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,
\[ \Gamma_n \sim \text{Haar measure on } O_n \]

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,
\[ \Gamma_n \sim \text{Haar measure on } O_n \]

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 \( Y \);
- \( \Gamma_n \overset{d}{=} Y(Y^T Y)^{-1/2} \).

Any row or column of \( \Gamma_n \) has uniform distribution on \( S^{n-1} \)
Observations of independence

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

Observation: When \( n \) is large, \( \gamma_{ij} \)'s are \textit{roughly} independent \( N(0, \frac{1}{n}) \)
1) Borel (1914): Kinetic theory of gas;
\( P(\sqrt{n}\gamma_{11} \leq x) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt \)
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;
\( y_1, y_2, \ldots, y_n \): independent \( N(0, 1) \). Then
\[
\| \mathcal{L}(\sqrt{n}(\gamma_{11}, \gamma_{21}, \ldots, \gamma_{m1})) - \mathcal{L}(y_1, y_2, \ldots, y_m) \| \to 0
\]
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: independent \ 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:

\[ \| \mu - \nu \| = \sup_A |\mu(A) - \nu(A)| \]

m = o(n)
4) Diaconis, Eaton and Lauritzen (1992)

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

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

$G_n = (g_{ij})_{p \times q}$, where $g_{ij}$ are independent $N(0, 1)$. Then
4) Diaconis, Eaton and Lauritzen (1992)

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

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

$G_n = (g_{ij})_{p \times q}$, 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 $\alpha$ s.t. above holds?*
Open problem by Diaconis (Bulletin of AMS, 2003):

What is largest $\alpha$ s.t. above holds?

Speculation: $O(n^{1/3})$, $o(n^{1/2})$, $O(\frac{n}{\log n})$. 
Open problem by Diaconis (Bulletin of AMS, 2003):

What is largest $\alpha$ s.t. above holds?

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

Solution by J. (Ann. Probab, 2006): $o(n^{1/2})$
• $Z_n$: upper-left $p \times q$ block of $\Gamma_n$.
• $G_n$: $p \times q$ matrix with indept $N(0, 1)$ entries.

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

The following holds:

(1) If $p_n = o(\sqrt{n})$ and $q_n = o(\sqrt{n})$, then $\|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$. 
• $Z_n$: upper-left $p \times q$ block of $\Gamma_n$.
• $G_n$: $p \times q$ 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) $p_n = \lfloor xn^{1/2} \rfloor$ and $q_n = \lfloor yn^{1/2} \rfloor$. 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, $\|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\| \to 0$?
(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*
(1) If \( \frac{pq}{n} \to 0 \), \( p \) 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?
(1) If $pq/n \to 0$, $p$ and $q$ not same scale, $\|L\sqrt{n}Z_n) - L(G_n)\| \to 0$?

*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

$\mu$ and $\nu$ 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

\( \mu \) and \( \nu \) 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:

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

- Hilbert-Schmidt norm: \( A = (a_{ij}) \) has norm

\[
\|A\|_{HS} = (\sum |a_{ij}|^2)^{1/2}
\]
Our Result

**Theorem**

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

- $\lim_{n \to \infty} d(\sqrt{n}Z_n, G_n) = 0$ for any $p \geq 1$ and $q \geq 1$ with $\lim_{n \to \infty} pq^n = 0$;
- $\liminf_{n \to \infty} d(\sqrt{n}Z_n, G_n) > 0$ if $\lim_{n \to \infty} pq^n = \sigma \in (0, \infty)$. 
Our Result

Theorem

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

- $\lim_{n \to \infty} d(\sqrt{n}Z_n, G_n) = 0$ for any $p \geq 1$ and $q \geq 1$ with $\lim_{n \to \infty} pq/n = 0$;
Our Result

Theorem

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

- $\lim_{n \to \infty} d(\sqrt{n}Z_n, G_n) = 0$ for any $p \geq 1$ and $q \geq 1$ with $\lim_{n \to \infty} \frac{pq}{n} = 0$;
- $\liminf_{n \to \infty} d(\sqrt{n}Z_n, G_n) > 0$ if $\lim_{n \to \infty} \frac{pq}{n} = \sigma \in (0, \infty)$. 
Let $Y_n = (y_1, \cdots, y_n) = (y_{ij})_{n \times n}$, $y_{ij}$'s are indept $N(0, 1)$. 

Perform Gram-Schmidt algorithm on $y_1, \cdots, y_n$. We then get $\Gamma_n = (\gamma_1, \cdots, \gamma_n) = (\gamma_{ij})_{n \times n}$, an Haar-invariant orthogonal matrix.

Set $\Gamma_{p \times q} = (\gamma_{ij})_{1 \leq i \leq p, 1 \leq j \leq q}$ and $Y_{p \times q} = (y_{ij})_{1 \leq i \leq p, 1 \leq j \leq q}$. Then

$$\|\sqrt{n} \Gamma_{p \times q} - Y_{p \times q}\|_{2 \text{HS}} = \sum_{i=1}^{p} \sum_{j=1}^{q} (\sqrt{n} \gamma_{ij} - y_{ij})^2.$$
Let $Y_n = (y_1, \cdots, y_n) = (y_{ij})_{n \times n}$, $y_{ij}$’s are indept $N(0, 1)$.

Perform Gram-Schmidt algorithm on $y_1, \cdots, y_n$. 

Let $Y_n = (y_1, \cdots, y_n) = (y_{ij})_{n \times n}$, $y_{ij}$'s are indept $N(0, 1)$.

Perform Gram-Schmidt algorithm on $y_1, \cdots, y_n$.

We then get $\Gamma_n = (\gamma_1, \cdots, \gamma_n) = (\gamma_{ij})_{n \times n}$, an Haar-invariant orthogonal matrix.
Let $Y_n = (y_1, \cdots, y_n) = (y_{ij})_{n \times n}$, $y_{ij}$’s are indep $N(0, 1)$.

Perform Gram-Schmidt algorithm on $y_1, \cdots, y_n$.

We then get $\Gamma_n = (\gamma_1, \cdots, \gamma_n) = (\gamma_{ij})_{n \times n}$, an Haar-invariant orthogonal matrix.

Set $\Gamma_{p \times q} = (\gamma_{ij})_{1 \leq i \leq p, 1 \leq j \leq q}$ and $Y_{p \times q} = (y_{ij})_{1 \leq i \leq p, 1 \leq j \leq q}$. Then

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

\[
\| \sqrt{n} \Gamma_{p \times q} - Y_{p \times q} \|_{HS} \xrightarrow{p} 0.
\]
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
\[
\| \sqrt{n} \Gamma_{p \times q} - Y_{p \times q} \|_{HS} \overset{p}{\to} 0.
\]

• If \( 1 \leq p, q \leq n \) satisfy \( \lim_{n \to \infty} \frac{pq^2}{n} = \sigma \in (0, \infty) \), then
\[
\| \sqrt{n} \Gamma_{p \times q} - Y_{p \times q} \|_{HS} \nrightarrow 0
\]
<table>
<thead>
<tr>
<th>distance $d$</th>
<th>order of $(p, q)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>total variation</td>
<td>$(\sqrt{n}, \sqrt{n})$</td>
</tr>
<tr>
<td>Hellinger</td>
<td>$(\sqrt{n}, \sqrt{n})$</td>
</tr>
<tr>
<td>Kullback-Leibler</td>
<td>$(\sqrt{n}, \sqrt{n})$</td>
</tr>
<tr>
<td>Euclidean</td>
<td>$(\sqrt[3]{n}, \sqrt[3]{n})$</td>
</tr>
<tr>
<td>weak</td>
<td>$(n, \frac{n}{\log n})$</td>
</tr>
<tr>
<td>distance $d$</td>
<td>order of $(p, q)$</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>total variation</td>
<td>$(\sqrt{n}, \sqrt{n})$</td>
</tr>
<tr>
<td>Hellinger</td>
<td>$(\sqrt{n}, \sqrt{n})$</td>
</tr>
<tr>
<td>Kullback-Leibler</td>
<td>$(\sqrt{n}, \sqrt{n})$</td>
</tr>
<tr>
<td>Euclidean</td>
<td>$(\sqrt[3]{n}, \sqrt[3]{n})$</td>
</tr>
<tr>
<td>weak</td>
<td>$(n, \frac{n}{\log n})$</td>
</tr>
</tbody>
</table>

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

$$\|\sqrt{n}\mathbf{T}_{n \times q} - \mathbf{Y}_{n \times q}\|_{max} \to 0 \text{ if } q = o\left(\frac{n}{\log n}\right);$$

it doesn’t if $q = c \cdot \frac{n}{\log n}$.  

---

**Summary**

- Distance $d$ and the 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: $(n, \frac{n}{\log n})$
1. *Random matrices*: Jacobi/MANOVA/Matrix beta dist. is classical matrix
*(Dyson’s 3-fold classification: Hermite, Laguerre, Jacobi ensemble)*
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.

3. Wireless communications:
   - Li, B., Kumar, H. and Petropulu, A. P. (2016)
   - Li, B. and Petropulu, A. (2016)
• 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

$$\|\mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n)\|_{TV} = \int_{\mathbb{R}^{pq}} |f_1(z) - g(z)| \, dz$$

$$= \int_{\mathbb{R}^{pq}} \left| \frac{f_1(z)}{g(z)} - 1 \right| g(z) \, dz$$

$$= E \left| \frac{f_1(Z)}{g(Z)} - 1 \right| ,$$

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

\[ \| \mathcal{L}(\sqrt{n}Z_n) - \mathcal{L}(G_n) \| = \int_{\mathbb{R}^{pq}} |f_1(z) - g(z)| \, dz \]

\[ = E \left| \frac{g(Z)}{f_1(Z)} - 1 \right|, \]

where \( Z \) is the \( p \times q \) block of Haar-orthogonal matrix.
Lemma by D,E,L: Joint density of entries of $Z$ is $f(z) =$

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

if $\lambda_i$, 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

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

if $\lambda_i$, 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 $\Gamma_n = (\gamma_{ij})_{n \times n}$ be Haar-orthogonal matrix. Then

\begin{align*}
\text{a)} \quad & 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)}. \\
\text{b)} \quad & E(\gamma_{11} \gamma_{12} \gamma_{22} \gamma_{23} \gamma_{31} \gamma_{33}) = \frac{2}{(n-2)(n-1)n(n+2)(n+4)}. 
\end{align*}
To compute trace of submatrix, we need to compute following:

**Theorem**

Let $\Gamma_n = (\gamma_{ij})_{n \times n}$ 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) \ldots

d) \ldots

e) \ldots
Thank you for attending the talk!