Linear functions on the compact classical groups

Elizabeth Meckes
Stanford

Haar measure on $\mathcal{O}_n$

Construction:

1. Fill an $n \times n$ array with independent standard normal random variables, and perform the Gram-Schmidt process.

2. Pick a point at random from S^{n-1} and fill the first column with its coordinates. Fill the second column with a random point in the orthogonal complement (in S^{n-1}) of the first column, and so on.

<u>Meta-theorem</u>: Random orthogonal matrices are "like" Gaussian matrices.

Some actual theorems along these lines:

<u>Theorem</u> (Borel, 1906): Let X_1 be the first coordinate of a randomly chosen point on the sphere S^{n-1}. Then

$$\mathbb{P}(\sqrt{n}\, X_1 \leq t) \xrightarrow[n \to \infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2}\, dx$$

i.e., $\sqrt{n}\, X_1$ is asymptotically Gaussian.

* By construction 2. of Haar measure and its translation invariance, this means that if $M = (m_{ij}) \in \mathcal{O}_n$ is a Haar-distributed matrix, then

$$\mathbb{P}(\sqrt{n}\, m_{k\ell} \leq t) \xrightarrow[n \to \infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2}\, dx$$

for each k, ℓ.

A (much) later strengthening:

Theorem (Diaconis & Freedman, 1987):

Let X_1 be as before and $Z \sim N(0,1)$.
Then

$$d_{T.V.}(\sqrt{n}\, X_1, Z) \leq \frac{4}{n-1}$$

and the bound is sharp up to the value of the constant.

Theorem (Jiang, 2004):

Let $M \sim \text{Haar}(\mathcal{O}_n)$, $k = o(\sqrt{n})$ and Γ_k a $k \times k$ Gaussian matrix. Then if M_k is the top-left $k \times k$ block of M,

$$d_{TV}(M_k, \Gamma_k) \to 0 \quad \text{as } n \to \infty.$$

Theorem (Jiang, 2004):

For each $n \geq 2$ there are random matrices $M = (m_{ij})$ and $\Gamma = (\gamma_{ij})$ defined on the same probability space s.t.

1. $M \sim \text{Haar}(\mathcal{O}_n)$, Γ is Gaussian
2. If $\varepsilon_n(m) = \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} |\sqrt{n}\, m_{ij} - \gamma_{ij}|$, then if $m = o\left(\frac{n}{\log n}\right)$,

$$\varepsilon_n(m) \to 0 \text{ in probability as } n \to \infty.$$

Theorem (d'Aristotile/Diaconis/Newman, 2003):

Let M be a random $n \times n$ orthogonal matrix, and let $\{k_n\}$ be a sequence of numbers with $k_n \xrightarrow[n\to\infty]{} \infty$.

Let $\beta_1, \dots, \beta_{k_n}$ be a collection of k_n of the entries of M. Write

$$S_j^n = \sqrt{\frac{n}{k_n}} \sum_{i=1}^{j} \beta_i \,, \qquad W_n(t) = S^n_{\lfloor k_n t \rfloor}$$

Then

$$W_n \underset{n\to\infty}{\Rightarrow} W, \text{ a standard Brownian motion.}$$

Theorem (M., 2005): Let $M \in \mathcal{O}_n$ be distributed according to Haar measure and $A \in M_n(\mathbb{R})$ fixed, with $\mathrm{tr}(AA^t) = n$. Let $W = \mathrm{tr}(AM)$ and $Z \sim N(0,1)$. Then

$$d_{T.V.}(W, Z) \le \frac{2\sqrt{3}}{n-1}$$

and the rate is sharp up to the constant.

Background for the proof: Stein's method for normal approximati[on]

Idea: the normal distribution is characterized by the differential operator

$$f \overset{T}{\longmapsto} f'(x) - x f(x)$$

in the sense that:

1. $$\mathbb{E}[f'(Z) - Z f(Z)] \equiv 0 \quad \text{for } Z \sim N(0,1).$$

2. If Y is a random variable s.t.

$$\mathbb{E}[f'(Y) - Y f(Y)] \equiv 0$$

then $\quad Y \sim N(0,1).$

This follows because the differential equation

$$f'(x) - x f(x) = g(x) - \mathbb{E} g(Z)$$

can be solved for f in terms of g:

$$f(x) = U_0 g(x) = e^{x^2/2} \int_{-\infty}^{x} [g(t) - \mathbb{E} g(Z)] e^{-t^2/2} \, dt.$$

$\Rightarrow$ given g, if $f = U_0 g$, then

$$0 = \mathbb{E}[f'(Y) - Y f(Y)] = \mathbb{E} g(Y) - \mathbb{E} g(Z).$$

Next idea:

if $\mathbb{E}[f'(Y) - Yf(Y)]$ is <u>small</u> for a large class of f, then $Y \dot{\sim} N(0,1)$.

Fix a test function g, and let $f = U_0 g$.

$$\mathbb{E}g(Y) - \mathbb{E}g(Z) = \mathbb{E}[f'(Y) - Yf(Y)] \leftarrow \text{small}$$

Many notions of distance between random variables (or their distributions) can be expressed as

$$d_{\mathcal{F}}(X,Y) = \sup_{f \in \mathcal{F}} |\mathbb{E}f(X) - \mathbb{E}f(Y)|$$

for some class $\mathcal{F}$. For example:

$\mathcal{F}_1$ = continuous functions vanishing at ∞ and bounded by 1 $\longleftrightarrow$ total variation distance

$\mathcal{F}_2 = \{ f \in C^1(\mathbb{R}) : \|f\|_\infty \le 1, \ \|f'\|_\infty \le 1 \}$ $\longleftrightarrow$ dual-Lipschitz distance

$\mathcal{F}_3 = \{ f \in C^1(\mathbb{R}) : \|f'\|_\infty \le 1 \}$ $\longleftrightarrow$ Wasserstein distance

$\mathcal{F}_4 = \{ \mathbb{1}_{(-\infty, x]} : x \in \mathbb{R} \}$ $\longleftrightarrow$ distance between distribution functions

Exchangeable pairs

Have: W, conjectured to be normal (approximately).

From W, make a small change to get W' such that (W, W') is exchangeable; i.e., $(W, W') \sim (W', W)$.

Suppose that there is a number λ s.t.

$$\mathbb{E}[W' - W \mid W] = -\lambda W$$

$$\mathbb{E}[(W' - W)^2 \mid W] = 2\lambda$$

Then:

$$0 = \mathbb{E}[(W' - W)(f(W') + f(W))]$$

$$= \mathbb{E}[(W' - W)(f(W') - f(W)) + 2(W' - W) f(W)]$$

$$= \mathbb{E}[(W' - W)^2 f'(W) + 2(W' - W) f(W) + R]$$

$$= \mathbb{E}\left[\mathbb{E}[(W' - W)^2 \mid W] f'(W) + 2\,\mathbb{E}[W' - W \mid W] f(W) + R\right]$$

$$= \mathbb{E}[2\lambda f'(W) + 2(-\lambda W) f(W) + R]$$

$$\Longrightarrow \mathbb{E}[f'(W) - W f(W)] = \mathbb{E}\left[\frac{-R}{2\lambda}\right].$$

Theorem: $M \sim \text{Haar}(\mathcal{O}_n)$, $A \in M_n(\mathbb{R})$, $\operatorname{tr}(AA^t) = n$, $W := \operatorname{tr}(AM)$.

$$\Rightarrow \quad d_{TV}(W, Z) \leq \frac{2\sqrt{3}}{n-1}$$

<u>Proof</u>:

Define a family of exchangeable pairs (W, W_ε) as follows.

Define
$$A_\varepsilon = \begin{bmatrix} \sqrt{1-\varepsilon^2} & \varepsilon \\ -\varepsilon & \sqrt{1-\varepsilon^2} \end{bmatrix} \oplus I_{n-2}.$$
$$= I_n + \left[\left(-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right) I_2 + \varepsilon C_2\right] \oplus 0_{n-2}$$

$C_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

Let $U \sim \text{Haar}(\mathcal{O}_n)$, independent of M.

then $M_\varepsilon := U A_\varepsilon U^t M$

M_ε is a rotation of M by $\arcsin(\varepsilon)$ in a random 2-dimensional subspace.

Let $W_\varepsilon = \operatorname{tr}(AM_\varepsilon)$; (W, W_ε) is exchangeable for each fixed ε.

As before, fix g and let $f = U_o g$.

$$0 = \mathbb{E}\left[(W_\varepsilon - W)(f(W_\varepsilon) + f(W))\right]$$

$$= \mathbb{E}\left[(W_\varepsilon - W)^2 f'(W) + 2(W_\varepsilon - W) f(W) + R\right]$$

$$M_\varepsilon - M = U\left[\left(-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right) I_2 + \varepsilon C_2\right] U^t M$$

$$= \left(-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right) KK^t M + \varepsilon K C_2 K^t M$$

$$K = \begin{bmatrix} u_{11} & u_{12} \\ \vdots & \vdots \\ u_{n1} & u_{n2} \end{bmatrix} \qquad C_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

$$\Rightarrow W_\varepsilon - W = \mathrm{tr}(A(M_\varepsilon - M))$$

$$= \left(-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right) \mathrm{tr}(AKK^t M) + \varepsilon\, \mathrm{tr}(AKC_2K^t M)$$

Now, $(KK^t)_{ij} = u_{i1}u_{j1} + u_{i2}u_{j2} \Rightarrow \mathbb{E}\left[(KK^t)_{ij}\right] = \frac{2}{n}\delta_{ij}$

$(KC_2K^t)_{ij} = u_{i1}u_{j2} - u_{i2}u_{j1} \Rightarrow \mathbb{E}\left[(KC_2K^t)_{ij}\right] = 0.$

So $\mathbb{E}[W_\varepsilon - W \mid W]$

$$= \mathbb{E}\left[\left(-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right)\mathbb{E}\left[\mathrm{tr}(AKK^t M) \mid M\right] + \varepsilon\, \mathbb{E}\left[\mathrm{tr}(AKC_2K^t M) \mid M\right] \,\middle|\, W\right]$$

$$= \left(-\frac{\varepsilon^2}{n} + O(\varepsilon^4)\right)\mathbb{E}\left[\mathrm{tr}(AM) \mid W\right]$$

$$= \left(-\frac{\varepsilon^2}{n} + O(\varepsilon^4)\right) W.$$

We had:

$$0 = \mathbb{E}\left[\mathbb{E}[(W_\varepsilon - W)^2 | W] f'(W) + 2\mathbb{E}[W_\varepsilon - W | W] f(W) + R\right]$$

$$\Rightarrow 0 = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} \mathbb{E}\left[\mathbb{E}[(W_\varepsilon - W)^2 | W] f'(W) + 2\mathbb{E}[W_\varepsilon - W | W] f(W) + R\right] \quad *$$

(provided this makes sense)

from the last page,

$$\frac{1}{\varepsilon^2}\mathbb{E}[W_\varepsilon - W | W] = -\frac{1}{n} W + O(\varepsilon^2).$$

By Taylor's theorem,

$$|R| \le \|f''\|_\infty |W_\varepsilon - W|^3 = O(\varepsilon^3)$$

$$\Rightarrow \frac{1}{\varepsilon^2}\mathbb{E}|R| \to 0 \text{ as } \varepsilon \to 0.$$

From $W_\varepsilon - W = \left(-\frac{\varepsilon^2}{2} + O(\varepsilon^4)\right) \mathrm{tr}(AKK^tM) + \varepsilon\, \mathrm{tr}(AKC_2K^tM)$,

$$\mathbb{E}[|W_\varepsilon - W|^2 | W] = \varepsilon^2 \mathbb{E}[(\mathrm{tr}(AKC_2K^tM))^2 | W] + O(\varepsilon^3).$$

$$\Rightarrow \frac{1}{\varepsilon^2}\mathbb{E}[|W_\varepsilon - W|^2 | W] = \frac{2}{n} - \frac{2}{n(n-1)}\mathbb{E}[\mathrm{tr}((AM)^2) | W] + O(\varepsilon).$$

$$\Rightarrow \text{by } * \quad 0 = \mathbb{E}\left[\frac{2}{n} f'(W) - \frac{2}{n} W f(W) - \frac{2}{n(n-1)}\mathbb{E}[\mathrm{tr}((AM)^2) | W] f'(W)\right]$$

$$\Rightarrow \underbrace{\mathbb{E}[f'(W) - Wf(W)]}_{\mathbb{E}g(W) - \mathbb{E}g(Z)} = \frac{1}{n-1}\mathbb{E}\Big(\underbrace{\mathbb{E}[\mathrm{tr}((AM)^2) | W]}_{\text{bounded.}} \underbrace{f'(W)}_{\text{bounded for } g \text{ bounded}}\Big)$$

Things to take away from the proof

1. Stein's method is clever and powerful.

2. Taking advantage of the continuous symmetries of $\mathcal{O}_n$ led to a major improvement (it allowed the derivative approximation to become exact in the limit.)

3. The argument is quite general and doesn't use that much about Haar measure, at least to get started.

Other directions

1. $\mathcal{U}_n$: M a random unitary matrix, $A \in M_n(\mathbb{C})$, $\mathrm{tr}(AA^*) = n$, $W := \mathrm{tr}(AM)$.

   W is approximately normal but this is a multivariate problem, so the approach has to be modified.

2. Multivariate versions for $\mathcal{O}_n$, $\mathcal{U}_n$

   We can do it, but only for the weaker dual-Lipschitz metric on [illegible] metrizes the weak* topology!

I [illegible]:

Let M be a random orthogonal matrix and let $A_1, \dots, A_k$ be $n \times n$ matrices (fixed) with $\mathrm{tr}(A_i A_j^t) = n\delta_{ij}$. Consider the random vector

$$W = (\mathrm{tr}(A_1 M), \dots, \mathrm{tr}(A_k M)).$$

If $Z \in \mathbb{R}^k$ is a standard Gaussian random vector, then

$$d_{L^*}(W, Z) := \sup_{\substack{f, \\ \|f\|_\infty + \|f'\|_\infty \le 1}} |\mathbb{E}f(W) - \mathbb{E}f(Z)| \le \frac{c\, k^{3/2}}{n}$$

for an explicit constant c independent of k and n.

A very similar proof to the proof of the orthogonal result gives:

<u>Theorem</u> (M., 2008):

Let $\mathbb{X}$ be a compact locally symmetric space, and let $f : \mathbb{X} \to \mathbb{R}$ be an eigenfunction, with non-zero eigenvalue, of the Laplacian Δ on $\mathbb{X}$. Let X be a random point of $\mathbb{X}$, and consider the random variable $W = f(X)$. If $Z \sim N(0,1)$, the following bound holds

$$d_{TV}(W, Z) \leq \frac{1}{\lambda}\sqrt{\operatorname{Var}(\|\nabla f\|^2)}.$$

(convention: $\lambda > 0$)