Concentration and convergence rates for spectral measures

Joint work with Elizabeth Meckes

Mark Meckes

Case Western Reserve University

High Dimensional Probability VI
BIRS, October 14, 2011

Background

Given an $n \times n$ matrix M,

$$\mu_M = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(M)}$$

is the empirical spectral distribution of M.

$\nu =$ uniform measure on $\mathbb{T} = \{z : |z| = 1\}$.

Background

Given an $n \times n$ matrix M,

$$\mu_M = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(M)}$$

is the empirical spectral distribution of M.

$\nu =$ uniform measure on $\mathbb{T} = \{z : |z| = 1\}$.

Theorem (Diaconis–Shahshahani (1994))

For each n, let U_n be uniformly distributed in $\mathbb{U}(n)$, $\mathbb{O}(n)$, or $\mathbb{Sp}(2n)$. Then $\mu_{U_n} \Rightarrow \nu$ in probability.

Background

Theorem (Hiai–Petz (2006))

For each n, let U_n be uniformly distributed in $\mathbb{U}(n)$. Then μ_{U_n} satisfies the LDP in the scale n^{-2} with rate function

$$I(\mu) = -\int\int \log|x-y| \; d\mu(x)\, d\mu(y).$$

Roughly, if A is a set of measures on $\mathbb{T}$, then

$$\lim_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}\left[\mu \in A\right] = -\inf\{I(\rho) : \rho \in A\}.$$

Background

Theorem (Hiai–Petz (2006))

For each n, let U_n be uniformly distributed in $\mathbb{U}(n)$. Then μ_{U_n} satisfies the LDP in the scale n^{-2} with rate function

$$I(\mu) = -\int\int \log|x-y| \; d\mu(x)\, d\mu(y).$$

Roughly, if A is a set of measures on $\mathbb{T}$, then

$$\lim_{n\to\infty} \frac{1}{n^2} \log \mathbb{P}\left[\mu \in A\right] = -\inf\{I(\rho) : \rho \in A\}.$$

Corollary

For each n, let U_n be uniformly distributed in $\mathbb{U}(n)$. Then $\mu_{U_n} \Rightarrow \nu$ almost surely.

Similar results for $\mathbb{SU}(n)$ were proved by Hiai–Petz–Ueda (2006).

Main results

The Wasserstein distance between μ_1, μ_2 is

$$d_W(\mu_1, \mu_2) = \sup \{\mu_1(f) - \mu_2(f) \mid f : \mathbb{C} \to \mathbb{R} \text{ is 1-Lipschitz}\} .$$

Main results

The Wasserstein distance between μ_1, μ_2 is

$$d_W(\mu_1, \mu_2) = \sup \left\{ \mu_1(f) - \mu_2(f) \mid f : \mathbb{C} \to \mathbb{R} \text{ is 1-Lipschitz} \right\}.$$

Theorem (E. Meckes and M. M. (2011))

Let U be uniformly distributed in $\mathbb{U}(n)$, $\mathbb{SU}(n)$, $\mathbb{O}(n)$, $\mathbb{SO}(n)$, or $\mathbb{Sp}(2n)$. Then for $t > 0$,

$$\mathbb{P}\left[d_W(\mu_U, \nu) \geq C n^{-2/3} + t \right] \leq e^{-c n^2 t^2}.$$

Main results

The Wasserstein distance between μ_1, μ_2 is

$$d_W(\mu_1, \mu_2) = \sup \{\mu_1(f) - \mu_2(f) \mid f : \mathbb{C} \to \mathbb{R} \text{ is 1-Lipschitz}\}.$$

Theorem (E. Meckes and M. M. (2011))

Let U be uniformly distributed in $\mathbb{U}(n)$, $\mathbb{SU}(n)$, $\mathbb{O}(n)$, $\mathbb{SO}(n)$, or $\mathbb{Sp}(2n)$. Then for $t > 0$,

$$\mathbb{P}\left[d_W(\mu_U, \nu) \geq Cn^{-2/3} + t\right] \leq e^{-cn^2 t^2}.$$

Corollary

For each n, let U_n be uniformly distributed in $\mathbb{U}(n)$, $\mathbb{SU}(n)$, $\mathbb{O}(n)$, $\mathbb{SO}(n)$, or $\mathbb{Sp}(2n)$. Then almost surely, for large enough n,

$$d_W(\mu_{U_n}, \nu) \leq Cn^{-2/3}.$$

Not quite correct sketch of proof

Not quite correct sketch of proof

Step 1: By symmetry, $\mathbb{E}\mu_U = \nu$.

Not quite correct sketch of proof

Step 1: By symmetry, $\mathbb{E}\mu_U = \nu$.

Step 2: $M \mapsto \mu_M$ is a $n^{-1/2}$-Lipschitz map from $\{$normal matrices$\}$ with the Hilbert–Schmidt norm to $\{$probability measures on $\mathbb{C}\}$ with d_W.

This follows from the Hoffmann–Wielandt inequality for normal matrices and (most easily) the dual definition

$$d_W(\mu_1, \mu_2) = \inf\left\{ \int |x - y| \ d\pi(x, y) \middle| \pi \text{ has marginals } \mu_1, \mu_2 \right\}.$$

Not quite correct sketch of proof

Step 1: By symmetry, $\mathbb{E}\mu_U = \nu$.

Step 2: $M \mapsto \mu_M$ is a $n^{-1/2}$-Lipschitz map from $\{\text{normal matrices}\}$ with the Hilbert–Schmidt norm to $\{\text{probability measures on } \mathbb{C}\}$ with d_W.

This follows from the Hoffmann–Wielandt inequality for normal matrices and (most easily) the dual definition

$$d_W(\mu_1, \mu_2) = \inf \left\{ \int |x - y| \; d\pi(x, y) \middle| \pi \text{ has marginals } \mu_1, \mu_2 \right\}.$$

Two particular consequences:

- $M \mapsto \mu_M(f)$ is $n^{-1/2} |f|_L$-Lipschitz.
- $M \mapsto d_W(\mu_M, \mu)$ is $n^{-1/2}$-Lipschitz.

Not quite correct sketch of proof, continued

Step 3: For random U, define $X_f = \mu_U(f) - \nu(f)$. The Gromov–Milman concentration phenomenon for classical matrix groups shows that

$$\mathbb{P}\left[\left|X_f - X_g\right| \geq t\right] \leq 2e^{-cn^2t^2/|f-g|_L^2}.$$

The mean Wasserstein distance

$$\mathbb{E}d_W(\mu_U, \nu) = \mathbb{E} \sup_{|f|_L \leq 1} X_f$$

can now be estimated by classical entropy methods.

Not quite correct sketch of proof, continued

Step 3: For random U, define $X_f = \mu_U(f) - \nu(f)$. The Gromov–Milman concentration phenomenon for classical matrix groups shows that

$$\mathbb{P}\big[\big|X_f - X_g\big| \geq t\big] \leq 2e^{-cn^2t^2/|f-g|_L^2}.$$

The mean Wasserstein distance

$$\mathbb{E}d_W(\mu_U, \nu) = \mathbb{E} \sup_{|f|_L \leq 1} X_f$$

can now be estimated by classical entropy methods.

Step 4: Gromov–Milman again shows that

$$\mathbb{P}\big[d_W(\mu_U, \nu) \geq \mathbb{E}d_W(\mu_U, \nu) + t\big] \leq e^{-cn^2t^2}.$$

Problems, and how to fix them

Step 1: The symmetry argument that $\mathbb{E}\mu_U = \nu$ only works for $\mathbb{U}(n)$. For the other groups, we need to estimate

$$(\mathbb{E}\mu_U)(f) = \mathbb{E}(\mu_U(f)) = \frac{1}{n}\mathbb{E}\operatorname{Tr} f(U).$$

Results of Diaconis–Mallows, Diaconis–Shahshahani, and Rains show that if $1 \leq |k| < n$, $\left|\mathbb{E}\operatorname{Tr} U^k\right| \leq 1$.

Problems, and how to fix them

Step 1: The symmetry argument that $\mathbb{E}\mu_U = \nu$ only works for $\mathbb{U}(n)$. For the other groups, we need to estimate

$$(\mathbb{E}\mu_U)(f) = \mathbb{E}(\mu_U(f)) = \frac{1}{n}\mathbb{E}\operatorname{Tr} f(U).$$

Results of Diaconis–Mallows, Diaconis–Shahshahani, and Rains show that if $1 \leq |k| < n$, $\left|\mathbb{E}\operatorname{Tr} U^k\right| \leq 1$.

Classical results on approximation of Lipschitz functions on $\mathbb{T}$ by polynomials imply that

$$d_W(\mathbb{E}\mu_U, \nu) \leq C\frac{\log n}{n}.$$

For $\mathbb{SU}(n)$, the $\log n$ can be removed.

Problems, and how to fix them, continued

Step 2 is correct as stated.

Problems, and how to fix them, continued

Step 2 is correct as stated.

Steps 3 and 4: The Gromov–Milman theorem doesn't apply to $\mathbb{U}(n)$ or $\mathbb{O}(n)$.

This can be addressed by conditioning on $\det U$. Alternatively, for $\mathbb{U}(n)$, a coupling argument shows that

$$d_W(\mu_U, \nu)$$

has the same distribution for $U \in \mathbb{U}(n)$ and $U \in \mathbb{SU}(n)$.

Problems, and how to fix them, concluded

Step 3: Gromov–Milman shows that the stochastic process

$$X_f = \mu_U(f) - \mathbb{E}\mu_U(f)$$

indexed by $\{|f|_L \leq 1\}$ has subgaussian increments w.r.t. the seminorm $|\cdot|_L$. But the metric entropy of the ball of an infinite-dimensional space w.r.t. its own norm is infinite, so entropy methods don't directly apply.

Problems, and how to fix them, concluded

Step 3: Gromov–Milman shows that the stochastic process

$$X_f = \mu_U(f) - \mathbb{E}\mu_U(f)$$

indexed by $\{|f|_L \leq 1\}$ has subgaussian increments w.r.t. the seminorm $|\cdot|_L$. But the metric entropy of the ball of an infinite-dimensional space w.r.t. its own norm is infinite, so entropy methods don't directly apply.

We first approximate f in the sup norm by, say, a piecewise affine function.

Problems, and how to fix them, concluded

Step 3: Gromov–Milman shows that the stochastic process

$$X_f = \mu_U(f) - \mathbb{E}\mu_U(f)$$

indexed by $\{|f|_L \leq 1\}$ has subgaussian increments w.r.t. the seminorm $|\cdot|_L$. But the metric entropy of the ball of an infinite-dimensional space w.r.t. its own norm is infinite, so entropy methods don't directly apply.

We first approximate f in the sup norm by, say, a piecewise affine function.

Up to constants, the estimate on $\mathbb{E} \sup X_f$ appears to be the only nonoptimal part of our results.

Circular ensembles

Let $U \in \mathbb{U}(n)$ be uniformly distributed.

- The Circular Unitary Ensemble is distributed as U.
- The Circular Orthogonal Ensemble is distributed as $U^T U$.
- The Circular Symplectic Ensemble is distributed as $JU^T JU$, $J = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \cdots \oplus \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$. (Here $U \in \mathbb{U}(2n)$.)

Circular ensembles

Let $U \in \mathbb{U}(n)$ be uniformly distributed.

- The Circular Unitary Ensemble is distributed as U.
- The Circular Orthogonal Ensemble is distributed as $U^T U$.
- The Circular Symplectic Ensemble is distributed as $JU^T JU$, $J = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \cdots \oplus \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$. (Here $U \in \mathbb{U}(2n)$.)

Theorem

Let V be drawn from the COE(n), CUE(n), or CSE(2n). Then for $t > 0$,

$$\mathbb{P}\left[d_W(\mu_V, \nu) \geq Cn^{-2/3} + t\right] \leq e^{-cn^2 t^2}.$$

Circular ensembles

Let $U \in \mathbb{U}(n)$ be uniformly distributed.

- The Circular Unitary Ensemble is distributed as U.
- The Circular Orthogonal Ensemble is distributed as $U^T U$.
- The Circular Symplectic Ensemble is distributed as $JU^T JU$, $J = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right] \oplus \cdots \oplus \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$. (Here $U \in \mathbb{U}(2n)$.)

Theorem

Let V be drawn from the COE(n), CUE(n), or CSE(2n). Then for $t > 0$,

$$\mathbb{P}\left[d_W(\mu_V, \nu) \geq Cn^{-2/3} + t\right] \leq e^{-cn^2 t^2}.$$

Corollary

For each n, let V_n be drawn from the COE(n), CUE(n), or CSE(2n). Then almost surely, for large enough n,

$$d_W(\mu_{V_n}, \nu) \leq Cn^{-2/3}.$$

Randomized sums

Theorem

Fix Hermitian $n \times n$ matrices A, B with $\|A\|, \|B\| < K$, and define

$$M = UAU^* + B,$$

where $U \in \mathbb{U}(n)$ is uniformly distributed. Then for $t > 0$,

$$\mathbb{P}\left[d_W(\mu_M, \mathbb{E}\mu_M) \geq Cn^{-2/3} + t\right] \leq e^{-cn^2t^2}.$$

Randomized sums

Theorem

Fix Hermitian $n \times n$ matrices A, B with $\|A\|, \|B\| < K$, and define

$$M = UAU^* + B,$$

where $U \in \mathbb{U}(n)$ is uniformly distributed. Then for $t > 0$,

$$\mathbb{P}\left[d_W(\mu_M, \mathbb{E}\mu_M) \geq Cn^{-2/3} + t\right] \leq e^{-cn^2t^2}.$$

This also holds for random A and B satisfying a concentration hypothesis.

A similar result (for fixed A and B) was proved by Kargin (2011).

Random compressions

Theorem

Fix a Hermitian $n \times n$ matrix A with $\|A\| < K$, and define

$$M = P_k U A U^* P_k^*,$$

where $U \in \mathbb{U}(n)$ is uniformly distributed and $P_k : \mathbb{R}^n \to \mathbb{R}^k$ is the usual projection. Then for $t > 0$,

$$\mathbb{P}\left[d_W(\mu_M, \mathbb{E}\mu_M) \geq Cn^{-2/3} + t\right] \leq e^{-cn^2 t^2}.$$

Random compressions

Theorem

Fix a Hermitian $n \times n$ matrix A with $\|A\| < K$, and define

$$M = P_k U A U^* P_k^*,$$

where $U \in \mathbb{U}(n)$ is uniformly distributed and $P_k : \mathbb{R}^n \to \mathbb{R}^k$ is the usual projection. Then for $t > 0$,

$$\mathbb{P}\left[d_W(\mu_M, \mathbb{E}\mu_M) \geq Cn^{-2/3} + t\right] \leq e^{-cn^2 t^2}.$$

This also holds for random A satisfying a concentration hypothesis.

This improves an earlier result of Meckes–M. (2011). When $k = n$ and A is random, this sharpens results of Guionnet–Zeitouni (2000). When $k = n$ and A is a Wigner matrix sharper results were proved by Götze–Tikhomirov (2011).

Thank you.