Some progress using Stein's method for strong embeddings

Chinmoy Bhattacharjee, Larry Goldstein
University of Southern California

Workshop on New Directions in Stein's method, May 25th, 2015
[arXiv1505.03199]

Strong Embedding

Let $\epsilon, \epsilon_1, \epsilon_2 \ldots$ be i.i.d, mean zero, variance one, and for $1 \leq k \leq n$ and $t \in [0,1]$ let

$$S_k = \sum_{i=1}^{k} \epsilon_i \quad \text{and} \quad X_n(t) = \frac{1}{\sqrt{n}}[S_{[nt]} + (nt - [nt])\epsilon_{[nt]+1}].$$

Then $X_n(t)$ converges weakly to Brownian Motion $(B_t)_{t \geq 0}$ on $[0,1]$ by Donsker.

Find the best (smallest) rate of growth for $f(n)$ that can be achieved by a coupling

$$\max_{0 \leq k \leq n} |S_k - B_k| = O_p(f(n)).$$

Strong Embedding

We say strong embedding (SE) holds for $\mathcal{L}(\epsilon)$ if there exist positive constants C, K and λ such that

$$P\left(\max_{0\le k\le n}|S_k - B_k| \ge C\log n + x\right) \le Ke^{-\lambda x} \quad \text{for all } x \ge 0.$$

Komlós, Major and Tusnády (or KMT) 1975, SE holds for $\mathcal{L}(\epsilon)$ with finite moment generating function in a neighborhood of zero.

Best possible rate, Bártfai 1966.

Numerous applications, see the texts of Shorack and Wellner, 2009, or Csörgő and Révész, 2014.

Coupling constructions used are quite complicated.

Previous Results using Stein for KMT

Chatterjee 2012 shows SE holds for Rademacher summands.

And a 'collateral' new result that for $\epsilon_1, \ldots, \epsilon_n$ exchangeable variables taking values in $\{-1, +1\}$, there exist a coupling of

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

and $(B_t)_{t \in [0,1]}$, a standard Brownian Bridge, and positive constants λ_0, C and K such that for all $\lambda \leq \lambda_0$

$$E \exp(\lambda \max_{0 \leq k \leq n} |W_k - \sqrt{n} B_{k/n}|) \leq E \exp\left(C \log n + \frac{K \lambda^2 S_n^2}{n} \right).$$

Will extend these results to variables with vanishing third moment taking values in a finite set $\mathcal{A}$ not containing zero.

Previous Results using Stein for KMT

Chatterjee 2012 shows SE holds for Rademacher summands.

And a 'collateral' new result that for $\epsilon_1, \ldots, \epsilon_n$ exchangeable variables taking values in $\{-1, +1\}$, there exist a coupling of

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

and $(B_t)_{t \in [0,1]}$, a standard Brownian Bridge, and positive constants λ_0, C and K such that for all $\lambda \leq \lambda_0$

$$E \exp(\lambda \max_{0 \leq k \leq n} |W_k - \sqrt{n} B_{k/n}|) \leq E \exp\left(C \log n + \frac{K \lambda^2 S_n^2}{n}\right).$$

Will extend these results to variables with vanishing third moment taking values in a finite set $\mathcal{A}$ not containing zero.

Stein Coefficient/Kernel

Modify Stein identity for normal $Z \sim \mathcal{N}(0,1)$

$$E[Zf(Z)] = E[f'(Z)] \quad \text{to} \quad E[Wf(W)] = E[Tf'(W)]$$

for W with mean zero and variance 1, some 'Stein coefficient' T. See Cacoullos and Papathanasiou 1992.

Compare to zero bias transformation (G. and Reinert 1997)

$$E[Wf(W)] = E[f'(W^*)] \quad \text{we see} \quad E[T|W] = \frac{dF^*}{dF}$$

As W^* always has a density, T cannot exist when the distribution of W has no absolutely continuous component, e.g. for symmetric Bernoulli variables. Smoothing required.

Generally such T are valuable, and not so easy to construct.

T yields Marginal Coupling Bound

Theorem 1.2, Chatterjee 2012

Let W be mean zero with finite second moment and suppose that T is a Stein coefficient for W with $|T|$ almost surely bounded by a constant. Then, given any $\sigma^2 > 0$, we can construct a version of W and $Z \sim \mathcal{N}(0, \sigma^2)$ on the same probability space such that

$$E \exp(\theta|W - Z|) \leq 2E \exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right) \quad \text{for all } \theta \in \mathbb{R}.$$

Note bound is tight in that $T = \sigma^2$ if and only if $W \sim \mathcal{N}(0, \sigma^2)$, and that $\sigma^2 > 0$ is arbitrary, which will be important later on.

How does T yield a coupling?

T yields Marginal Coupling Bound

Theorem 1.2, Chatterjee 2012

Let W be mean zero with finite second moment and suppose that T is a Stein coefficient for W with $|T|$ almost surely bounded by a constant. Then, given any $\sigma^2 > 0$, we can construct a version of W and $Z \sim \mathcal{N}(0, \sigma^2)$ on the same probability space such that

$$E \exp(\theta|W - Z|) \leq 2E \exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right) \quad \text{for all } \theta \in \mathbb{R}.$$

Note bound is tight in that $T = \sigma^2$ if and only if $W \sim \mathcal{N}(0, \sigma^2)$, and that $\sigma^2 > 0$ is arbitrary, which will be important later on.

How does T yield a coupling?

Sampling a Random Walk

Sample a random walk by generating $\epsilon_1, \ldots, \epsilon_n$ independent variables with $\mathcal{L}(\epsilon)$ and summing

$$S_k = \sum_{i=1}^{k} \epsilon_i, \quad 1 \le k \le n.$$

Convoluted, but for our purposes a 'better' way to sample:

1. Sample the value S at the terminal time n of the walk.
2. Sample a multiset $\{\epsilon_1, \ldots, \epsilon_n\}$ of independent $\mathcal{L}(\epsilon)$ variables conditional on their sum being S.
3. Now vary about the line connecting $(0,0)$ to (S,n) by a discrete Brownian Bridge, that is, by sampling an independent, uniformly random permutation π, and forming

$$W_k = S_k - \frac{k}{n}S \quad \text{where} \quad S_k = \sum_{j=1}^{k} \epsilon_{\pi(j)}$$

Sampling a Random Walk

Sample a random walk by generating $\epsilon_1, \ldots, \epsilon_n$ independent variables with $\mathcal{L}(\epsilon)$ and summing

$$S_k = \sum_{i=1}^{k} \epsilon_i, \quad 1 \leq k \leq n.$$

Convoluted, but for our purposes a 'better' way to sample:

1. Sample the value S at the terminal time n of the walk.
2. Sample a multiset $\{\epsilon_1, \ldots, \epsilon_n\}$ of independent $\mathcal{L}(\epsilon)$ variables conditional on their sum being S.
3. Now vary about the line connecting $(0, 0)$ to (S, n) by a discrete Brownian Bridge, that is, by sampling an independent, uniformly random permutation π, and forming

$$W_k = S_k - \frac{k}{n} S \quad \text{where} \quad S_k = \sum_{j=1}^{k} \epsilon_{\pi(j)}$$

Coupling Overview

1. Couple the marginals (S, Z) at the terminal time n of the sum and the Brownian motion using Theorem 1.2.
2. Given S, sample multiset $\{\epsilon_1, \ldots, \epsilon_n\}$ according to the conditional probability that i.i.d. variables sum to S.
3. Given $\{\epsilon_1, \ldots, \epsilon_n\}$, couple

$$W_k = S_k - \frac{k}{n} S, \quad \text{where} \quad S_k = \sum_{j=1}^{k} \epsilon_{\pi(j)},$$

to a discrete time Brownian Bridge $\widetilde{Z}_1, \ldots, \widetilde{Z}_n$, i.e. a mean zero Gaussian vector with covariance $(i \wedge j)(n - i \vee j)/n$.

Can check that the processes

$$S_i = W_i + \frac{i}{n} S \quad \text{and} \quad Z_i = \widetilde{Z}_i + \frac{i}{n} Z$$

are coupled and have the correct marginals (and mgf bound).

Brownian Bridge, Induction

Step 3. Couple

$$W_k = S_k - \frac{k}{n} S, \quad \text{where} \quad S_k = \sum_{j=1}^{k} \epsilon_{\pi(j)},$$

to a discrete Brownian Bridge $\widetilde{Z}_1, \ldots, \widetilde{Z}_n$.

Induct on n, using Theorem 1.2 to couple W_k and $\widetilde{Z}_k$ at time $k = [n/2]$.

Rademacher

In the Rademacher case:

1. Sum $S_n = \epsilon_1 + \cdots + \epsilon_n$ determines $\{\epsilon_1, \ldots, \epsilon_n\}$.
2. Smoothing variable Y is known.
3. Hidden 'variance parameter' $\gamma^2 = n^{-1} \sum_{i=1}^n \epsilon_i^2$ is identically one, *and also for all subsets of variables.*

Rademacher

In the Rademacher case:

1. Sum $S_n = \epsilon_1 + \cdots + \epsilon_n$ determines $\{\epsilon_1, \ldots, \epsilon_n\}$.
2. Smoothing variable Y is known.
3. Hidden 'variance parameter' $\gamma^2 = n^{-1} \sum_{i=1}^n \epsilon_i^2$ is identically one, *and also for all subsets of variables.*

Relax assumptions to allow more general distributions.

Rademacher

In the Rademacher case:

1. Sum $S_n = \epsilon_1 + \cdots + \epsilon_n$ determines $\{\epsilon_1, \ldots, \epsilon_n\}$.
   Combinatorial accounting
2. Smoothing variable Y is known.
3. Hidden 'variance parameter' $\gamma^2 = n^{-1} \sum_{i=1}^n \epsilon_i^2$ is identically one, *and also for all subsets of variables.*

Relax assumptions to allow more general distributions.

Smoothing

Did not attain more: '...we do not know yet how to use Theorem 1.2 to prove the KMT theorem in its full generality, because we do not know how to generalize the smoothing technique of Example 3.'

Smoothing needed in Step 1 for coupling values at terminal time of random walk S_n and Brownian motion, and in Step 3 for coupling values at time $k = [n/2]$ of W_k and Brownian bridge.

Generalizing 'Example 3' to the 'smoothing lemma' is perhaps not the largest obstacle.

Smoothing

Did not attain more: '...we do not know yet how to use Theorem 1.2 to prove the KMT theorem in its full generality, because we do not know how to generalize the smoothing technique of Example 3.'

Smoothing needed in Step 1 for coupling values at terminal time of random walk S_n and Brownian motion, and in Step 3 for coupling values at time $k = [n/2]$ of W_k and Brownian bridge.

Generalizing 'Example 3' to the 'smoothing lemma' is perhaps not the largest obstacle.

Use of Smoothing to obtain T

For X Rademacher, there does not exist T such that

$$E[Xf(X)] = E[Tf'(X)].$$

But one can find an independent ‘smoothing’ Y such that

$$E[Xf(X+Y)] = E[(1-XY)f'(X+Y)].$$

Smoother for Step 1, terminal time of random walk.

If $X_1, \ldots, X_n$ are independent Rademacher, S_n their sum, and $S_n^{(i)} = S_n - X_i$,

$$E[S_n f(S_n + Y)] = \sum_{i=1}^{n} E[X_i f(X_i + Y + S_n^{(i)})]$$
$$= \sum_{i=1}^{n} E[(1 - X_i Y) f'(X_i + Y + S_n^{(i)})] = E[T f'(S_n + Y)]$$

where

$$T = \sum_{i=1}^{n} (1 - X_i Y).$$

Only added one variable, Y, to the sum.

Smoother for Step 1, terminal time of random walk.

If $X_1, \ldots, X_n$ are independent Rademacher, S_n their sum, and $S_n^{(i)} = S_n - X_i$,

$$E[S_n f(S_n + Y)] = \sum_{i=1}^{n} E[X_i f(X_i + Y + S_n^{(i)})]$$
$$= \sum_{i=1}^{n} E[(1 - X_i Y) f'(X_i + Y + S_n^{(i)})] = E[T f'(S_n + Y)]$$

where

$$T = \sum_{i=1}^{n} (1 - X_i Y).$$

Only added one variable, Y, to the sum.

Smoothing Lemma

When X is symmetric $\{+1, -1\}$, taking Y to be an independent $\mathcal{U}[-1, 1]$ yields

$$E[Xf(X+Y)] = E[(1-XY)f'(X+Y)].$$

What is Y for general X?

Smoothing Lemma

For X Rademacher the smoothing Y is $\mathcal{U}[-1,1]$ that yields

$$E[Xf(X+Y)] = E[(1-XY)f'(X+Y)].$$

What is it in general?

Smoothing Lemma: If X has mean zero and finite, non-zero variance, and Y is independent of X and has the X-zero bias distribution, then

$$E[Xf(X+Y)] = E[(X^2-XY)f'(X+Y)].$$

Exercise: use that

$$Y =_d UX^\square \quad \text{where} \quad U \sim \mathcal{U}[0,1] \quad \text{and} \quad \frac{dF^\square}{dF} = \frac{x^2}{EX^2}.$$

Smoothing in Step 3, Bridge at midpoint

The sum

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

may be rewritten as

$$W_k = \frac{1}{n} \sum_{1 \le i \le k < j \le n} \left(\epsilon_{\pi(j)} - \epsilon_{\pi(i)} \right).$$

Half differences of $\epsilon_{\pi(j)} - \epsilon_{\pi(i)} \in \{-2, 0, 2\}$, can be smoothed by a single $Y \sim \mathcal{U}[-1, 1]$ in Rademacher case.

More General Discrete Variable

When ϵ takes values in a finite set $\mathcal{A} \subset \mathbb{R}$, then

$$\epsilon_{\pi(j)} - \epsilon_{\pi(i)} \in \mathcal{D} := \mathcal{A} - \mathcal{A}.$$

Add independent smoothers $Y_d \sim \mathcal{U}[-d/2, d/2]$, one for each $d \in \mathcal{D} \cap (0, \infty)$.

Only finitely many soothers as $n \to \infty$.

Excludes $\mathcal{A}$ infinite and continuous variables.

Can one smooth in such cases using a finite number of smoothers as n tends to infinity?

Variance Parameter, Bridge

Rademacher case, KMT follows from a bound on

$$E[\exp(\lambda \max_{1 \le i \le n} |W_i - \sqrt{n} B_{i/n}|)]$$

Variance Parameter, Bridge

In general, KMT follows from a bound on

$$E[\exp(\lambda \max_{1 \le i \le n} |W_i - \gamma \sqrt{n} B_{i/n}|)]$$

where $\gamma^2 = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2$.

Variance Parameter, Bridge

In general, KMT follows from a bound on

$$E[\exp(\lambda \max_{1 \le i \le n} |W_i - \gamma \sqrt{n} B_{i/n}|)]$$

where $\gamma^2 = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2$.

For Rademacher, $\gamma^2 = 1$, and also for segments of the path standardized by their length.

Variance Parameter, Bridge

In general, KMT follows from a bound on

$$E[\exp(\lambda \max_{1 \le i \le n} |W_i - \gamma\sqrt{n}B_{i/n}|)]$$

where $\gamma^2 = \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2$.

For Rademacher, $\gamma^2 = 1$, and also for segments of the path standardized by their length.

Problem: In the induction step the variance parameters γ_1,γ_2 for the left and right segments of the path will not in general equal γ.

Effect of Variance Parameter

Rademacher variable bound on Laplace transform for $\lambda \leq \lambda_0$ of

$$\max_{0 \leq i \leq n} |W_i - \sqrt{n} B_{i/n}| \quad \text{of the form} \quad \exp(C \log n) E \exp\left(\frac{K \lambda^2 S_n^2}{n}\right),$$

where $S_n = \sum_{i=1}^n \epsilon_i$.

Effect of Variance Parameter

Rademacher variable bound on Laplace transform for $\lambda \leq \lambda_0$ of

$$\max_{0 \leq i \leq n} |W_i - \sqrt{n} B_{i/n}| \quad \text{of the form} \quad \exp(C \log n) E \exp\left(\frac{K\lambda^2 S_n^2}{n}\right),$$

where $S_n = \sum_{i=1}^n \epsilon_i$, becomes bound on

$$\max_{0 \leq i \leq n} |W_i - \sqrt{n}\eta B_{i/n}|,$$

of the form

$$\exp(C \log n) E \exp\left(\frac{K_1 \lambda^2 S_n^2}{n} + K_2 \lambda^2 n (\gamma^2 - \eta^2)^2\right).$$

Recall $\sigma^2 > 0$ in Theorem 1.2 is arbitrary.

Effect of Variance Parameter

Variance parameter γ_1^2 for first half of path has expectation η^2, parameter for entire path, yields term in induction

$$E\exp\left(\theta^2 k(\gamma_1^2-\eta^2)^2\right) = E\exp\left(\theta^2 k^{-1}\left(\sum_{i=1}^{k}(\epsilon_{\pi(i)}^2-\eta^2)\right)^2\right).$$

Expression is identically zero in the Rademacher case. Control in general using summand variables are negatively associated.

T yields marginal coupling

For $T:\mathbb{R}^n \to \mathbb{R}^{n\times n}$ p.s.d, uniformly bounded over $x \in \mathbb{R}^n$, for $\epsilon > 0$ and a measure μ let $\mathcal{T}_\epsilon \mu$ be the measure

$$(1-\epsilon)X + \sqrt{2\epsilon T(X)Z} \quad \text{for } X \sim \mu \text{ and } Z \sim \mathcal{N}(0, I) \text{ independent.}$$

Transformation $\mathcal{T}_\epsilon \mu$ has fixed point μ_ϵ, take $\epsilon \to 0$ obtain measure for a random vector X that satisfies

$$E[X \cdot \nabla f(X)] = E[\mathrm{Tr}\left\{T(X)D^2 f(X)\right\}].$$

So, given 'nice' $T(x)$, have random vector X for which $T(x)$ is the 'Stein matrix.'

Existence of marginal coupling

For such T and X, use the Stein matrix identity to obtain moment bounds

$$E(X_i - X_j)^{2k} \leq (2k-1)^k E v_{ij}(X)^k$$

where $v_{ij}(x) = t_{ii}(x) + t_{jj}(x) - 2t_{ij}(x)$, and thus

$$E\exp(\theta|X_i - X_j|) \leq 2E\exp(2\theta^2 v_{ij}(X)).$$

With $h(x_1)$ Stein coefficient of W, apply in $\mathbb{R}^2$ with

$$T(x) = \begin{pmatrix} h(x_1) & \sigma\sqrt{h(x_1)} \\ \sigma\sqrt{h(x_1)} & \sigma^2 \end{pmatrix},$$

for which $v_{12}(x) = h(x_1) + \sigma^2 - 2\sigma\sqrt{h(x_1)} = (\sqrt{h(x_1)} - \sigma)^2$.

Multivariate approach, Stein Matrix, Dependence

Marginal argument appears 'uncooperative' and 'brittle'.

Can avoid induction, relax conditions and handle dependence with multivariate version of Theorem 1.2 if given $X \in \mathbb{R}^n$ one could produce Stein matrix T such that

$$E[X \cdot \nabla f(X)] = E[\mathrm{tr}(T(X)D^2 f(X))].$$