KMT strong embedding theorems

Manjunath Krishnapur
IISc, Department of mathematics

Coupling Binomial with Gaussian

- Let $S_n = X_1 + \ldots + X_n$ with X_k i.i.d. ± 1 with equal probability
- Let Z be a standard Gaussian
- CLT $\Leftrightarrow$ Can couple so that $\left| S_n - Z\sqrt{n} \right| = o(\sqrt{n})$ a.s.

TUSNADY'S LEMMA. For large enough n, there is a coupling so that

$$|S_n| \le |Z|\sqrt{n} + 3 \quad \text{and} \quad |S_n - Z\sqrt{n}| \le Z^2 + 11$$

- Tusnády had better constants. Irrelevant for us. Power 2 on Z important
- $\mathbb{P}\{|S_n - Z\sqrt{n}| \ge x\} \le e^{-cx}$ which is far better than what CLT gave
- Key ingredient in the proof of KMT theorems

KMT for simple symmetric random walk

- $S = (S_0, S_1, \ldots)$ SSRW as before
- W a standard Brownian motion in one dimension
- Donsker's theorem $\Leftrightarrow$ Can couple so that $\max_{0 \le k \le n} |S_k - W_k| = o(\sqrt{n})$ a.s.

Komlós Major Tusnády (KMT-RW). There is a $C < \infty$ and a coupling so that

$$\max_{0 \le k \le n} |S_k - W_k| \le C(\log n + x) \quad \text{w.p. } \ge 1 - e^{-x}$$

- Improves upon the $\sqrt{n}$ in Donsker's theorem. Bound $\log n$ is optimal
- $\log n$ as opposed to $O(1)$ in univariate coupling
- Komlós, Major, Tusnády showed this for X_i with exponential tail

KMT for uniform empirical process

- $V_1, V_2, \ldots$ i.i.d. uniform[0,1]
- Empirical CDF: $F_n(t)$ = proportion of $k \le n$ for which $V_k \le t$
- Empirical process: $G_n(t) = \sqrt{n}(F_n(t) - t)$ for $0 \le t \le 1$
- Brownian bridge: $\overline{W}(t) = W_t - tW_1$ for $0 \le t \le 1$
- Donsker (Kolmogorov-Smirnov) implies coupling so that $\max\{\sqrt{n}\,|\,G_n(t) - \overline{W}(t)\,| \; : \; 0 \le t \le 1\} = o(\sqrt{n})$ a.s.

Komlos Major Tusnády (KMT-EP). There is a $C < \infty$ and a coupling so that

$$\max_{0 \le t \le 1} \sqrt{n}\,|\,G_n(t) - \overline{W}(t)\,| \le C(\log n + x) \quad \text{w.p. } \ge 1 - e^{-x}$$

Broad outline of the proof

Proof strategies

Combinatorial: KMT, Csörgõ-Révész, Bretagnolle and Massart, Dudley, Massart, Carter—Pollard, Pollard (both KMT-RW and KMT-EP). Excellent reference - Pollard's book **UGMPT**

Analytic: Chatterjee (KMT-RW for SSRW). Extension by Bhattacharjee and Goldstein.

Steps in both kinds of proof

1. Univariate coupling lemmas such as Tusnády's lemma or other similar coupling lemmas (will see in second lecture). Combinatorial proof compares Binomial to Gaussian by Stirling's etc. (but way more involved). Analytic proof uses a form of *Stein's method*.

2. Go from univariate coupling to coupling of paths. In the dyadic approach, couple $(G_n(1/2), \overline{W}(1/2))$, $(G_n(1/4), \overline{W}(1/4))$, $(G_n(3/4), \overline{W}(3/4))$, … for $\log n$ generations. Chatterjee craftily frames a statement that can be proved by induction on n.

An idea: Use the Cauchy criterion

- To show $x_n \to x$, enough to show $|x_n - x_{n+1}| \leq \dfrac{1}{n^2}$ or $|x_n - x_{2n}| \leq \dfrac{1}{n^\varepsilon}$. Then,
$$|x_n - x| \leq \sum_{j \geq n} |x_j - x_{j+1}| \lesssim \frac{1}{n} \text{ or } |x_n - x| \leq \sum_{j \geq 0} |x_{n2^j} - x_{n2^{j+1}}| \lesssim \frac{1}{n^\varepsilon}$$

- Analogously, to show closeness of S_n to $Z\sqrt{n}$ (*like with unlike*), suffices to show closeness of $2S_n$ to S_{4n} (*like with like*).

- As $S_m \sim$ Binomial, hope for purely combinatorial proof (this lecture) or using finite Markov chains (second lecture)

Our version of the combinatorial proof

Tusnády-like lemma

TUSNADY TYPE LEMMA. For large enough even n, there is a coupling so that

$$|2S_n| \le |S_{4n}| + 2 \quad \text{and} \quad |2S_n - S_{4n}| \le \frac{|S_{4n}|^2}{8n} + 9$$

Deducing Tusnády's lemma

- Fix a large enough even number n and let $Z_0 = \frac{S_n}{\sqrt{n}}, Z_1 = \frac{S_{4n}}{\sqrt{4n}}, \ldots, Z_k = \frac{S_{4^k n}}{\sqrt{4^k n}}, \ldots$
- Successively couple so that $|Z_k| \le |Z_{k+1}| + \frac{2}{2^k\sqrt{n}}$ and $|Z_k - Z_{k+1}| \le \frac{1}{2^{k+2}\sqrt{n}}|Z_{k+1}|^2 + \frac{9}{2^{k+1}\sqrt{n}}$
- By Bernstein/Hoeffding, $|Z_k| \le 2^{k/10}$ with high probability.
- Hence, a.s., $\{Z_k\}$ is Cauchy and $Z_k \to Z \sim N(0,1)$
- Summing, $|Z_k| \le |Z| + \frac{2}{\sqrt{n}}$ and $|Z_k - Z| \le \frac{1}{4\sqrt{n}}Z^2 + \frac{9}{\sqrt{n}}$ which is Tusnády's lemma (for even n)

How to couple $2S_n$ and S_{4n}? Preparation

FACT (easy to prove)

X, Y - $\mathbb{Z}_+$-valued. Assume $\mathbb{P}\{X \geq k - f(k)\} \geq \mathbb{P}\{Y \geq k\}$ and $\mathbb{P}\{Y \geq k - g(k)\} \geq \mathbb{P}\{X \geq k\}$ for all k.
Then, they can be coupled so that $X \geq Y - f(Y)$ and $Y \geq X - g(X)$.

- Towards the goal of coupling $2S_n$ with S_{4n}, let $n = 2m$.
- Let $\alpha_m(k) = \mathbb{P}\{S_{2m} = 2k\}$ and $\beta_m(k) = \mathbb{P}\{S_{8m} = 4k \text{ or } 4k-2\}$. PMFs of $X = \frac{S_{2m}}{2}$ and $Y = \left[\frac{S_{8m}}{4}\right]_+$
- Let $\bar{\alpha}_m(k) = \sum_{j \geq k} \alpha(j)$ and $\bar{\beta}_m(k) = \sum_{j \geq k} \beta_m(j)$ be the tails
- From the fact, it suffices to prove that

$$\bar{\alpha}_m(k) \leq \bar{\beta}_m(k) \text{ for all } k \quad \text{and} \quad \bar{\alpha}_m(k) \geq \bar{\beta}_m(\ell) \text{ for } k \leq \ell - \frac{\ell^2}{2m} - 1$$

Because $g(k) = 0$ and $f(\ell) = \frac{\ell^2}{2m} + 1$

How to couple $2S_n$ and S_{4n}? Reduction

- $\bar{\alpha}_m(k) = \sum_{j \geq k} \alpha(j)$ where $\alpha_m(k) = \mathbb{P}\{S_{2m} = 2k\} = \binom{2m}{m+k} \frac{1}{2^{2m}}$
- $\bar{\beta}_m(k) = \sum_{j \geq k} \beta_m(j)$ where $\beta_m(k) = \mathbb{P}\{S_{8m} = 4k \text{ or } 4k-2\} = \binom{8m+1}{4m+2k} \frac{1}{2^{8m}}$
- To show (symmetry etc.): $\bar{\alpha}_m(k) \leq \bar{\beta}_m(k)$ for all k and $\bar{\alpha}_m(k) \geq \bar{\beta}_m(\ell)$ for $k \leq \ell - \frac{\ell^2}{2m} - 1$

Binomial tails don't have simple closed form expressions. But we do have estimates such as

$$A\frac{\sqrt{n}}{\sqrt{k(n-k)}}e^{-nD(k/n)} \leq \frac{1}{2^n}\binom{n}{k} \leq B\frac{\sqrt{n}}{\sqrt{k(n-k)}}e^{-nD(k/n)}$$

$D(p) = p\log(2p) + (1-p)\log(2-2p)$

$$A\frac{\sqrt{n}}{\sqrt{k(n-k)}}e^{-nD(k/n)} \leq \sum_{j=k}^{n} \frac{1}{2^n}\binom{n}{j} \leq e^{-nD(k/n)}$$

These only suffice for comparison in the far tail

How to couple $2S_n$ and S_{4n}? The key lemma

Key Combinatorial Lemma. $\alpha_m(k) \le \beta_m(k)$ for all k and $\alpha_m(k) \ge \beta_m(\ell)$ for $k \le \ell - \dfrac{\ell^3}{4m^2} - 1$

- Compares mass functions instead of tails
- Must be combined with estimates for binomial coefficients to complete the proof
- As $\dfrac{\ell^3}{m^2} \lesssim \dfrac{\ell^2}{m}$ it suggests Carter-Pollard improvement: $\dfrac{1}{n}|Z|^3 \le Z^2$ in Tusnády's lemma bound
- Bijective proof? (see next slide)

Proof of the key combinatorial lemma

Key Combinatorial Lemma. $\alpha_m(k) \le \beta_m(k)$ for all k and $\alpha_m(k) \ge \beta_m(\ell)$ for $k \le \ell - \dfrac{\ell^3}{4m^2} - 1$

Let $f(m,k) := \dfrac{\beta_m(k)}{\alpha_m(k)} = \dfrac{\binom{8m+1}{4m+2k}}{2^{6m}\binom{2m}{m+k}}$ and $g_h(m,k) := \dfrac{\beta_m(k)}{\alpha_m(k-h)} = \dfrac{\binom{8m+1}{4m+2k}}{2^{6m}\binom{2m}{m+k-h}}$

Lemma follows from following claims

$f(m,k) \ge 1$

1. $f(m,k+1) \ge f(m,k)$ for $1 \le k \le m-1$ and $m \ge 1$
2. $f(m,1) \ge f(m+1,1)$ for $m \ge 1$
3. $f(m,1) \to 1$ as $m \to \infty$

$g_h(m,k) \le 1$ if $h+1 \le k \le [(4h-1)m^2]^{1/3}$

1. $g_h(m,k+1) \le g_h(m,k)$ for $h+1 \le k \le [(4h-1)m^2]^{1/3}$
2. $g_h(m,h+1) \ge g_h(m+1,h+1)$ for $m \ge h+1$
3. $g_h(m,h+1) \to 1$ as $m \to \infty$ for h fixed

Proof of the key combinatorial lemma

Proof that $f(m,k) \geq 1$

$f(m,1) \to 1$ as $m \to \infty$ is trivial (Stirling's or invoke local limit theorem).

In the other two steps, we get to take ratios and cancel most of the factorials

$$f(m,1) = \frac{\binom{8m+1}{4m+2}}{2^{6m}\binom{2m}{m+1}}$$

To show that $f(m,1) \geq f(m+1,1)$, compute $f(m+1,1)/f(m,1)$ to get

$$\frac{m(m+2) \times (8m+2)\ldots(8m+9)}{2^6(2m+1)(2m+2) \times (4m+3)\ldots(4m+6) \times (4m)\ldots(4m+3)} = \frac{p(x)}{q(x)}$$

where $x = 8m$ and

$p(x) = (x+16)(x+9)(x+7)(x+5)(x+3)$

$q(x) = (x+12)(x+10)(x+8)(x+6)(x+4)$

Majorization explained

16>**12**

16+9=**25**>**22**=12+10

16+9+7=**32**>**30**=12+10+8

16+9+7+5=**37**>**36**=12+10+8+6

16+9+7+5+3=**40**=12+10+8+6+4

(16,9,7,5,3) majorizes (12,10,8,6,4), hence $p(x) \leq q(x)$ for $x > 0$ by *Schur concavity*

Proof of the key combinatorial lemma

Remains to show that $f(m, k+1) \geq f(m, k)$ for $1 \leq k \leq m-1$. Again take ratios

$$\frac{f(m,k+1)}{f(m,k)} = \frac{(x+2a+4)(x-a)(x-a+1)}{(x-2a)(x+a+1)(x+a+2)} \quad \text{with } x = 4m,\ a = 2k$$

Do not see any majorization, but Numerator-Denominator gives

$$2x(x-2a) + 2x + 4a^2(a+2) \geq 0$$

since $x - 2a = 4(m-k) > 0$

Remark. It is in this step for the second claim that we get $g_h(m, k+1) \leq g_h(m, k)$ provided

$$2((4h-1)x^2 - 2a^3) + 2x(8h + 2a - 1) + 8(a^2(h-1) + ah + h) \geq 0$$

Second and third terms are positive. For the first to be positive, require $k^3 \lesssim (4h-1)m^2$

Our version of the analytic proof

KMT for simple symmetric random walk

- $S = (S_0, S_1, \ldots)$ SSRW
- W a standard Brownian motion in one dimension
- Donsker's theorem $\Leftrightarrow$ Can couple so that $\max_{0 \le k \le n} |S_k - W_k| = o(\sqrt{n})$ a.s.

Komlós Major Tusnády (KMT-RW). There is a $C < \infty$ and a coupling so that

$$\max_{0 \le k \le n} |S_k - W_k| \le C(\log n + x) \quad \text{w.p.} \ \ge 1 - e^{-x}$$

- Improves from $\sqrt{n}$ in Donsker's theorem to $\log n$ which is optimal
- Komlós, Major, Tusnády showed this for X_i with exponential tail

Chatterjee's approach to KMT for SSRW

Univariate coupling lemmas In place of Tusnády's lemma, Chatterjee proves two lemmas coupling binomial and hypergeometric to Gaussian.

Subject of the rest of the talk

From univariate coupling to that of paths Induction on n (length of the random walk) applied to the following statement: For any probable value t of S_n, there is a coupling of the random walk conditioned to have $S_n = t$ (a random walk bridge) with Brownian motion conditioned to have $W(n) = t$ (Brownian bridge) such that the maximum difference $M_{n,t}$ between the two paths satisfies

$$\mathbb{E}[e^{\lambda M_{n,t}}] \leq e^{A \log n + B\lambda^2 \frac{t^2}{n}} \text{ for } \lambda \leq \lambda_0 .$$

Here λ_0, A, B are fixed constants.

We do not talk about this second step

Coupling lemma for Binomial

- Let $S_k = X_1 + \ldots + X_k$ where X_j are i.i.d. ± 1. SRSWR of k coupons from a box with equal number of coupons labelled $+1$ and -1

Chatterjee's binomial coupling lemma. For some θ_0, κ_0, there is a coupling for any $n \geq 1$ such that $\mathbb{E}\left[e^{\theta_0 |S_n - Z\sqrt{n}|}\right] \leq \kappa_0$.

- Like in Tusnády's lemma, two random variables of std. dev. $\sqrt{n}$ are coupled within unit distance
- Tusnády's lemma implies this because $\mathbb{E}[e^{\theta_0 Z^2}] < \infty$ for small enough θ_0
- Converse not possible, as largeness of the difference not related with the value of Z (or S_n)

Coupling lemma for hypergeometric

- Let $S_k[n,s]$ be the sum of k coupons drawn without replacement from a box containing n coupons labelled ± 1 whose sum is s i.e., $(n \pm s)/2$ coupons of either kind. (Hypergeometric)
- Let $W_k[n,s] = S_k[n,s] - \frac{k}{n}s$ be the centred version. Its variance is $\left(1 - \frac{s^2}{n^2}\right)\frac{k(n-k)}{n-1}$
- Let $\sigma^2_{n,k} = \dfrac{k(n-k)}{n-1}$ be the variance of $W_k[n,0]$ (case of unbiased box)

Chatterjee's hypergeometric coupling lemma. For some θ_1, M_1, there is a coupling for $n \geq 1$ and $\frac{1}{3}n \leq k \leq \frac{2}{3}n$ such that $\mathbb{E}\left[e^{\theta|W_k[n,s]-\sigma_{n,k}Z|}\right] \leq e^{1+M_1\theta^2\frac{s^2}{n}}$ for $\theta \leq \theta_1$.

- Supplements binomial coupling and feeds into the induction step (which is about bridges)
- Enough to have $k \approx \frac{n}{2}$ (nearest integer, for eg.)

Univariate couplings using Cauchy criterion

From the following lemmas, it is easy to deduce Chatterjee's coupling lemmas, using the Cauchy criterion as in the first part of the talk

BINOMIAL COUPLING LEMMA. For some θ_0, κ_0, there is a coupling for any $n \geq 1$ such that

$$\mathbb{E}\left[e^{\theta_0 |2S_n - S_{4n}|}\right] \leq \kappa_0$$

HYPERGEOMETRIC COUPLING LEMMA. For some θ_1, κ_1, M_1, for even n and $\frac{1}{3}n \leq k \leq \frac{2}{3}n$, with $W_1 = W_k[n,0]$, $W_2 = W_{4k}[4n,0]$, $W = W_k[n,s] - \frac{sk}{n}$, there are couplings such that

(1) $\mathbb{E}\left[e^{\theta_1 |2W_1 - W_2|}\right] \leq \kappa_1$

(2) $\mathbb{E}\left[e^{\theta |W_1 - W|}\right] \leq e^{1 + M_1 \theta^2 \frac{s^2}{n}}$ for any $\theta \leq \theta_1$

Coupling distributions on $\mathbb{Z}$ using Markov chains

The general strategy of coupling

- α, β - probability distributions on $\mathbb{Z}$ that we want to couple. Eg., $2S_n$, S_{4n}
- We shall need nearest neighbour Markov chains on $\mathbb{Z}$ with these stationary distributions. This means no gaps in support
- Examples: $\frac{1}{2}S_{4n}$ has support $[-2n, 2n] \cap \mathbb{Z}$ - Good!
  But S_n has gaps. Modify to $S_n + R$ where R is an independent r.v. taking values -1,0,1 w.p. $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$. Enough to couple $S_n + R$ with $\frac{1}{2}S_{4n}$
- Similar games (scaling, additive perturbation) for hypergeometric coupling

The general strategy of coupling

- Assume that α has **finite connected support** in $\mathbb{Z}$. Many choices of rates $\lambda_i^{\pm}$ for $i \mapsto i \mp 1$
- **Ehrenfest-like chain**: If α is reversible for $\lambda_i^{\pm} = T(i) \mp i$ for some $T : \mathbb{Z} \mapsto \mathbb{R}$
- Does it exist? The equations $\alpha(i)(T(i) - i) = \alpha(i+1)(T(i+1) + i + 1)$ can be solved from right end of support to get $T(i) = i + \dfrac{2}{\alpha(i)} \sum_{j>i} j\alpha(j)$
- Satisfied at left end if and only if α has **zero mean**
- **T is called the Stein coefficient of α**

Example: If α is the distribution of $S_{4n}/2$, then $T(i) = 2n$ for $-2n \leq i \leq 2n$ (The true Ehrenfest chain)

Example: If β is the distribution of $S_n + R$, then

$$T(j) = \begin{cases} 2n+1 & \text{if } j \in \{-n, -n+2, \ldots, n-2, n\} \\ 2n + 2 - \dfrac{j^2}{n+1} & \text{if } j \in \{-n-1, -n+1, \ldots, n-1, n+1\} \end{cases}$$

Relationship to Stein's method

- Stein's method: $\mathbb{E}[Wf(W)] = \mathbb{E}[\sigma^2 f'(W)]$ for a large class of f if and only if $W \sim N(0,\sigma^2)$
- The Ornstein-Uhlenbeck process has generator $Lg(x) = g''(x) - xg'(x)$ and stationary distribution $\pi = N(0,1)$. Stein's equation is $\mathbb{E}_\pi[Lg] = 0$, with $f = g'$
- If a random variable W satisfied $\mathbb{E}[Wf(W)] = \mathbb{E}[Tf'(W)]$, Chatterjee calls T the Stein coefficient of W. Closeness of T to constant related to closeness of W to normal.

- Stein's equation for Binomial: $\mathbb{E}[n(g(W+2) - g(W))] = \mathbb{E}[W(g(W+2) - g(W))]$ for a large class of g if and only if $W \sim S_n$
- Ehrenfest chain has generator $Lg(i) = (n-i)(g(i+2) - g(i)) + (n+i)(g(i-2) - g(i))$ and stationary distribution $\pi = \mathscr{L}(S_n)$. Stein's equation is $\mathbb{E}_\pi[Lh] = 0$, with $h(i) = f(i) - f(i-2)$
- If a random variable W satisfied $\mathbb{E}[T(g(W+2) - g(W))] = \mathbb{E}[W(g(W+2) - g(W))]$, we call T the Stein coefficient of W. Closeness of T to constant related to closeness of W to normal.
- Equivalently, $\mathscr{L}(W)$ is stationary for a Markov chain with rates $\lambda_i^\pm = T(i) \mp i$

Coupling nearest neighbour Markov chains on $\mathbb{Z}$

- Let α, β be stationary distributions for rates $\lambda_i^{\pm}, \mu_i^{\pm}$ (n.n, continuous times chains on $\mathbb{Z}$)
- Define a Markov chain on $\mathbb{Z}^2$ with rates as follows. Basic idea: Try to move together as much as possible. Never move in opposing directions
- Generator

$$\begin{aligned} Lf(i,j) = {} & \theta_{i,j}^{+,+}(f(i+1,j+1) - f(i,j)) + \theta_{i,j}^{+,\circ}(f(i+1,j) - f(i,j)) + \theta_{i,j}^{\circ,+}(f(i,j+1) - f(i,j)) \\ & +\theta_{i,j}^{-,-}(f(i-1,j-1) - f(i,j)) + \theta_{i,j}^{-,\circ}(f(i-1,j) - f(i,j)) + \theta_{i,j}^{\circ,-}(f(i,j-1) - f(i,j)) \end{aligned}$$

$$\theta_{i,j}^{\circ,+} = (\mu_j^+ - \lambda_i^+)_+ \qquad \theta_{i,j}^{+,+} = \lambda_i^+ \wedge \mu_j^+$$

$$\theta_{i,j}^{-,\circ} = (\lambda_i^- - \mu_j^-)_+ \qquad (i,j) \qquad \theta_{i,j}^{+,\circ} = (\lambda_i^+ - \mu_j^+)_+$$

$$\theta_{i,j}^{-,-} = \lambda_i^- \wedge \mu_j^- \qquad \theta_{i,j}^{\circ,-} = (\mu_j^- - \lambda_i^-)_+$$

Coupling nearest neighbour Markov chains on $\mathbb{Z}$

- Let $Z = (X, Y)$ denote the Markov chain on $\mathbb{Z}^2$ with generator
$$Lf(i,j) = \theta^{+,+}_{i,j}(f(i+1,j+1) - f(i,j)) + \theta^{+,\circ}_{i,j}(f(i+1,j) - f(i,j)) + \theta^{\circ,+}_{i,j}(f(i,j+1) - f(i,j))$$
$$+\theta^{-,-}_{i,j}(f(i-1,j-1) - f(i,j)) + \theta^{-,\circ}_{i,j}(f(i-1,j) - f(i,j)) + \theta^{\circ,-}_{i,j}(f(i,j-1) - f(i,j))$$
- If γ is a stationary distribution of Z, then $\mathbb{E}_\gamma[Lf(Z)] = 0$ for all $f : \mathbb{Z}^2 \mapsto \mathbb{R}$
- $f(x,y) = \varphi(x)$: $Lf(i,j) = \lambda_i^+(\varphi(i+1) - \varphi(i)) + \lambda_i^-(\varphi(i-1) - \varphi(i)) \Longrightarrow$ First marginal of γ is α
- $f(x,y) = \psi(y)$: $Lf(i,j) = \mu_j^+(\psi(j+1) - \psi(j)) + \mu_j^-(\psi(j-1) - \psi(j)) \Longrightarrow$ Second marginal of γ is β
- Let $H = X - Y$. If $f(x,y) = \varphi(x-y)$ then $\mathbb{E}_\gamma[Lf(Z)] = 0$ reduces to
$$\mathbb{E}_\gamma[(B - |A|)(\psi(H) - \psi(H-1)] = 2\mathbb{E}_\gamma[A\_\psi(H-1) - A_+\psi(H)] \qquad \text{— (\#)}$$
$$\psi(h) = \varphi(h+1) - \varphi(h) \qquad A(i,j) = \frac{1}{2}(\lambda_i^+ - \mu_j^+ - \lambda_i^- + \mu_j^-) \qquad B(i,j) = \frac{1}{2}(|\lambda_i^+ - \mu_j^+| + |\lambda_i^- - \mu_j^-|)$$

Coupling Ehrenfest-like chains

- **Ehrenfest-like chains:** $\lambda_i^\pm = S(i) \mp i$ and $\mu_j^\pm = T(j) \mp j$. Then
$$A(i,j) = j - i \text{ and } B(i,j) = |\,i - j\,| + (\,|\,T(j) - S(i)\,| - |\,j - i\,|\,)_+$$
- Writing $Q = |\,S(X) - T(Y)\,|$, (#) becomes
$$\mathbb{E}_\gamma[(Q - |\,H\,|)_+(\psi(H) - \psi(H-1))] = 2\mathbb{E}_\gamma[H_+\psi(H-1) - H_-\psi(H)]$$
- Use? Put $\psi(x) = \mathbf{1}_{x \geq a}$ to get $\mathbb{E}_\gamma[H_+\mathbf{1}_{H \geq a+1}] \leq \mathbb{E}_\gamma[(Q-a)_+]$ or $\mathbb{P}_\gamma[H \geq a+1\} \leq \mathbb{E}_\gamma[(Q-a)_+]$

Example: $X = \dfrac{1}{2}S_{4n}$ and $Y = S_n + R$ where $R \sim \dfrac{1}{2}S_2$ is an independent copy. Then
$$S(i) = 2n \text{ for } \;|\,i\,| \leq n \quad \text{and } T(j) = \begin{cases} 2n+1 & \text{if } j \in \{-n, -n+1, \ldots, n-1, n\} \\ 2n + 2 - \dfrac{j^2}{n+1} & \text{if } i \in \{-n-1, -n+1, \ldots, n-1, n+1\} \end{cases}$$
Here $Q \leq 2 + \dfrac{Y^2}{n+1}$ which has exponential tail. Immediately gives *Binomial coupling lemma*.

Coupling Ehrenfest-like chains

- General bound. With $Q = |S(X) - T(Y)|$ and $H = X - Y$, for any function ψ
$$\mathbb{E}_\gamma[(Q - |H|)_+(\psi(H) - \psi(H-1))] = 2\mathbb{E}_\gamma[H_+\psi(H-1) - H_-\psi(H)]$$

- Exponential tail on H not good enough for hypergeometric coupling lemma. Better bound: For some universal constants K, κ, Θ (the quadratic θ^2 in the exponent is important)
$$\mathbb{E}\left[e^{\theta|H|}\right] \leq K\,\mathbb{E}\left[e^{\kappa\theta^2 Q}\right] \text{ for } \theta \leq \Theta \qquad \text{— (*)}$$

If f is convex, its dual is
$f^\dagger(y) := \sup\{xy - f(x)\}$
Clearly $\forall\, x, y,$
$xy \leq f(x) + f^\dagger(y)$

- **Proof of (*):** Taking $\psi(x) = e^{\theta(x+1)}\mathbf{1}_{x>0}$ and $\psi(x) = e^{\theta(-x+1)}\mathbf{1}_{x<0}$ and add to get
$$2\mathbb{E}_\gamma[|H|e^{\theta|H|}] \leq (e^\theta - 1)\mathbb{E}_\gamma[Qe^{\theta|H|}]$$
$$\leq (e^\theta - 1)\left\{\beta\theta\mathbb{E}\left[|H|e^{\theta|H|}\right] + \beta e^{-1}\mathbb{E}\left[e^{Q/\beta}\right]\right\}$$
because $\beta x \log x$ and $\beta e^{-1+\frac{x}{\beta}}$ are Legendre convex duals of each other. Choose $\beta = 1/\theta^2$ and rearrange.
Use for small enough θ to get $\mathbb{E}_\gamma[\theta|H|e^{\theta|H|}] \leq K\mathbb{E}_\gamma[e^{\kappa\theta^2 Q}]$. Now use $e^x \leq e^b + \frac{1}{b}xe^x$ with any b. ∎

- The essential point is this: In the Markov coupling, $H = X - Y$ is controlled by $Q = |S - T|$ (the difference in Stein coefficients)

Binomial coupling

- Let $S_{n,p} \sim \text{Binomial}(n,p)$ (centered) taking values in $\{-np, -np+1, \ldots, nq-1, nq\}$
- It has Stein coefficient $T(x) = 2pqn + (q-p)x$
- $X = S_{4n,\frac{1}{2}}$ has $T_X(x) = 2n$ and $Y = 2S_{n,\frac{1}{2}} + R$ has $T_Y(y) = 2n + O(y^2/n)$. Thus $Q \leq \dfrac{Y^2}{n}$ in this case. Gives the Binomial coupling lemma.
- $X = S_{n,\frac{1}{2}}$ has $T_X(x) = n/2$ and $Y = S_{n,p}$ has $T_Y(y) = 2pqn + (q-p)y$. Thus $Q \leq \dfrac{n}{2}|1-4pq| + |p-q|Y$ which is bounded by $\dfrac{s^2}{n} + \dfrac{s}{\sqrt{n}}\dfrac{|Y|}{\sqrt{n}}$ with $s = n(p-q)$
- In the Markov chain coupling $\mathbb{E}[e^{\theta|H|}] \leq \mathbb{E}[e^{\kappa\theta^2 Q}] \leq e^{c\theta^2\frac{s^2}{n}}$

Hypergeometric coupling lemma

HYPERGEOMETRIC COUPLING LEMMA. For some θ_1, κ_1, M_1, for even n and $\frac{1}{3}n \leq k \leq \frac{2}{3}n$, with $W_1 = W_k[n,0]$, $W_2 = W_{4k}[4n,0]$, $W = W_k[n,s] - \frac{sk}{n}$, there are couplings such that

(1) $\mathbb{E}\left[e^{\theta_1|2W_1 - W_2|}\right] \leq \kappa_1$

(2) $\mathbb{E}\left[e^{\theta|W_1 - W|}\right] \leq e^{1+M_1\theta^2\frac{s^2}{n}}$ for any $\theta \leq \theta_1$

- Entirely analogous to the previous situation $(W_1, W_2, W) \Leftrightarrow (S_{n,\frac{1}{2}}, S_{4n,\frac{1}{2}}, S_{n,p})$ with $s = n(p-q)$
- Can compute explicitly the Stein coefficients of W_1, W_2, W
- Bounding exponential moments of Q made easy by a lemma of Hoeffding that says $\mathbb{E}[f(W)] \leq \mathbb{E}[f(S_{n,p})]$ for any convex function f

Thank you for listening!