverse*.

RURV

Starting with a matrix A, generate a decomposition $A = URV$ with R upper triangular, U, V orthogonal/unitary.

- Generate a random Gaussian B.
- $[V, \hat{R}] = \mathrm{QR}(B)$ (generate a Haar orthogonal/unitary V).
- $\hat{A} = A \cdot V^H$
- $[U, R] = \mathrm{QR}(\hat{A})$.
- Output U, R, V.

Why not **QR** outright?

Because

- if numerical rank is small, unless one does pivoting, not guaranteed to work well
- recall we want it to work on products of matrices and inverses; how to do **QR** on that without doing the product?

Generalized RURV (GRURV)

Want to find a rank-revealing factorization for $A^{-1}B$, but only need the left invariant spaces for our applications.

- $[U_2, R_2, V]$ =RURV(B);
- $R_1 U_1$ =RQ$(U_2^H A)$,
- Output U_1.

Note that

$$A^{-1}B = (U_2 R_1 U_1)^{-1}(U_2 R_2 V) = U_1^H (R_1^{-1} R_2) V$$

and we only need U_1 for our applications.

Why it works

Theorem (Ballard, Demmel, D., Melgaard '16+)

GRURV computes the RURV for $A^{-1}B$ and it is backward stable.

Theorem (BDDM'16+)

RURV computes a strong rank-revealing decomposition for A and it is backward stable.

RURV is *strong*

Let A be of numerical rank k (with a large gap between σ_k and σ_{k+1}).

Pick a Haar matrix V and then do QR on AV^H to get U, R. Then $A = URV; R = \begin{bmatrix} R_{11} & R_{12} \\ & R_{22} \end{bmatrix}$ and the following

- $\sigma_{min}(R_{11})$ is a good approximation to σ_k
- $\sigma_{max}(R_{22})$ is a good approximation to σ_{k+1}
- $||R_{11}^{-1}R_{12}||$ is small

All this happens with probability $1 - \delta$; making δ smaller increases the arithmetic costs.

The analysis hinges on knowing the distribution of the smallest singular value of the $k \times k$ principal minor for the Haar matrix V.

The smallest singular value of a $k \times k$ minor of V

It is known (Collins '03,'05, Sutton '06) that a $k \times k$ principal minor of a Haar matrix has eigenvalues $\lambda_1, \ldots, \lambda_k$ distributed like the Jacobi ensembles:

$$f(\lambda_1, \ldots, \lambda_k) \propto \prod_{i=1}^{k} \lambda_i^{\beta/2-1}(1-\lambda_i)^{\beta(n-2k+1)/2-1} \prod_{i<j} |\lambda_i - \lambda_j|^{\beta} ,$$

where $\beta = 1, 2$ for real/complex.

The smallest singular value of a $k \times k$ minor of V

Theorem (D.)

The pdf of the smallest singular value for a Jacobi ensemble as above, $\beta = 2$, is

$$f_{k,n}(x) \propto x^{-1/2}(1-x)^{\frac{1}{2}k(n-k)-1} \, {}_2F_1\left(\frac{1}{2}(n-k-1), \frac{1}{2}(k-1); \frac{1}{2}(n-1)+1; (1-x)\right) .$$

How do we get usable formulae/asymptotics?

Recall that the pdf of λ_{min} is

$$
\begin{aligned}
f(\lambda) \quad &\propto \quad \lambda^{\frac{\beta}{2}(a+1)-1}(1-\lambda)^{\frac{\beta}{2}m(b+m)-1} \times \\
& \quad {}_2F_1{}^{2/\beta}\left(1-\frac{\beta}{2}(a+1), \frac{\beta}{2}(b+m-1); \frac{\beta}{2}(b+m-1)+1; (1-\lambda)I_{m-1}\right) .
\end{aligned}
$$

The issue here is the $(1-\lambda)$.

Making the hypergeometric a polynomial, or simple

$$ {}_2F_1{}^{2/\beta}\left(1-\frac{\beta}{2}(a+1), \frac{\beta}{2}(b+m-1); \frac{\beta}{2}(b+m-1)+1; (1-\lambda)I_{m-1}\right) $$

- If $\frac{\beta}{2}(a+1)-1 \in \mathbb{Z}_{\geq 0}$, series terminates. Kummer relationships (Forrester) allow you to use a slightly different formula for the hypergeometric integral, which can be analyzed asymptotically
- If $1-\frac{\beta}{2}(a+1)=\frac{\beta}{2}$, then the hypergeometric becomes a classical one.
- It stands to reason that there may be other cases that are analyzable; the problem is open.

Case 1: $\frac{\beta}{2}(a+1) - 1 = k \in \mathbb{Z}_{\geq 0}$

Can obtain the distribution of the smallest eigenvalue:

$$\begin{aligned} f_m(\lambda) \quad &\propto \quad \lambda^{k-1}(1-\lambda)^{\frac{\beta}{2}m(b+m)-1} \\ &\quad \times {}_2F_1{}^{4/\beta}\left(1-m, -m-b+1; 2+\frac{2}{\beta}(k-1); \{\lambda\}^{k-1}\right), \end{aligned}$$

Asymptotics: m fixed, $b \to \infty$; scale $y = (b+m)\lambda$ to get

$$f_m(y) \propto y^{k-1} e^{-\beta m y/2} {}_1F_1{}^{4/\beta}\left(1-m, 2+\frac{2}{\beta}(k-1); \{-y\}^{k-1}\right).$$

If $\beta = 2, k = 1$ (Haar unitary matrix!), get exactly $f_m(y) = me^{-my/2}$.

Case 1: $\frac{\beta}{2}(a+1) - 1 = k \in \mathbb{Z}_{\geq 0}$

Can obtain the distribution of the smallest eigenvalue:

$$\begin{aligned} f_m(\lambda) \quad &\propto \quad \lambda^{k-1}(1-\lambda)^{\frac{\beta}{2}m(b+m)-1} \\ & \quad \times {}_2F_1{}^{4/\beta}\left(1-m, -m-b+1; 2+\frac{2}{\beta}(k-1); \{\lambda\}^{k-1}\right), \end{aligned}$$

Asymptotics: $m, (b+m) \to \infty$; scale $y = m(b+m)\lambda$ to get

$$f(y) \propto y^{k-1} e^{-\beta y/2} {}_0F_1{}^{4/\beta}\left(2 + \frac{2}{\beta}(k-1); y^{k-1}\right).$$

If $\beta = 2, k = 1$ (Haar unitary matrix!), get exactly $f(y) = e^{-y}$.

Case 2: $a = \frac{2}{\beta} - 2$

After a bit of manipulation, can obtain that

$$\begin{aligned} f_{\beta,b,m}(\lambda) \quad &= \quad \tilde{C}_{\beta,b,m}\, \lambda^{-\beta/2}\, (1-\lambda)^{\beta m(b+m)/2-1} \\ &\times \Bigg(\frac{1}{\Gamma(-\frac{\beta}{2})\Gamma(\frac{\beta m}{2}+1)}\, {}_2F_1\Big(\frac{\beta(b+m-1)}{2}, \frac{\beta(m-1)}{2}; -\frac{\beta}{2}; \lambda\Big) \\ &\quad - \lambda^{1+\beta/2} \frac{1}{\Gamma(\frac{\beta}{2}+2)\Gamma(\frac{\beta(m-1)}{2})}\, \frac{\Gamma(\frac{\beta(b+m)}{2}+1)}{\Gamma(\frac{\beta(b+m-1)}{2})}\, {}_2F_1\Big(\frac{\beta m}{2}+1, \frac{\beta(b+m)}{2}+1; 2+\frac{\beta}{2}; \lambda\Big) \Bigg) \end{aligned}$$

Asymptotics: m fixed, $b \to \infty$; scale $y = (b+m)\lambda$ to get

$$f_m(y) \propto y^{-\beta/2} e^{-my}\, U\left(\frac{\beta}{2}(m-1); -\frac{\beta}{2}; y\right) ,$$

with U the Tricomi function.

Case 2: $a = \frac{2}{\beta} - 2$

After a bit of manipulation, can obtain that

$$
\begin{aligned}
f_{\beta,b,m}(\lambda) \quad = \quad & \tilde{C}_{\beta,b,m}\, \lambda^{-\beta/2}\,(1-\lambda)^{\beta m(b+m)/2-1} \\
& \times \left(\frac{1}{\Gamma(-\frac{\beta}{2})\Gamma(\frac{\beta m}{2}+1)}\, {}_2F_1\!\left(\frac{\beta(b+m-1)}{2}, \frac{\beta(m-1)}{2}; -\frac{\beta}{2}; \lambda\right) \right. \\
& \left. - \lambda^{1+\beta/2} \frac{1}{\Gamma(\frac{\beta}{2}+2)\Gamma(\frac{\beta(m-1)}{2})} \, \frac{\Gamma(\frac{\beta(b+m)}{2}+1)}{\Gamma(\frac{\beta(b+m-1)}{2})}\, {}_2F_1\!\left(\frac{\beta m}{2}+1, \frac{\beta(b+m)}{2}+1; 2+\frac{\beta}{2}; \lambda\right) \right)
\end{aligned}
$$

Asymptotics: m fixed, $b+m \to \infty$; scale $y = \beta m(b+m)\lambda/2$ to get

$$
f(y) \propto y^{\frac{1}{2}-\beta/4} e^{-y} K_{1+\frac{\beta}{2}}(\sqrt{2\beta y}) ,
$$

with K the modified Bessel function. This corresponds to complex Haar and (if wanted) quaternion Haar matrices.

Pretty pictures

The following tests were made possible by the cool multivariate hypergeometric package **mgh**, by Plamen Koev; and also by the β-Jacobi tridiagonal model due to Brian Sutton and Alan Edelman.

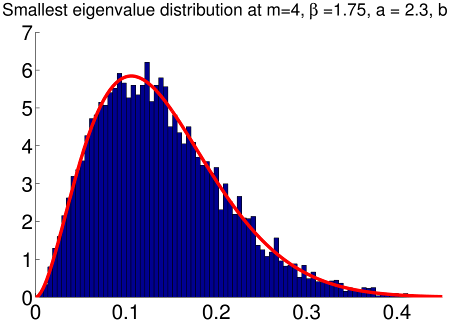


Figure: The solid red line represents the theoretical distribution; the normalized histogram represents the results of a Monte Carlo experiment with 10,000 trials.

Pretty pictures

The following tests were made possible by the cool multivariate hypergeometric package **mgh**, by Plamen Koev; and also by the β-Jacobi tridiagonal model due to Brian Sutton and Alan Edelman.

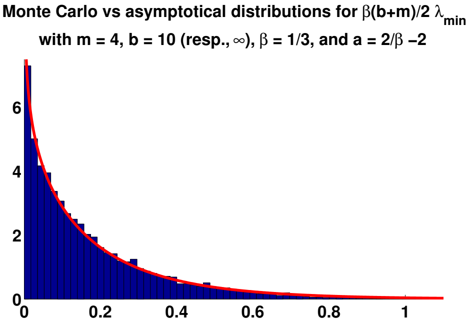


Figure: The solid red line represents the asymptotical ($b = \infty$) distribution, while the normalized histogram represents the results of a Monte Carlo experiment for $b = 10$, with 10, 000 trials.

Pretty pictures

The following tests were made possible by the cool multivariate hypergeometric package **mgh**, by Plamen Koev; and also by the β-Jacobi tridiagonal model due to Brian Sutton and Alan Edelman.

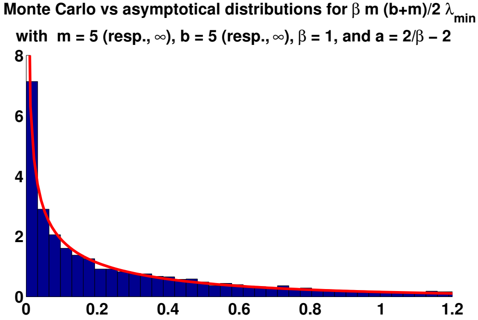


Figure: The solid red line represents the asymptotical $(m, b = \infty)$ distribution, while the normalized histogram represents the results of a Monte Carlo experiment for $m = 5$, $b = 5$, with $5,000$ trials.

Pretty pictures

The following tests were made possible by the cool multivariate hypergeometric package **mgh**, by Plamen Koev; and also by the β-Jacobi tridiagonal model due to Brian Sutton and Alan Edelman.

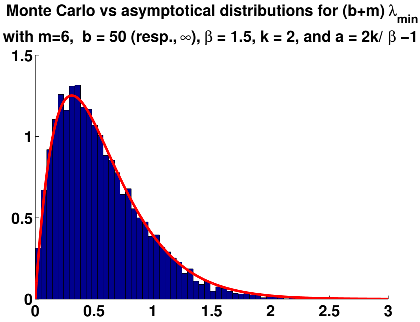


Figure: The solid red line represents the asymptotical ($b = \infty$) distribution, while the normalized histogram represents the results of a Monte Carlo experiment for $b = 50$, with $10,000$ trials.

Pretty pictures

The following tests were made possible by the cool multivariate hypergeometric package **mgh**, by Plamen Koev; and also by the β-Jacobi tridiagonal model due to Brian Sutton and Alan Edelman.

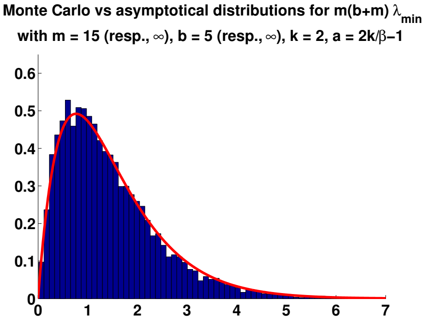


Figure: The solid red line represents the asymptotical $(m, b = \infty)$ distribution, while the normalized histogram represents the results of a Monte Carlo experiment for $m = 15$, $b = 50$, with $10,000$ trials.

What to take home

- Still plenty of problems in computing extremal eigenvalue distributions, either for n fixed or asymptotically
- Hypergeometric functions are cool, but slightly unsatisfying; computable (but not for very large matrix sizes); work well in only some cases; more to uncover
- RMT has unexpected and interesting applications in scientific computing
- There's a world full of *potential* out-there.