Distances between Random Orthogonal Matrices and Independent Normals

Tiefeng Jiang

School of Statistics, University of Minnesota

Joint with Yutao Ma (Beijing Normal University)

July 19, 2017

Outline

- What are Haar-orthogonal matrices?
- Observations of independence
- Questions and solutions
- Application
- Proof

What are Haar-orthogonal matrices?

Mathematically,
$\Gamma_n \sim$ Haar measure on O_n

Statistically,
Let $Y = (y_{ij})_{n \times n}$ where y_{ij}'s are independent $N(0, 1)$.

Two ways to generate such matrices.

What are Haar-orthogonal matrices?

Mathematically,
$\Gamma_n \sim$ Haar measure on O_n

Statistically,
Let $Y = (y_{ij})_{n\times n}$ where y_{ij}'s are independent $N(0,1)$.

Two ways to generate such matrices.

- Orthogonal matrix by Gram-Schmidt algorithm on columns of Y;
- $\Gamma_n \stackrel{d}{=} \mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1/2}$.

Any row or column of Γ_n has uniform distribution on S^{n-1}

Observations of independence

$$\Gamma_n = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1q} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2q} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots & & \\ \gamma_{p1} & \gamma_{p2} & \cdots & \gamma_{pq} & \cdots & \gamma_{pn} \\ \vdots & \vdots & \ddots & \vdots & & \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nq} & \cdots & \gamma_{nn} \end{pmatrix}$$

Observation: When n is large, γ_{ij}'s are *roughly* independent $N(0, \frac{1}{n})$

1) Borel (1914): Kinetic theory of gas;

$P(\sqrt{n}\gamma_{11} \leq x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$

1) Borel (1914): Kinetic theory of gas;
$P(\sqrt{n}\gamma_{11} \leq x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$

2) Stam (1982): Geometric Probability;
$y_1, y_2, \cdots, y_n$: indpendent $N(0, 1)$. Then

$$\|\mathcal{L}(\sqrt{n}(\gamma_{11}, \gamma_{21}, \cdots, \gamma_{m1})) - \mathcal{L}(y_1, y_2, \cdots, y_m)\| \to 0$$

if $m = o(\sqrt{n})$, where $\|\cdot\|$ is variation distance:
$\|\mu - \nu\| = \sup_A |\mu(A) - \nu(A)|$

1) Borel (1914): Kinetic theory of gas;
$P(\sqrt{n}\gamma_{11} \leq x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$

2) Stam (1982): Geometric Probability;
$y_1, y_2, \cdots, y_n$: indpendent $N(0,1)$. Then

$$\|\mathcal{L}(\sqrt{n}(\gamma_{11}, \gamma_{21}, \cdots, \gamma_{m1})) - \mathcal{L}(y_1, y_2, \cdots, y_m)\| \to 0$$

if $m = o(\sqrt{n})$, where $\|\cdot\|$ is variation distance:
$\|\mu - \nu\| = \sup_A |\mu(A) - \nu(A)|$

3) Diaconis and Freedman (1987): Finite de Finetti Theorem;
$m = o(n)$

4) Diaconis, Eaton and Lauritzen (1992)

Z_n : upper-left $p \times q$ block of Γ_n :

$$\Gamma_n = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1q} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2q} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots & & \\ \gamma_{p1} & \gamma_{p2} & \cdots & \gamma_{pq} & \cdots & \gamma_{pn} \\ \vdots & \vdots & \ddots & \vdots & & \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nq} & \cdots & \gamma_{nn} \end{pmatrix}$$

$G_n = (g_{ij})_{p\times q}$, where g_{ij} are independent $N(0, 1)$. Then

4) Diaconis, Eaton and Lauritzen (1992)

Z_n : upper-left $p \times q$ block of Γ_n :

$$\Gamma_n = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1q} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2q} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots & & \\ \gamma_{p1} & \gamma_{p2} & \cdots & \gamma_{pq} & \cdots & \gamma_{pn} \\ \vdots & \vdots & \ddots & \vdots & & \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nq} & \cdots & \gamma_{nn} \end{pmatrix}$$

$G_n = (g_{ij})_{p\times q}$, where g_{ij} are independent $N(0,1)$. Then

$$\|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$$

if $p = o(n^\alpha)$ and $q = o(n^\alpha)$ for $\alpha = 1/3$.

Open problem and solution

Open problem by Diaconis (Bulletin of AMS, 2003):

What is largest α *s.t. above holds?*

Open problem and solution

Open problem by Diaconis (Bulletin of AMS, 2003):

What is largest α s.t. above holds?

Speculation: $O(n^{1/3}),\ o(n^{1/2}),\ O(\frac{n}{\log n})$.

Open problem and solution

Open problem by Diaconis (Bulletin of AMS, 2003):

What is largest α *s.t. above holds?*

Speculation: $O(n^{1/3}),\ o(n^{1/2}),\ O(\frac{n}{\log n})$.

Solution by J. (Ann. Probab, 2006): $o(n^{1/2})$

Solution by J. (2006)

- Z_n : upper-left $p \times q$ block of Γ_n.
- G_n : $p \times q$ matrix with indept $N(0,1)$ entries.

The following holds:

Solution by J. (2006)

- Z_n : upper-left $p \times q$ block of Γ_n.
- G_n : $p \times q$ matrix with indept $N(0,1)$ entries.

The following holds:

(1) If $p_n = o(\sqrt{n})$ and $q_n = o(\sqrt{n})$, then $\|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$.

Solution by J. (2006)

- Z_n : upper-left $p \times q$ block of Γ_n.
- G_n : $p \times q$ matrix with indept $N(0,1)$ entries.

The following holds:

(1) If $p_n = o(\sqrt{n})$ and $q_n = o(\sqrt{n})$, then $\|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$.

(2) $p_n = [xn^{1/2}]$ and $q_n = [yn^{1/2}]$. Then

$$\liminf_{n\to\infty} \|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \geq \phi(x,y) > 0.$$

New Questions

(1) If $pq/n \to 0$, p and q not same scale, $\|\mathcal{L}\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$?

New Questions

(1) If $pq/n \to 0$, p and q not same scale, $\|\mathcal{L}\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$?

New technology on $p/q \to 0$ is needed, a feature of Big Data

New Questions

(1) If $pq/n \to 0$, p and q not same scale, $\|\mathcal{L}\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$?

New technology on $p/q \to 0$ is needed, a feature of Big Data

(2) What if $\|\cdot\|$ is replaced by
Hellinger, Kullback-Leibler or Euclidean distances?

New Questions

(1) If $pq/n \to 0$, p and q not same scale, $\|\mathcal{L}\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$?

New technology on $p/q \to 0$ is needed, a feature of Big Data

(2) What if $\|\cdot\|$ is replaced by
Hellinger, Kullback-Leibler or Euclidean distances?

Distances

μ and ν have density f and g, respectively.

- Hellinger distance $H(\mu, \nu)$

$$H^2(\mu, \nu) = \frac{1}{2} \int_{\mathbb{R}^m} |\sqrt{f(x)} - \sqrt{g(x)}|^2 dx.$$

Distances

μ and ν have density f and g, respectively.

- Hellinger distance $H(\mu, \nu)$

$$H^2(\mu, \nu) = \frac{1}{2} \int_{\mathbb{R}^m} |\sqrt{f(x)} - \sqrt{g(x)}\,|^2 dx.$$

- Kullback-Leibler distance:

$$D_{\mathrm{KL}}(\mu, \nu) = \int_{\mathbb{R}^m} \frac{d\mu}{d\nu} \log \frac{d\mu}{d\nu} d\nu.$$

- Hilbert-Schmidt norm: $\mathbf{A} = (a_{ij})$ has norm $\|\mathbf{A}\|_{HS} = (\sum |a_{ij}|^2)^{1/2}$

Our Result

Theorem

Let $d(\sqrt{n}\mathbf{Z}_n, \mathbf{G}_n)$ *be total variation, Hellinger or Kullback-Leibler distance between* $\sqrt{n}\mathbf{Z}_n$ *and* $\mathbf{G}_n$*. Then*

Our Result

Theorem

Let $d(\sqrt{n}\mathbf{Z}_n, \mathbf{G}_n)$ be total variation, Hellinger or Kullback-Leibler distance between $\sqrt{n}\mathbf{Z}_n$ and $\mathbf{G}_n$. Then

- $\lim_{n\to\infty} d(\sqrt{n}\mathbf{Z}_n, \mathbf{G}_n) = 0$ *for any* $p \geq 1$ *and* $q \geq 1$ *with* $\lim_{n\to\infty} \frac{pq}{n} = 0$;

Our Result

Theorem

Let $d(\sqrt{n}\mathbf{Z}_n, \mathbf{G}_n)$ be total variation, Hellinger or Kullback-Leibler distance between $\sqrt{n}\mathbf{Z}_n$ and $\mathbf{G}_n$. Then

- $\lim_{n\to\infty} d(\sqrt{n}\mathbf{Z}_n, \mathbf{G}_n) = 0$ *for any* $p \geq 1$ *and* $q \geq 1$ *with* $\lim_{n\to\infty} \frac{pq}{n} = 0$;
- $\liminf_{n\to\infty} d(\sqrt{n}\mathbf{Z}_n, \mathbf{G}_n) > 0$ *if* $\lim_{n\to\infty} \frac{pq}{n} = \sigma \in (0, \infty)$.

Let $\mathbf{Y}_n = (\mathbf{y}_1, \cdots, \mathbf{y}_n) = (y_{ij})_{n\times n},\ y_{ij}$'s are indept $N(0, 1)$.

Let $\mathbf{Y}_n = (\mathbf{y}_1, \cdots, \mathbf{y}_n) = (y_{ij})_{n\times n}$, y_{ij}'s are indept $N(0, 1)$.

Perform Gram-Schmidt algorithm on $\mathbf{y}_1, \cdots, \mathbf{y}_n$.

Let $\mathbf{Y}_n = (\mathbf{y}_1, \cdots, \mathbf{y}_n) = (y_{ij})_{n\times n}$, y_{ij}'s are indept $N(0, 1)$.

Perform Gram-Schmidt algorithm on $\mathbf{y}_1, \cdots, \mathbf{y}_n$.

We then get $\boldsymbol{\Gamma}_n = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_n) = (\gamma_{ij})_{n\times n}$, an Haar-invariant orthogonal matrix.

Let $\mathbf{Y}_n = (\mathbf{y}_1, \cdots, \mathbf{y}_n) = (y_{ij})_{n\times n}$, y_{ij}'s are indept $N(0,1)$.

Perform Gram-Schmidt algorithm on $\mathbf{y}_1, \cdots, \mathbf{y}_n$.

We then get $\mathbf{\Gamma}_n = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_n) = (\gamma_{ij})_{n\times n}$, an Haar-invariant orthogonal matrix.

Set $\mathbf{\Gamma}_{p\times q} = (\gamma_{ij})_{1\le i\le p, 1\le j\le q}$ and $\mathbf{Y}_{p\times q} = (y_{ij})_{1\le i\le p, 1\le j\le q}$. Then

$$\|\sqrt{n}\mathbf{\Gamma}_{p\times q} - \mathbf{Y}_{p\times q}\|_{\text{HS}}^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}(\sqrt{n}\gamma_{ij} - y_{ij})^2.$$

Theorem

- *If* $p = p_n, q = q_n$ *satisfy* $1 \leq p, q \leq n$ *and* $\lim_{n\to\infty} \frac{pq^2}{n} = 0$*, then* $\|\sqrt{n}\mathbf{\Gamma}_{p\times q} - \mathbf{Y}_{p\times q}\|_{\mathrm{HS}} \xrightarrow{p} 0$.

Theorem

- *If* $p = p_n, q = q_n$ *satisfy* $1 \le p, q \le n$ *and* $\lim_{n\to\infty} \frac{pq^2}{n} = 0$, *then* $\|\sqrt{n}\mathbf{\Gamma}_{p\times q} - \mathbf{Y}_{p\times q}\|_{\mathrm{HS}} \xrightarrow{p} 0.$

- *If* $1 \le p, q \le n$ *satisfy* $\lim_{n\to\infty} \frac{pq^2}{n} = \sigma \in (0, \infty)$, *then* $\|\sqrt{n}\mathbf{\Gamma}_{p\times q} - \mathbf{Y}_{p\times q}\|_{\mathrm{HS}} \overset{p}{\nrightarrow} 0$

Summary

distance d	order of (p, q)
total variation	$(\sqrt{n}, \sqrt{n})$
Hellinger	$(\sqrt{n}, \sqrt{n})$
Kullback-Leibler	$(\sqrt{n}, \sqrt{n})$
Euclidean	$(\sqrt[3]{n}, \sqrt[3]{n})$
weak	$(n, \frac{n}{\log n})$

Summary

distance d	order of (p, q)
total variation	$(\sqrt{n}, \sqrt{n})$
Hellinger	$(\sqrt{n}, \sqrt{n})$
Kullback-Leibler	$(\sqrt{n}, \sqrt{n})$
Euclidean	$(\sqrt[3]{n}, \sqrt[3]{n})$
weak	$(n, \frac{n}{\log n})$

Weak norm: $\|\mathbf{A}\|_{max} = \max_{1\le i\le n, 1\le j\le q} |a_{ij}|$ if $\mathbf{A} = (a_{ij})_{n\times q}$

$\|\sqrt{n}\mathbf{\Gamma}_{n\times q} - \mathbf{Y}_{n\times q}\|_{max} \to 0$ if $q = o(\frac{n}{\log n})$;
it doesn't if $q = c \cdot \frac{n}{\log n}$.

Applications

1. *Random matrices*: Jacobi/MANOVA/Matrix beta dist. is classical matrix
(Dyson's 3-fold classification: Hermite, Laguerre, Jacobi ensemble)

Applications

1. *Random matrices*: Jacobi/MANOVA/Matrix beta dist. is classical matrix
(Dyson's 3-fold classification: Hermite, Laguerre, Jacobi ensemble)

J. (2009):
• Largest eigenvalue is asymptotic Tracy-Widom law for extreme case.
• Empirical dist. of all eigenvalues is asymptotic Machenko-Pastur law.
• Eigenvalues of square truncation $\Gamma_{p\times p}$ empirically follow circular law.

2. *Big Data*: Data storage by Chen K. and Liu, L. (2011)

3. Wireless communications:
- Li, B., Kumar, H. and Petropulu, A. P. (2016)
- Li, Y., Nguyen, H. L. and Woodruff, D. P. (2014)
- Li, B. and Petropulu, A. (2016)

Comments

• Similar work are true for unitary group, symplectic group

• These are connected to Riemannian symmetric space (Cartan's classification)

• There are many probab. distances. What are about them?

Proof for Total variation dist.

$f_1(z)$: joint density of $\sqrt{n}Z_n$.
$g(z)$: joint density of pq i.i.d. $N(0,1)$. Then

$$\begin{aligned}\|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\|_{TV} &= \int_{R^{pq}} |f_1(z) - g(z)|\, dz \\ &= \int_{R^{pq}} \left|\frac{f_1(z)}{g(z)} - 1\right| g(z)\, dz \\ &= E\left|\frac{f_1(Z)}{g(Z)} - 1\right|,\end{aligned}$$

where entries of Z are i.i.d. $N(0,1)$. Work on Wishart matrices

Similarly,

$$\begin{aligned} \|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\| &= \int_{R^{pq}} |f_1(z) - g(z)|\, dz \\ &= E\left|\frac{g(Z)}{f_1(Z)} - 1\right|, \end{aligned}$$

where Z is the $p \times q$ block of Haar-orthogonal matrix.

Lemma by D,E,L: Joint density of entries of Z is $f(z) =$

$$(\sqrt{2\pi})^{-pq}\frac{\omega(n-p,q)}{\omega(n,q)}\left\{\prod_{i=1}^{q}(1-\lambda_i)^{(n-p-q-1)/2}\right\}$$

if λ_i, eigenvalues of $z'z$, are in $(0,1)$, where $\omega(\cdot,\cdot)$ is

$$\frac{1}{\omega(r,s)} = \pi^{s(s-1)/4}2^{rs/2}\prod_{j=1}^{s}\Gamma\left(\frac{r-j+1}{2}\right).$$

Lemma by D,E,L: Joint density of entries of Z is $f(z) =$

$$(\sqrt{2\pi})^{-pq}\frac{\omega(n-p,q)}{\omega(n,q)}\left\{\prod_{i=1}^{q}(1-\lambda_i)^{(n-p-q-1)/2}\right\}$$

if λ_i, eigenvalues of $z'z$, are in $(0,1)$, where $\omega(\cdot,\cdot)$ is

$$\frac{1}{\omega(r,s)} = \pi^{s(s-1)/4}2^{rs/2}\prod_{j=1}^{s}\Gamma\left(\frac{r-j+1}{2}\right).$$

Key steps:

- CLT for $(tr(Z'Z), tr((Z'Z)^2)\ )$ by Jonsson (1982); J. and Li (2015).
- Expansion of Gamma function
- subsequence argument on $p_n \equiv c$, $p_n/q_n \to c$ and $p_n/q_n \to \infty$.
- Pinsker ineq: connections among TV, Hellinger and KL

To compute trace of submatrix, we need to compute following:

Theorem

Let $\mathbf{\Gamma}_n = (\gamma_{ij})_{n\times n}$ *be Haar-orthogonal matrix. Then*

a) $\mathbb{E}(\gamma_{11}^2\gamma_{21}^2\gamma_{22}^2) = \frac{1}{(n-1)n(n+2)} - \frac{3}{(n-1)n(n+2)(n+4)}$.

b) $\mathbb{E}(\gamma_{11}\gamma_{12}\gamma_{22}\gamma_{23}\gamma_{31}\gamma_{33}) = \frac{2}{(n-2)(n-1)n(n+2)(n+4)}$.

To compute trace of submatrix, we need to compute following:

Theorem

Let $\mathbf{\Gamma}_n = (\gamma_{ij})_{n\times n}$ *be Haar-orthogonal matrix. Then*

a) $\mathbb{E}(\gamma_{11}^2\gamma_{21}^2\gamma_{22}^2) = \frac{1}{(n-1)n(n+2)} - \frac{3}{(n-1)n(n+2)(n+4)}$.

b) $\mathbb{E}(\gamma_{11}\gamma_{12}\gamma_{22}\gamma_{23}\gamma_{31}\gamma_{33}) = \frac{2}{(n-2)(n-1)n(n+2)(n+4)}$.

c) $\cdots$
d) $\cdots$
e) $\cdots$

Thank you for attending the talk!