A new approach to strong embeddings

Sourav Chatterjee

UC Berkeley

Coupling of random walks with Brownian motion

- Suppose $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random variables with $\mathbb{E}(\varepsilon_1) = 0$ and $\mathbb{E}(\varepsilon_1^2) = 1$. For each k, let

$$S_k = \sum_{i=1}^{k} \varepsilon_i.$$

- Goal is to construct a standard Brownian motion $(B_t)_{t \geq 0}$ on the same probability space so as to minimize the growth rate of

$$\max_{1 \leq k \leq n} |S_k - B_k|.$$

Skorokhod embeddings

- Skorokhod (1961), Strassen (1966): Start with $(B_t)_{t\geq 0}$, and construct stopping times $T_1 \leq T_2 \leq \cdots$ such that
  - B_{T_1} has the same distribution as ε_1.
  - $B_{T_1}, B_{T_2} - B_{T_1}, B_{T_3} - B_{T_2}, \cdots$ are i.i.d. In other words, $(B_{T_k})_{k\geq 1}$ has the same law as $(S_k)_{k\geq 1}$.
  - $E(T_{i+1} - T_i \mid (B_t)_{t\leq T_i}) = 1$ for each i.
- Show that $T_k \simeq k$ in an average sense, to bound $\max_{1\leq k\leq n} |B_{T_k} - B_k|$.
- Strassen (1966) proved that

$$\max_{1\leq k\leq n} |B_{T_k} - B_k| = O((n \log\log n)^{1/4}(\log n)^{1/2}),$$

  and conjectured that this is the best possible rate under $\mathbb{E}(\varepsilon_1^2) < \infty$. Proved by Kiefer (1969).
- Is it possible to improve, assuming stronger moment conditions on the summands?

The KMT embedding theorem

Indeed, yes. If ε_1 has a finite moment generating function in a neighborhood of zero, then one can get

$$\max_{k\leq n} |S_k - B_k| = O(\log n).$$

Moreover, this is the best possible.

Theorem (Komlós-Major-Tusnády, 1975)

Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. random variables with $\mathbb{E} \exp \theta|\varepsilon_1| < \infty$ for some $\theta > 0$. Let $S_k := \sum_{i=1}^k \varepsilon_i$, $k = 0, 1, \ldots$ be the corresponding random walk. It is possible to construct a version of the sequence $(S_k)_{k\geq 0}$ and a standard Brownian motion $(B_t)_{t\geq 0}$ on the same probability space such that for every n and every $t \geq 0$,

$$\mathbb{P}\bigl(\max_{k\leq n} |S_k - B_k| \geq C \log n + t\bigr) \leq K e^{-\lambda t},$$

where C, K, and λ do not depend on n.

Further developments

- Numerous applications in both applied and theoretical problems.
- Invaluable for understanding fine properties of simple random walk, e.g. in the works of Dembo-Peres-Rosen-Zetouni, Lawler, etc.
- There is a different KMT theorem for empirical processes, in the same paper. Of great interest to statisticians.
- No version yet for the non-i.i.d. case.
- Original paper still considered to be very hard to read. We will give a new (simpler?) proof for Bernoulli summands.

The first step

Lemma

Suppose $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are i.i.d. symmetric ± 1-valued r.v. Let $S_n = \sum_{i=1}^n \varepsilon_i$. Then it is possible to construct a version of S_n and a gaussian r.v. Z_n with mean 0 and variance n on the same probability space such that for all $t \geq 0$,

$$\mathbb{P}(|S_n - Z_n| \geq t) \leq K e^{-\lambda t},$$

where K and λ do not depend on n.

Question: Can we have a version of this in more complex normal approximation problems?

An abstraction

Lemma

Suppose W is a random variable with $\mathbb{E}(W) = 0$ and $\mathbb{E}(W^2) < \infty$. Let T be another random variable, defined on the same probability space as W, satisfying

$$\mathbb{E}(W\varphi(W)) = \mathbb{E}(\varphi'(W)T)$$

for all Lipschitz φ. Suppose $|T|$ is almost surely bounded by a constant. Then, given any $\sigma^2 > 0$, we can construct $Z \sim N(0, \sigma^2)$ on the same probability space such that for any $\theta \in \mathbb{R}$,

$$\mathbb{E}\exp(\theta|W - Z|) \leq 2\,\mathbb{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right).$$

- Key idea, inspired by Stein's method: $T \simeq \sigma^2 \implies W$ is approximately $N(0, \sigma^2)$. I call T a Stein coefficient of W.
- However, classical Stein's method can only give bounds on quantities like $\sup_{f \in \mathcal{F}} |\mathbb{E}f(W) - \mathbb{E}f(Z)|$, for various classes $\mathcal{F}$. The above result seems to be of a fundamentally different nature.

Examples

- Suppose X is a random variable with $\mathbb{E}(X) = 0$, $\mathbb{E}(X^2) < \infty$, and density ρ. Let
$$h(x) = \frac{\int_x^\infty y\rho(y)dy}{\rho(x)}.$$
- Then, by integration-by-parts, we have
$$\mathbb{E}(X\varphi(X)) = \mathbb{E}(\varphi'(X)h(X)).$$
Thus, $h(X)$ is a Stein coefficient for X.
- Suppose $X_1, \ldots, X_n$ are i.i.d. copies of X, and let $W = n^{-1/2}\sum_{i=1}^n X_i$. Then by the above result,
$$\begin{aligned}\mathbb{E}(W\varphi(W)) &= n^{-1/2}\sum_{i=1}^n \mathbb{E}(X_i\varphi(W)) \\ &= n^{-1}\sum_{i=1}^n \mathbb{E}(h(X_i)\varphi'(W)).\end{aligned}$$
- Thus, $n^{-1}\sum h(X_i)$ is a Stein coefficient for W.

A discrete example

- Suppose $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. symmetric ± 1-valued r.v. Let $S_n = \sum_{i=1}^n \varepsilon_i$.
- Let $Y \sim \mathrm{Uniform}[-1, 1]$. Let $W = S_n + Y$.
- Let
$$T = n - S_n Y + \frac{1 - Y^2}{2}.$$
- It follows from a calculation involving integration-by-parts that for all φ,
$$\mathbb{E}(W\varphi(W)) = \mathbb{E}(\varphi'(W)T),$$
that is, T is a Stein coefficient for W.
- Letting $\sigma^2 = n$, the abstract lemma tells us that it is possible to construct $Z \sim N(0, n)$ such that
$$\mathbb{E}\exp(\theta|W - Z|) \leq 2\,\mathbb{E}\exp\left(\frac{2\theta^2(T - n)^2}{n}\right).$$
Since $T = n + O(\sqrt{n})$, this proves the lemma.

A general class of examples

- Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ be a vector of i.i.d. standard gaussian r.v.
- Let $W = f(\mathbf{X})$, where f is absolutely continuous. Suppose $\mathbb{E}(W) = 0$.
- Let $\mathbf{X}' = (X_1', \ldots, X_n')$ be an independent copy of $\mathbf{X}$.
- Let
$$T = \int_0^1 \frac{1}{2\sqrt{t}} \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{X}) \frac{\partial f}{\partial x_i}(\sqrt{1-t}\mathbf{X} + \sqrt{t}\mathbf{X}')dt.$$
- Then, one can show that T is a Stein coefficient for W.
- This can be used, e.g., to prove CLTs for linear statistics of eigenvalues of random matrices. (Different talk.)

Proof of abstract lemma: Step I

- Recall setup: (W, T) is a pair of r.v. such that for all φ,

$$\mathbb{E}(W\varphi(W)) = \mathbb{E}(\varphi'(W)T).$$

Given σ^2, we are trying to construct $Z \sim N(0, \sigma^2)$ such that

$$\mathbb{E}\exp(\theta|W - Z|) \le 2\,\mathbb{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right).$$

- Let $h(W) = \mathbb{E}(T \mid W)$. Then for all φ,

$$\mathbb{E}(W\varphi(W)) = \mathbb{E}(\varphi'(W)h(W)).$$

The function h characterizes the distribution of W via this equation.

- One can show that it suffices to construct (W, Z) such that $Z \sim N(0, \sigma^2)$ and for all $\theta > 0$,

$$\mathbb{E}\exp(\theta|W - Z|) \le 2\,\mathbb{E}\exp(2\theta^2(\sqrt{h(W)} - \sigma)^2).$$

Step II: Construction of the coupling

- Fix a function $r : \mathbb{R}^2 \to \mathbb{R}$.
- For $f \in C^2(\mathbb{R}^2)$, let

$$\mathcal{L}f(x,y) := h(x)\frac{\partial^2 f}{\partial x^2} + 2r(x,y)\frac{\partial^2 f}{\partial x \partial y} + \sigma^2 \frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

- Suppose there exists a probability measure μ on $\mathbb{R}^2$ such that for all f,

$$\int_{\mathbb{R}^2} \mathcal{L}f \; d\mu = 0. \qquad (*)$$

- We will show that every choice of r that allows a μ satisfying $(*)$ gives rise to a coupling of W and Z.

Step II contd.

- Recap: we have μ such that for all f

$$(*) \quad \int_{\mathbb{R}^2} \mathcal{L}f \, d\mu = 0, \quad \text{where}$$

$$\mathcal{L}f(x,y) := h(x)\frac{\partial^2 f}{\partial x^2} + 2r(x,y)\frac{\partial^2 f}{\partial x \partial y} + \sigma^2 \frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

- Let $(X,Y) \sim \mu$.
- Take any $\Phi \in C^2(\mathbb{R})$, and let $\varphi = \Phi'$. Putting $f(x,y) = \Phi(x)$ in $(*)$, we get

$$\mathbb{E}(h(X)\varphi'(X) - X\varphi(X)) = 0.$$

  Thus, X has the same law as W.
- Similarly, putting $f(x,y) = \Phi(y)$, we get $\mathbb{E}(Y\varphi(Y)) = \sigma^2\mathbb{E}(\varphi'(Y))$, and thus, $Y \sim N(0,\sigma^2)$.
- Note that these deductions are independent of the choice of $r(x,y)$, as long as μ exists. Each valid choice of $r(x,y)$ gives a coupling of W and Z.

Step III: The key lemma

Lemma

Suppose that the matrix valued function

$A(x,y) := \begin{pmatrix} h(x) & r(x,y) \\ r(x,y) & \sigma^2 \end{pmatrix}$ *is bounded, positive semidefinite, and continous everywhere. Then there exists a probability measure* μ *such that for all* $f \in C^2(\mathbb{R}^2)$ *satisfying certain mild conditions, we have*

$$(*) \quad \int_{\mathbb{R}^2} \mathcal{L}f \; d\mu = 0, \quad \textit{where}$$

$$\mathcal{L}f(x,y) := h(x)\frac{\partial^2 f}{\partial x^2} + 2r(x,y)\frac{\partial^2 f}{\partial x \partial y} + \sigma^2 \frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

Note that

- The operator $\mathcal{L}$ is not uniformly elliptic.
- The domain of f is unbounded.
- The functions h and r need not be Lipschitz.
- Can be solved by stochastic process techniques. Will come back to give a different, self-contained proof.

Step IV: Positive definite completion

- Thus, finding $r(x, y)$ is a problem of positive definite completion for the incomplete matrix

$$\begin{pmatrix} h(x) & ? \\ ? & \sigma^2 \end{pmatrix}.$$

- Intuition: the more nearly singular the completion, the tighter the coupling between the two coordinates.
- The 'most singular choice' is given by the geometric mean

$$r(x, y) = \sigma\sqrt{h(x)}.$$

- Another choice: $r(x, y) = h(x) \wedge \sigma^2$. This is not good: It brings us back to Skorokhod embeddings. (Will not elaborate now.)

Step V: Getting bounds

- With $r(x,y) = \sigma\sqrt{h(x)}$, we have

$$\mathcal{L}f(x,y) = h(x)\frac{\partial^2 f}{\partial x^2} + 2\sigma\sqrt{h(x)}\frac{\partial^2 f}{\partial x \partial y} + \sigma^2\frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

- With $f(x,y) = \frac{1}{2k}(x-y)^{2k}$, we get

$$\mathcal{L}f(x,y) = (2k-1)(x-y)^{2k-2}(\sqrt{h(x)} - \sigma)^2 - (x-y)^{2k}.$$

- So, if $\mathbb{E}(\mathcal{L}f(X,Y)) = 0$ for all f, then

$$\begin{aligned}\mathbb{E}(X-Y)^{2k} &= (2k-1)\mathbb{E}((X-Y)^{2k-2}(\sqrt{h(X)}-\sigma)^2) \\ &\le (2k-1)(\mathbb{E}(X-Y)^{2k})^{\frac{k-1}{k}}(\mathbb{E}(\sqrt{h(X)}-\sigma)^{2k})^{1/k}.\end{aligned}$$

- This gives

$$\mathbb{E}(X-Y)^{2k} \le (2k-1)^k\mathbb{E}(\sqrt{h(X)}-\sigma)^{2k}.$$

The proof of the lemma is completed by summing over $k \ge 1$.

Proof of the key lemma

- Let A be a continous bounded map from $\mathbb{R}^2$ into the set of all 2×2 positive semidefinite matrices.
- Take any probability measure μ on $\mathbb{R}^2$. Suppose $\mathbf{X} \sim \mu$, and let $\mathbf{Z}$ be a standard gaussian random vector independent of $\mathbf{X}$.
- For each $\varepsilon > 0$, let $T_\varepsilon \mu$ be the law of the random vector

$$(1-\varepsilon)\mathbf{X} + \sqrt{2\varepsilon A(\mathbf{X})}\mathbf{Z},$$

  where $\sqrt{A}$ denotes the positive definite square root of A. Then T_ε is a continuous map.
- Suppose $\|A(\mathbf{x})\| \le b$ for all $\mathbf{x} \in \mathbb{R}^2$. Let K be the set of all probability measures on $\mathbb{R}^2$ satisfying

$$\int \mathbf{x} d\mu(\mathbf{x}) = 0 \quad \text{and} \quad \int \exp\langle \mathbf{u}, \mathbf{x}\rangle d\mu(\mathbf{x}) \le \exp(b\|\mathbf{u}\|^2) \text{ for all } \mathbf{u} \in \mathbb{R}^2.$$

  Easy to verify that K is non-empty, compact, and convex.
- Main observation: For any $0 < \varepsilon < 1$, $T_\varepsilon(K) \subseteq K$.
- So, by the Schauder-Tychonoff fixed point theorem for loally convex spaces, for every $\varepsilon \in (0,1)$, $\exists \mu_\varepsilon \in K$ such that $T_\varepsilon \mu_\varepsilon = \mu_\varepsilon$.

Proof of key lemma contd.

- Since K is compact, $\exists \mu \in K$ such that for some sequence $\varepsilon_n \downarrow 0$, $\mu_{\varepsilon_n} \to \mu$.
- Suppose $\mathbf{X}_\varepsilon \sim \mu_\varepsilon$. Using the identity $T_\varepsilon \mu_\varepsilon = \mu_\varepsilon$, we get that for any f,

$$\mathbb{E}f((1-\varepsilon)\mathbf{X}_\varepsilon + \sqrt{2\varepsilon A(\mathbf{X}_\varepsilon)}\mathbf{Z}) = \mathbb{E}f(\mathbf{X}_\varepsilon).$$

- Taylor expanding the LHS around $\mathbf{X}_\varepsilon$, dividing both sides by ε, and letting $\varepsilon \downarrow 0$ along $\{\varepsilon_n\}$, we end up with

$$(*) \quad \int_{\mathbb{R}^2} \mathcal{L}f \, d\mu = 0, \quad \text{where}$$

$$\mathcal{L}f(x,y) := A_{11}(x,y)\frac{\partial^2 f}{\partial x^2} + 2A_{12}(x,y)\frac{\partial^2 f}{\partial x \partial y} + A_{22}(x,y)\frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

- Thus, μ satisfies our quest and completes the proof of the lemma.

Completing the proof of the KMT theorem

We prove the following by induction on n.

Theorem

There exist positive universal constants C, K and λ_0 such that the following is true. Take any integer $n \geq 2$. Suppose $\varepsilon_1, \ldots, \varepsilon_n$ are exchangeable ± 1 random variables. For $k = 0, 1, \ldots, n$, let $S_k = \sum_{i=1}^k \varepsilon_i$ and let

$$W_k = S_k - \frac{k}{n} S_n.$$

It is possible to construct a version of $W_0, \ldots, W_n$ and a standard Brownian bridge $(\widetilde{B}_t)_{0 \leq t \leq 1}$ on the same probability space such that for any $0 < \lambda < \lambda_0$,

$$\mathbb{E} \exp(\lambda \max_{k \leq n} |W_k - \sqrt{n} \widetilde{B}_{k/n}|) \leq \exp(C \log n) \mathbb{E} \exp\left(\frac{K \lambda^2 S_n^2}{n} \right).$$

How to carry out the induction

The main ingredient is the following result, which is proved using the abstract lemma. (I won't go into more details.)

Theorem

Let $\varepsilon_1, \ldots, \varepsilon_n$ *be* n *arbitrary elements of* $\{-1, 1\}$. *Let* π *be a uniform random permutation of* $\{1, \ldots, n\}$. *For each* $1 \leq k \leq n$, *let* $S_k = \sum_{\ell=1}^{k} \varepsilon_{\pi(\ell)}$, *and let*

$$W_k = S_k - \frac{kS_n}{n}.$$

There exist universal constants $c > 1$ *and* $\theta_0 > 0$ *satisfying the following. Take any* $n \geq 3$, *any possible value of* S_n, *and any* $n/3 \leq k \leq 2n/3$. *It is possible to construct a version of* W_k *and a gaussian random variable* Z_k *with mean* 0 *and variance* $k(n-k)/n$ *on the same probability space such that for any* $\theta \leq \theta_0$,

$$\mathbb{E} \exp(\theta |W_k - Z_k|) \leq \exp\left(1 + \frac{c\theta^2 S_n^2}{n}\right).$$

A few remarks about the intuition

- If $\mathbb{E}(X\varphi(X)) = \mathbb{E}(\varphi'(X)h(X))$ for all φ, then the law of X is an invariant measure for the diffusion

$$dX_t = -X_t dt + \sqrt{2h(X_t)}dB_t.$$

- Suppose we have independent diffusions

$$dX_t^i = -X_t^i dt + \sqrt{2h(X_t^i)}dB_t^i, \quad i = 1, \ldots, n.$$

- Let $W_t = n^{-1/2} \sum_{i=1}^n X_t^i$. Then

$$dW_t = -W_t dt + n^{-1/2} \sum_{i=1}^n \sqrt{2h(X_t^i)}dB_t^i.$$

Intuition contd.

- If we define another process $(\widetilde{B}_t)_{t\geq 0}$ by

$$d\widetilde{B}_t = \frac{\sum_{i=1}^n \sqrt{2h(X_t^i)}dB_t^i}{\sqrt{\sum_{i=1}^n 2h(X_t^i)}},$$

then $\widetilde{B}_t$ is again a standard Brownian motion, and

$$dW_t = -W_t dt + \sqrt{\frac{\sum_{i=1}^n 2h(X_t^i)}{n}}d\widetilde{B}_t.$$

- Since $\mathbb{E}(h(X)) = \mathbb{E}(X^2) =: \sigma^2$, therefore $n^{-1}\sum_{i=1}^n 2h(X_t^i) \simeq 2\sigma^2$ with high probability. Thus, W_t is 'approximately' an Ornstein-Uhlenbeck process.
- If we define $dZ_t = -Z_t dt + \sqrt{2}\sigma d\widetilde{B}_t$, then Z_t is an actual O-U process.
- Moreover, Z_t and W_t 'come close' at infinity. Our abstract lemma gets its hands on (W_∞, Z_∞).

An important question

Recall:

Lemma

Suppose W is a random variable with $\mathbb{E}(W) = 0$ and $\mathbb{E}(W^2) < \infty$. Let T be another random variable, defined on the same probability space as W, satisfying

$$\mathbb{E}(W\varphi(W)) = \mathbb{E}(\varphi'(W)T)$$

for all Lipschitz φ. Suppose $|T|$ is almost surely bounded by a constant. Then, given any $\sigma^2 > 0$, we can construct $Z \sim N(0, \sigma^2)$ on the same probability space such that for any $\theta \in \mathbb{R}$,

$$\mathbb{E}\exp(\theta|W - Z|) \leq 2\,\mathbb{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right).$$

Does there exist a multidimensional version of this lemma?

Immediate consequence: Direct proof of KMT treating the problem as a coupling of random vectors; possibly many other implications.