Quantitative CLTs via Martingale Embeddings

Dan Mikulincer

Weizmann Institute of Science
Joint work with Ronen Eldan and Alex Zhai

The central limit theorem

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. copies of a random vector X in $\mathbb{R}^d$ with

$$\mathbb{E}[X] = 0 \text{ and } \mathrm{Cov}(X) = \Sigma.$$

If $S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ and $G \sim \mathcal{N}(0, \Sigma)$ then

$$S_n \underset{n \to \infty}{\longrightarrow} G,$$

in an appropriate sense.

- We usually normalize X to be isotropic, that is, $\Sigma = \mathrm{I}_d$.
- We are interested in bounding the convergence rate.

The central limit theorem

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. copies of a random vector X in $\mathbb{R}^d$ with

$$\mathbb{E}[X] = 0 \text{ and } \mathrm{Cov}(X) = \Sigma.$$

If $S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ and $G \sim \mathcal{N}(0, \Sigma)$ then

$$S_n \underset{n \to \infty}{\longrightarrow} G,$$

in an appropriate sense.

- We usually normalize X to be isotropic, that is, $\Sigma = \mathrm{I}_d$.
- We are interested in bounding the convergence rate.

The central limit theorem

Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. copies of a random vector X in $\mathbb{R}^d$ with

$$\mathbb{E}[X] = 0 \text{ and } \mathrm{Cov}(X) = \Sigma.$$

If $S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ and $G \sim \mathcal{N}(0, \Sigma)$ then

$$S_n \underset{n\to\infty}{\longrightarrow} G,$$

in an appropriate sense.

- We usually normalize X to be isotropic, that is, $\Sigma = \mathrm{I}_d$.
- We are interested in bounding the convergence rate.

Quantitative central limit theorem

Berry-Esseen is an early examples of a quantitative bound.

Theorem (Berry-Esseen)

In the 1*-dimensional case, for any* $t \in \mathbb{R}$,

$$|\mathbb{P}(S_n \leq t) - \mathbb{P}(G \leq t)| \leq \frac{\mathbb{E}\left[|X|^3\right]}{\sqrt{n}}.$$

This estimate is sharp.

Quantitative central limit theorem

Berry-Esseen is an early examples of a quantitative bound.

Theorem (Berry-Esseen)

In the 1*-dimensional case, for any* $t \in \mathbb{R}$,

$$|\mathbb{P}(S_n \leq t) - \mathbb{P}(G \leq t)| \leq \frac{\mathbb{E}\left[|X|^3\right]}{\sqrt{n}}.$$

This estimate is sharp.

Quantitative central limit theorem

In higher dimensions the current best known result is due to Bentkus.

Theorem (Bentkus, 2003)

In the d-dimensional case, for any convex set $K \subset \mathbb{R}^d$,

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{1}{4}} \mathbb{E}\left[||X||^3\right]}{\sqrt{n}}.$$

- The $d^{\frac{1}{4}}$ term is the maximal Gaussian surface area of a convex set in $\mathbb{R}^d$. If K^ε is the ε enlargement of K then
$$\mathbb{P}(G \in K^\varepsilon \setminus K) \leq 4\varepsilon d^{\frac{1}{4}}.$$
- Whether one can omit $d^{\frac{1}{4}}$ remains an open question.

Quantitative central limit theorem

In higher dimensions the current best known result is due to Bentkus.

Theorem (Bentkus, 2003)

In the d-dimensional case, for any convex set $K \subset \mathbb{R}^d$,

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{1}{4}}\mathbb{E}\left[||X||^3\right]}{\sqrt{n}}.$$

- The $d^{\frac{1}{4}}$ term is the maximal Gaussian surface area of a convex set in $\mathbb{R}^d$. If K^ε is the ε enlargement of K then

$$\mathbb{P}(G \in K^\varepsilon \setminus K) \leq 4\varepsilon d^{\frac{1}{4}}.$$

- Whether one can omit $d^{\frac{1}{4}}$ remains an open question.

Quantitative central limit theorem

In higher dimensions the current best known result is due to Bentkus.

Theorem (Bentkus, 2003)

In the d-dimensional case, for any convex set $K \subset \mathbb{R}^d$,

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{1}{4}} \mathbb{E}\left[||X||^3\right]}{\sqrt{n}}.$$

- The $d^{\frac{1}{4}}$ term is the maximal Gaussian surface area of a convex set in $\mathbb{R}^d$. If K^ε is the ε enlargement of K then

$$\mathbb{P}(G \in K^\varepsilon \setminus K) \leq 4\varepsilon d^{\frac{1}{4}}.$$

- Whether one can omit $d^{\frac{1}{4}}$ remains an open question.

Other metrics

We consider stronger notions of distance

Definition (Relative entropy between X and G)

$$\mathrm{Ent}(X||G) := \mathbb{E}[\ln(f(X))],$$

where f is the density of X with respect to G.

Definition (Wasserstein distance between X and G)

$$\mathcal{W}_2(X, G) := \inf_{\pi} \left\{ \mathbb{E}_{\pi} \left[||X - G||^2 \right] \right\}^{1/2}$$

where π ranges over all possible couplings of X and G.

Other metrics

We consider stronger notions of distance

Definition (Relative entropy between X and G)

$$\mathrm{Ent}(X||G) := \mathbb{E}[\ln(f(X))],$$

where f is the density of X with respect to G.

Definition (Wasserstein distance between X and G)

$$\mathcal{W}_2(X, G) := \inf_{\pi} \left\{ \mathbb{E}_{\pi} \left[||X - G||^2 \right] \right\}^{1/2}$$

where π ranges over all possible couplings of X and G.

Relative entropy

For relative entropy, if $A \subset \mathbb{R}^d$ is any measurable set, then by Pinsker's inequality,

$$|\mathbb{P}(S_n \in A) - \mathbb{P}(G \in A)| \le \sqrt{\mathrm{Ent}(S_n || G)}.$$

- In 84' Barron showed that if $\mathrm{Ent}(X||G) < \infty$ then

$$\lim_{n\to\infty} \mathrm{Ent}(S_n||G) = 0.$$

- In 2011, Bobkov, Chistyakov and Götze showed that if, in addition, X has a finite fourth moment then

$$\mathrm{Ent}(S_n||G) \le \frac{C}{n}.$$

- The above constant may depend on X as well as the dimension.

Relative entropy

For relative entropy, if $A \subset \mathbb{R}^d$ is any measurable set, then by Pinsker's inequality,

$$|\mathbb{P}(S_n \in A) - \mathbb{P}(G \in A)| \leq \sqrt{\mathrm{Ent}(S_n||G)}.$$

- In 84' Barron showed that if $\mathrm{Ent}(X||G) < \infty$ then

$$\lim_{n\to\infty} \mathrm{Ent}(S_n||G) = 0.$$

- In 2011, Bobkov, Chistyakov and Götze showed that if, in addition, X has a finite fourth moment then

$$\mathrm{Ent}(S_n||G) \leq \frac{C}{n}.$$

- The above constant may depend on X as well as the dimension.

Relative entropy

For relative entropy, if $A \subset \mathbb{R}^d$ is any measurable set, then by Pinsker's inequality,

$$|\mathbb{P}(S_n \in A) - \mathbb{P}(G \in A)| \leq \sqrt{\mathrm{Ent}(S_n || G)}.$$

- In 84' Barron showed that if $\mathrm{Ent}(X||G) < \infty$ then
$$\lim_{n\to\infty} \mathrm{Ent}(S_n||G) = 0.$$
- In 2011, Bobkov, Chistyakov and Götze showed that if, in addition, X has a finite fourth moment then
$$\mathrm{Ent}(S_n||G) \leq \frac{C}{n}.$$
- The above constant may depend on X as well as the dimension.

Relative entropy

For relative entropy, if $A \subset \mathbb{R}^d$ is any measurable set, then by Pinsker's inequality,

$$|\mathbb{P}(S_n \in A) - \mathbb{P}(G \in A)| \leq \sqrt{\mathrm{Ent}(S_n || G)}.$$

- In 84' Barron showed that if $\mathrm{Ent}(X||G) < \infty$ then
$$\lim_{n \to \infty} \mathrm{Ent}(S_n||G) = 0.$$
- In 2011, Bobkov, Chistyakov and Götze showed that if, in addition, X has a finite fourth moment then
$$\mathrm{Ent}(S_n||G) \leq \frac{C}{n}.$$
- The above constant may depend on X as well as the dimension.

Wasserstein distance

The approximation error on a convex set $K \subset \mathbb{R}^d$, can be related to the Wasserstein distance using the following inequality by Zhai

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq d^{\frac{1}{6}} \mathcal{W}_2(S_n, G)^{\frac{2}{3}}.$$

Proof.

Take the optimal coupling, so $\mathbb{E}\left[||S_n - G||^2\right] = \mathcal{W}_2(S_n, G)^2$.

$$\begin{aligned}
\mathbb{P}(S_n \in K) &\leq \mathbb{P}(||S_n - G|| \leq \varepsilon, S_n \in K) + \mathbb{P}(||S_n - G|| > \varepsilon) \\
&\leq \mathbb{P}(G \in K^\varepsilon) + \varepsilon^{-2} \mathcal{W}_2(S_n, G)^2 \\
&\leq \mathbb{P}(G \in K) + \varepsilon d^{\frac{1}{4}} + \varepsilon^{-2} \mathcal{W}_2(S_n, G)^2.
\end{aligned}$$

Now, optimize over ε. □

Wasserstein distance

The approximation error on a convex set $K \subset \mathbb{R}^d$, can be related to the Wasserstein distance using the following inequality by Zhai

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq d^{\frac{1}{6}} \mathcal{W}_2(S_n, G)^{\frac{2}{3}}.$$

Proof.

Take the optimal coupling, so $\mathbb{E}\left[||S_n - G||^2\right] = \mathcal{W}_2\left(S_n, G\right)^2$.

$$\begin{aligned}\mathbb{P}(S_n \in K) &\leq \mathbb{P}(||S_n - G|| \leq \varepsilon, S_n \in K) + \mathbb{P}(||S_n - G|| > \varepsilon) \\ &\leq \mathbb{P}(G \in K^{\varepsilon}) + \varepsilon^{-2}\mathcal{W}_2(S_n, G)^2 \\ &\leq \mathbb{P}(G \in K) + \varepsilon d^{\frac{1}{4}} + \varepsilon^{-2}\mathcal{W}_2(S_n, G)^2.\end{aligned}$$

Now, optimize over ε. □

Wasserstein distance

Theorem (Zhai)

If $||X|| \leq \beta$ *almost surely then*

$$\mathcal{W}_2(S_n, G) \leq \frac{\sqrt{d}\beta \log(n)}{\sqrt{n}}.$$

- Plugging this into the previous inequality shows

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{1}{2}}\beta^{\frac{2}{3}}}{n^{\frac{1}{3}}}.$$

- Substituting $\mathbb{E}\left[||X||^3\right]$ for β^3 in Bentkus' bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{1}{4}}\beta^3}{n^{\frac{1}{2}}}.$$

- the bounds are not comparable.

Wasserstein distance

Theorem (Zhai)

If $||X|| \leq \beta$ *almost surely then*

$$\mathcal{W}_2\left(S_n, G\right) \leq \frac{\sqrt{d}\beta\log(n)}{\sqrt{n}}.$$

- Plugging this into the previous inequality shows

$$|\mathbb{P}(S_n \in K) - \mathbb{P}\left(G \in K\right)| \leq \frac{d^{\frac{1}{2}}\beta^{\frac{2}{3}}}{n^{\frac{1}{3}}}.$$

- Substituting $\mathbb{E}\left[||X||^3\right]$ for β^3 in Bentkus' bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}\left(G \in K\right)| \leq \frac{d^{\frac{1}{4}}\beta^3}{n^{\frac{1}{2}}}.$$

- the bounds are not comparable.

Wasserstein distance

Theorem (Zhai)

If $||X|| \leq \beta$ *almost surely then*

$$\mathcal{W}_2\left(S_n, G\right) \leq \frac{\sqrt{d}\beta\log(n)}{\sqrt{n}}.$$

- Plugging this into the previous inequality shows

$$|\mathbb{P}(S_n \in K) - \mathbb{P}\left(G \in K\right)| \leq \frac{d^{\frac{1}{2}}\beta^{\frac{2}{3}}}{n^{\frac{1}{3}}}.$$

- Substituting $\mathbb{E}\left[||X||^3\right]$ for β^3 in Bentkus' bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}\left(G \in K\right)| \leq \frac{d^{\frac{1}{4}}\beta^3}{n^{\frac{1}{2}}}.$$

- the bounds are not comparable.

Wasserstein distance

Theorem (Zhai)

If $||X|| \leq \beta$ *almost surely then*

$$\mathcal{W}_2\left(S_n, G\right) \leq \frac{\sqrt{d}\beta\log(n)}{\sqrt{n}}.$$

- Plugging this into the previous inequality shows

$$|\mathbb{P}(S_n \in K) - \mathbb{P}\left(G \in K\right)| \leq \frac{d^{\frac{1}{2}}\beta^{\frac{2}{3}}}{n^{\frac{1}{3}}}.$$

- Substituting $\mathbb{E}\left[||X||^3\right]$ for β^3 in Bentkus' bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}\left(G \in K\right)| \leq \frac{d^{\frac{1}{4}}\beta^3}{n^{\frac{1}{2}}}.$$

- the bounds are not comparable.

Wasserstein distance

Consider X, distributed uniformly on $\pm\sqrt{d}e_i$. In this case, $\beta = \sqrt{d}$ and Zhai's bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{5}{6}}}{n^{\frac{1}{3}}}.$$

So, we can expect the CLT to hold whenever $d^{\frac{5}{2}} << n$.
On the other hand, Bentkus' bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{7}{4}}}{n^{\frac{1}{2}}}.$$

In this case, we would require $d^{\frac{7}{2}} << n$ for convergence.

Wasserstein distance

Consider X, distributed uniformly on $\pm\sqrt{d}e_i$. In this case, $\beta = \sqrt{d}$ and Zhai's bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{5}{6}}}{n^{\frac{1}{3}}}.$$

So, we can expect the CLT to hold whenever $d^{\frac{5}{2}} << n$.

On the other hand, Bentkus' bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{7}{4}}}{n^{\frac{1}{2}}}.$$

In this case, we would require $d^{\frac{7}{2}} << n$ for convergence.

Wasserstein distance

Consider X, distributed uniformly on $\pm\sqrt{d}e_i$. In this case, $\beta = \sqrt{d}$ and Zhai's bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{5}{6}}}{n^{\frac{1}{3}}}.$$

So, we can expect the CLT to hold whenever $d^{\frac{5}{2}} << n$.
On the other hand, Bentkus' bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{7}{4}}}{n^{\frac{1}{2}}}.$$

In this case, we would require $d^{\frac{7}{2}} << n$ for convergence.

Wasserstein distance

Consider X, distributed uniformly on $\pm\sqrt{d}e_i$. In this case, $\beta = \sqrt{d}$ and Zhai's bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{5}{6}}}{n^{\frac{1}{3}}}.$$

So, we can expect the CLT to hold whenever $d^{\frac{5}{2}} << n$.
On the other hand, Bentkus' bound gives

$$|\mathbb{P}(S_n \in K) - \mathbb{P}(G \in K)| \leq \frac{d^{\frac{7}{4}}}{n^{\frac{1}{2}}}.$$

In this case, we would require $d^{\frac{7}{2}} << n$ for convergence.

A new idea

Definition

A Skorokhod embedding of X is a Brownian motion B_t along with a stopping time τ such that B_τ has the same law as X.

Theorem (Skorokhod's embedding theorem)

If X is 1-dimensional and $\mathbb{E}[X] = 0$, there exists a Skorokhod embedding of X with $\mathbb{E}[\tau] = \mathbb{E}\left[X^2\right]$. Moreover, if X is bounded almost surely then τ has sub-exponential tails.

A new idea

Definition

A Skorokhod embedding of X is a Brownian motion B_t along with a stopping time τ such that B_τ has the same law as X.

Theorem (Skorokhod's embedding theorem)

If X is 1-dimensional and $\mathbb{E}[X] = 0$, there exists a Skorokhod embedding of X with $\mathbb{E}[\tau] = \mathbb{E}\left[X^2\right]$. Moreover, if X is bounded almost surely then τ has sub-exponential tails.

From Skorokhod embedding to CLT

Consider (B_t^i, τ_i), i.i.d. Skorokhod embeddings of X. We then have

$$S_n = \int_0^\infty \sum_{i=1}^n \frac{\mathbb{1}_{[0,\tau_i]}(t)}{\sqrt{n}} dB_t^i = \int_0^\infty \tilde{\mathbb{1}}(t) d\tilde{B}_t,$$

where $\tilde{\mathbb{1}} = \sqrt{\frac{\sum_{i=1}^n \mathbb{1}_{[0,\tau_i]}}{n}}$ and $\tilde{B}_t$ is a Brownian motion.

From Skorokhod embedding to CLT

Denote $G_n := \int_0^\infty \mathbb{E}\left[\tilde{\mathbb{1}}(t)\right] d\tilde{B}_t$, a rescaled Brownian motion. So that,

$$S_n = \int_0^\infty \tilde{\mathbb{1}}(t) d\tilde{B}_t = G_n + \int_0^\infty \tilde{\mathbb{1}}(t) - \mathbb{E}\left[\tilde{\mathbb{1}}(t)\right] d\tilde{B}_t.$$

This induces a natural coupling between S_n and G_n, which shows:

$$\begin{aligned}\mathcal{W}_2^2(S_n, G_n) &\leq \mathbb{E}\left[\left(\int_0^\infty \left(\tilde{\mathbb{1}}(t) - \mathbb{E}\left[\tilde{\mathbb{1}}(t)\right]\right) d\tilde{B}_t\right)^2\right] \\ &= \int_0^\infty \mathbb{E}\left[\left(\tilde{\mathbb{1}}(t) - \mathbb{E}\left[\tilde{\mathbb{1}}(t)\right]\right)^2 dt\right] = \int_0^\infty \mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) dt.\end{aligned}$$

From Skorokhod embedding to CLT

Denote $G_n := \int_0^\infty \mathbb{E}\left[\tilde{\mathbb{1}}(t)\right] d\tilde{B}_t$, a rescaled Brownian motion. So that,

$$S_n = \int_0^\infty \tilde{\mathbb{1}}(t) d\tilde{B}_t = G_n + \int_0^\infty \tilde{\mathbb{1}}(t) - \mathbb{E}\left[\tilde{\mathbb{1}}(t)\right] d\tilde{B}_t.$$

This induces a natural coupling between S_n and G_n, which shows:

$$\mathcal{W}_2^2\left(S_n, G_n\right) \leq \mathbb{E}\left[\left(\int_0^\infty \left(\tilde{\mathbb{1}}(t) - \mathbb{E}\left[\tilde{\mathbb{1}}(t)\right]\right) d\tilde{B}_t\right)^2\right]$$

$$= \int_0^\infty \mathbb{E}\left[\left(\tilde{\mathbb{1}}(t) - \mathbb{E}\left[\tilde{\mathbb{1}}(t)\right]\right)^2 dt\right] = \int_0^\infty \mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) dt.$$

Analysis of the coupling

- Recall $\tilde{\mathbb{1}}(t) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[0,\tau_i]}}$, so $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \to 0$.
- Moreover, one can show for any positive random variable Y

$$\mathrm{Var}\left(\sqrt{Y}\right) \leq \frac{\mathrm{Var}(Y)}{\mathbb{E}[Y]}.$$

  In our case, $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \leq \frac{1}{n}$.
- Also, $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \leq \mathbb{E}\left[\mathbb{1}_{[0,\tau]}(t)\right] = \mathbb{P}\left(t < \tau\right)$.
- So,

$$\mathcal{W}_2^2\left(S_n, G_n\right) \leq \int_0^\infty \min\left(\frac{1}{n}, \mathbb{P}\left(t < \tau\right)\right) dt.$$

Analysis of the coupling

- Recall $\tilde{\mathbb{1}}(t) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[0,\tau_i]}}$, so $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \to 0$.
- Moreover, one can show for any positive random variable Y

$$\mathrm{Var}\left(\sqrt{Y}\right) \leq \frac{\mathrm{Var}(Y)}{\mathbb{E}[Y]}.$$

  In our case, $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \leq \frac{1}{n}$.
- Also, $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \leq \mathbb{E}\left[\mathbb{1}_{[0,\tau]}(t)\right] = \mathbb{P}\left(t < \tau\right)$.
- So,

$$\mathcal{W}_2^2\left(S_n, G_n\right) \leq \int_0^\infty \min\left(\frac{1}{n}, \mathbb{P}\left(t < \tau\right)\right) dt.$$

Analysis of the coupling

- Recall $\tilde{\mathbb{1}}(t) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[0,\tau_i]}}$, so $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \to 0$.
- Moreover, one can show for any positive random variable Y

$$\mathrm{Var}\left(\sqrt{Y}\right) \leq \frac{\mathrm{Var}(Y)}{\mathbb{E}[Y]}.$$

  In our case, $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \leq \frac{1}{n}$.
- Also, $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \leq \mathbb{E}\left[\mathbb{1}_{[0,\tau]}(t)\right] = \mathbb{P}\left(t < \tau\right)$.
- So,

$$\mathcal{W}_2^2\left(S_n, G_n\right) \leq \int_0^\infty \min\left(\frac{1}{n}, \mathbb{P}\left(t < \tau\right)\right) dt.$$

Analysis of the coupling

- Recall $\tilde{\mathbb{1}}(t) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[0,\tau_i]}}$, so $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \to 0$.
- Moreover, one can show for any positive random variable Y

$$\mathrm{Var}\left(\sqrt{Y}\right) \leq \frac{\mathrm{Var}(Y)}{\mathbb{E}[Y]}.$$

  In our case, $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \leq \frac{1}{n}$.
- Also, $\mathrm{Var}\left(\tilde{\mathbb{1}}(t)\right) \leq \mathbb{E}\left[\mathbb{1}_{[0,\tau]}(t)\right] = \mathbb{P}\left(t < \tau\right)$.
- So,

$$\mathcal{W}_2^2\left(S_n, G_n\right) \leq \int_0^\infty \min\left(\frac{1}{n}, \mathbb{P}\left(t < \tau\right)\right) dt.$$

Extending to higher dimensions

- The Skorokhod embedding is a 1-dimensional construction.
- For random vectors we wouldn't expect such an embedding to exist.
- We are thus led to a more general notion:

Definition (Martingale embedding)

The triplet (M_t, Γ_t, τ) is a martingale embedding of X, if M_t is a martingale which satisfies $dM_t = \Gamma_t dB_t$ and M_τ has the same law as X.

Extending to higher dimensions

- The Skorokhod embedding is a 1-dimensional construction.
- For random vectors we wouldn't expect such an embedding to exist.
- We are thus led to a more general notion:

Definition (Martingale embedding)

The triplet (M_t, Γ_t, τ) is a martingale embedding of X, if M_t is a martingale which satisfies $dM_t = \Gamma_t dB_t$ and M_τ has the same law as X.

Extending to higher dimensions

- The Skorokhod embedding is a 1-dimensional construction.
- For random vectors we wouldn't expect such an embedding to exist.
- We are thus led to a more general notion:

Definition (Martingale embedding)

The triplet (M_t, Γ_t, τ) is a martingale embedding of X, if M_t is a martingale which satisfies $dM_t = \Gamma_t dB_t$ and M_τ has the same law as X.

Extending to higher dimensions

For martingale embeddings the same ideas used for the Skorokhod embedding yields

Theorem

If (M_t, Γ_t, τ) is a martingale embedding of X, and Γ_t is positive definite, then

$$\mathcal{W}_2^2(S_n, G) \leq \int_0^\infty \min\left(\frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\Gamma_t^4\right]\mathbb{E}\left[\Gamma_t^2\right]^{-1}\right), \mathrm{Tr}\left(\mathbb{E}\left[\Gamma_t^2\right]\right)\right) dt.$$

Note that if Γ_t is a projection matrix the bound simplifies to

$$\mathcal{W}_2^2(S_n, G) \leq d\int_0^\infty \min\left(\frac{1}{n}, \mathbb{P}(t \leq \tau)\right) dt.$$

Extending to higher dimensions

By repeatedly projecting a Brownian motion into lower dimensional spaces we are able to construct a martingale embedding with similar properties to the 1-dimensional Skorokhod embedding. In particular

- Γ_t is a projection matrix.
- $\mathbb{E}[\tau] \leq \mathbb{E}\left[||X||^2\right]$.
- If $||X|| \leq \beta$ almost surely, τ has sub exponential tails.

This leads to the following result

Theorem

If $||X|| \leq \beta$ almost surely

$$\mathcal{W}_2\left(S_n, G\right) \leq \frac{\sqrt{d\log(n)}\beta}{\sqrt{n}}.$$

Extending to higher dimensions

By repeatedly projecting a Brownian motion into lower dimensional spaces we are able to construct a martingale embedding with similar properties to the 1-dimensional Skorokhod embedding. In particular

- Γ_t is a projection matrix.
- $\mathbb{E}[\tau] \leq \mathbb{E}\left[||X||^2\right]$.
- If $||X|| \leq \beta$ almost surely, τ has sub exponential tails.

This leads to the following result

Theorem

If $||X|| \leq \beta$ almost surely

$$\mathcal{W}_2\left(S_n, G\right) \leq \frac{\sqrt{d \log(n)}\beta}{\sqrt{n}}.$$

Extending to higher dimensions - log concave measures

If X is log concave (it has a density f, such that $-\nabla^2 log(f) \geq 0$), then we can improve beyond anything directly implied by the previous theorem.

Denote $\kappa_d := \sup_Y \mathrm{Var}(||Y||)$, where the supremum is taken over all isotropic log concave random vectors in $\mathbb{R}^d$.

Theorem

If X is isotropic and log concave then, up to logarithmic factors

$$\mathcal{W}_2(S_n, G) \leq \sqrt{\frac{d}{n}}\kappa_d.$$

Moreover if X is $\frac{1}{\alpha}$-strongly log concave ($-\nabla^2 \log(f) \geq \alpha \mathrm{I}_d$) then

$$\mathcal{W}_2(S_n, G) \leq \sqrt{\frac{d}{n\alpha}}.$$

Extending to higher dimensions - log concave measures

If X is log concave (it has a density f, such that $-\nabla^2 log(f) \geq 0$), then we can improve beyond anything directly implied by the previous theorem.

Denote $\kappa_d := \sup_Y \mathrm{Var}\left(||Y||\right)$, where the supremum is taken over all isotropic log concave random vectors in $\mathbb{R}^d$.

Theorem

If X is isotropic and log concave then, up to logarithmic factors

$$\mathcal{W}_2\left(S_n, G\right) \leq \sqrt{\frac{d}{n}} \kappa_d.$$

Moreover if X is $\frac{1}{\alpha}$-strongly log concave ($-\nabla^2 \log(f) \geq \alpha \mathrm{I}_d$) then

$$\mathcal{W}_2\left(S_n, G\right) \leq \sqrt{\frac{d}{n\alpha}}.$$

Extending to higher dimensions - log concave measures

If X is log concave (it has a density f, such that $-\nabla^2 log(f) \geq 0$), then we can improve beyond anything directly implied by the previous theorem.

Denote $\kappa_d := \sup_Y \mathrm{Var}\left(||Y||\right)$, where the supremum is taken over all isotropic log concave random vectors in $\mathbb{R}^d$.

Theorem

If X is isotropic and log concave then, up to logarithmic factors

$$\mathcal{W}_2\left(S_n, G\right) \leq \sqrt{\frac{d}{n}}\kappa_d.$$

Moreover if X is $\frac{1}{\alpha}$-strongly log concave ($-\nabla^2 \log(f) \geq \alpha \mathrm{I}_d$) then

$$\mathcal{W}_2\left(S_n, G\right) \leq \sqrt{\frac{d}{n\alpha}}.$$

Martingale embeddings in the entropic CLT

We may also use martingale embeddings to obtain quantitative bounds in the entropic CLT:

Theorem

If $(M_t, \Gamma_t, 1)$ *is a martingale embedding of* X*, then*

$$\mathrm{Ent}(S_n || G) \leq \frac{1}{n} \int_0^1 \frac{\mathbb{E}\mathrm{Tr}\left(\left(\Gamma_t^2 - \mathbb{E}[\Gamma_t^2]\right)^2\right)}{(1-t)\sigma_t^4} dt,$$

where σ_t *is such that* $\mathbb{E}\left[\Gamma_t\right] \geq \sigma_t \mathrm{I}_d$.

Sketch of proof

- Denote $\tilde{\Gamma}_t = \sqrt{\frac{\sum \Gamma_t^2}{n}}$. As before

$$S_n = \int_0^1 \tilde{\Gamma}_t d\tilde{B}_t = \int_0^1 \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t + \int_0^1 \tilde{\Gamma}_t - \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t.$$

- Note that $G \overset{\text{law}}{=} \int_0^1 \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t$.
- Our goal is to reconstruct the discrepancy as an adapted drift to which Girsanov's theorem may apply.

Sketch of proof

- Denote $\tilde{\Gamma}_t = \sqrt{\frac{\sum \Gamma_t^2}{n}}$. As before

$$S_n = \int_0^1 \tilde{\Gamma}_t d\tilde{B}_t = \int_0^1 \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t + \int_0^1 \tilde{\Gamma}_t - \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t.$$

- Note that $G \overset{\text{law}}{=} \int_0^1 \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t$.
- Our goal is to reconstruct the discrepancy as an adapted drift to which Girsanov's theorem may apply.

Sketch of proof

- Denote $\tilde{\Gamma}_t = \sqrt{\frac{\sum \Gamma_t^2}{n}}$. As before

$$S_n = \int_0^1 \tilde{\Gamma}_t d\tilde{B}_t = \int_0^1 \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t + \int_0^1 \tilde{\Gamma}_t - \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t.$$

- Note that $G \overset{\text{law}}{=} \int_0^1 \sqrt{\mathbb{E}\left[\tilde{\Gamma}_t^2\right]} d\tilde{B}_t$.
- Our goal is to reconstruct the discrepancy as an adapted drift to which Girsanov's theorem may apply.

Sketch of proof

- Let $u_t := \int_0^t \frac{\tilde{\Gamma}_s - \sqrt{\mathbb{E}[\tilde{\Gamma}_s^2]}}{1-s} d\tilde{B}_s$ so that

$$\int_0^1 u_t dt = \int_0^1 \int_0^t \frac{\tilde{\Gamma}_s - \sqrt{\mathbb{E}\left[\tilde{\Gamma}_s^2\right]}}{1-s} d\tilde{B}_s dt = \int_0^1 \int_s^1 \frac{\tilde{\Gamma}_s - \sqrt{\mathbb{E}\left[\tilde{\Gamma}_s^2\right]}}{1-s} dt d\tilde{B}_s$$
$$= \int_0^1 \tilde{\Gamma}_s - \sqrt{\mathbb{E}\left[\tilde{\Gamma}_s^2\right]} d\tilde{B}_s.$$

- So $S_n = G + \int_0^1 u_t dt$. By Girsanov's theorem we get that, f, the density of S_n with respect to G satisfies

$$\mathbb{E}[\log(f)] \leq \frac{1}{2} \int_0^1 \mathbb{E}\left[\left|\left|\mathbb{E}\left[\tilde{\Gamma}_t^2\right]^{-\frac{1}{2}} u_t\right|\right|^2\right] dt.$$

Towards an embedding - the Föllmer drift

To find a good embedding we consider a solution to the following variational problem:

$$v_t := \arg\min_{u_t} \frac{1}{2} \int_0^1 \mathbb{E}\left[||u_t||^2\right] dt,$$

where u_t ranges over all adapted drifts for which $B_1 + \int_0^1 u_t dt$ has the same law as X.

Towards an embedding - the Föllmer drift

The process v_t goes back at least to the works of Föllmer (85'). In a later work by Lehec (13') it is shown that if X has finite entropy relative to the Gaussian, then v_t is well defined and

$$\mathrm{Ent}\,(X||G) = \frac{1}{2}\int_0^1 \mathbb{E}[||v_t||^2]dt.$$

In this case, v_t is a martingale and the process

$$Y_t = B_t + \int_0^t u_s ds,$$

is a Brownian bridge between 0 and X.

Towards an embedding - the Föllmer drift

The process v_t goes back at least to the works of Föllmer (85'). In a later work by Lehec (13') it is shown that if X has finite entropy relative to the Gaussian, then v_t is well defined and

$$\mathrm{Ent}\,(X||G) = \frac{1}{2}\int_0^1 \mathbb{E}[||v_t||^2]dt.$$

In this case, v_t is a martingale and the process

$$Y_t = B_t + \int_0^t u_s ds,$$

is a Brownian bridge between 0 and X.

Constructing an embedding

We use Y_t to construct a martingale embedding.

$$X_t := \mathbb{E}\left[Y_1 | \mathcal{F}_t\right].$$

The process X_t satisfies

$$X_t = \int_0^t \frac{\mathrm{Cov}\left(Y_1 | \mathcal{F}_s\right)}{1-s} dB_s = \int_0^t \Gamma_s dB_s.$$

This implies

$$v_t = \int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s} dB_s.$$

Constructing an embedding

We use Y_t to construct a martingale embedding.

$$X_t := \mathbb{E}\left[Y_1 | \mathcal{F}_t\right].$$

The process X_t satisfies

$$X_t = \int_0^t \frac{\mathrm{Cov}\left(Y_1 | \mathcal{F}_s\right)}{1-s} dB_s = \int_0^t \Gamma_s dB_s.$$

This implies

$$v_t = \int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s} dB_s.$$

Constructing an embedding

We use Y_t to construct a martingale embedding.

$$X_t := \mathbb{E}\left[Y_1 | \mathcal{F}_t\right].$$

The process X_t satisfies

$$X_t = \int_0^t \frac{\mathrm{Cov}\left(Y_1 | \mathcal{F}_s\right)}{1-s} dB_s = \int_0^t \Gamma_s dB_s.$$

This implies

$$v_t = \int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s} dB_s.$$

Entropic CLT for log concave vectors

$$\mathrm{Ent}(X||G) = \int_0^1 \mathbb{E}\Big|\Big| \int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s} dB_s \Big|\Big|^2 dt$$

$$= \int_0^1 \int_0^t \frac{\mathbb{E}\mathrm{Tr}\left((\Gamma_s - \mathrm{I}_d)^2\right)}{(1-s)^2} ds dt = \int_0^1 \frac{\mathbb{E}\mathrm{Tr}\left((\Gamma_t - \mathrm{I}_d)^2\right)}{1-t} dt.$$

We use this observation to prove:

Entropic CLT for log concave vectors

$$\mathrm{Ent}(X||G) = \int_0^1 \mathbb{E}\Big|\Big| \int_0^t \frac{\Gamma_s - \mathrm{I}_d}{1-s} dB_s \Big|\Big|^2 dt$$

$$= \int_0^1 \int_0^t \frac{\mathbb{E}\mathrm{Tr}\left((\Gamma_s - \mathrm{I}_d)^2\right)}{(1-s)^2} ds dt = \int_0^1 \frac{\mathbb{E}\mathrm{Tr}\left((\Gamma_t - \mathrm{I}_d)^2\right)}{1-t} dt.$$

We use this observation to prove:

Entropic CLT for log concave vectors

Theorem

1. *If X is log concave and isotropic then*

$$\mathrm{Ent}(S_n||G) \leq \frac{\mathrm{poly}(d)}{n}\mathrm{Ent}\left(X||G\right).$$

2. *If X is 1-strongly log concave (and not isotropic) then*

$$\mathrm{Ent}(S_n||G) \leq \frac{d}{n\sigma^4}\mathrm{Ent}\left(X||G\right),$$

where σ is the minimal eigenvalue of $\mathrm{Cov}\left(X\right)$.

Embeddings of log concave vectors

In the case where X is log concave, it turns out that Γ_t cannot be large.

- If X has density f, then $Y_1|\mathcal{F}_t$ has density proportional to

$$f(x)\exp\left(-\frac{t}{2(1-t)}||x||^2 + \frac{\langle X_t, x\rangle}{1-t}\right).$$

- In particular, if X is log concave then $X_1|\mathcal{F}_t$ is $\frac{1-t}{t}$-strongly log concave.
- Consequently, $\Gamma_t \leq \frac{1}{t}\mathrm{I}_d$.
- The same logic shows that $\Gamma_t \leq \mathrm{I}_d$ whenever X is 1-strongly log concave.

Embeddings of log concave vectors

Lemma

If X is 1-strongly log concave and $\mathrm{Cov}(X) \geq \sigma \mathrm{I}_d$ *then*

$$\mathbb{E}[\Gamma_t] \geq \sigma \mathrm{I}_d.$$

Proof.

First note

$$\mathrm{Cov}(Y_1|\mathcal{F}_t) = \mathbb{E}\left[Y_1^{\otimes 2}|\mathcal{F}_t\right] - \mathbb{E}[Y_1|\mathcal{F}_t]^{\otimes 2}.$$

Hence, by Itô's formula

$$\frac{d}{dt}\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] = -\frac{d}{dt}\mathbb{E}\left[\mathbb{E}[Y_1|\mathcal{F}_t]^{\otimes 2}\right] = -\mathbb{E}\left[\Gamma_t^2\right].$$

Embeddings of log concave vectors

Lemma

If X is 1-strongly log concave and $\mathrm{Cov}(X) \geq \sigma \mathrm{I}_d$ then

$$\mathbb{E}[\Gamma_t] \geq \sigma \mathrm{I}_d.$$

Proof.

First note

$$\mathrm{Cov}(Y_1|\mathcal{F}_t) = \mathbb{E}\left[Y_1^{\otimes 2}|\mathcal{F}_t\right] - \mathbb{E}[Y_1|\mathcal{F}_t]^{\otimes 2}.$$

Hence, by Itô's formula

$$\frac{d}{dt}\mathbb{E}[\mathrm{Cov}(Y_1|\mathcal{F}_t)] = -\frac{d}{dt}\mathbb{E}\left[\mathbb{E}[Y_1|\mathcal{F}_t]^{\otimes 2}\right] = -\mathbb{E}\left[\Gamma_t^2\right].$$

Proof (cont'd).

So,

$$\frac{d}{dt}\mathbb{E}\left[\Gamma_t\right] = \frac{d}{dt}\mathbb{E}\left[\frac{\mathrm{Cov}\left(Y_1|\mathcal{F}_t\right)}{1-t}\right] = \frac{\mathbb{E}\left[\mathrm{Cov}\left(Y_1|\mathcal{F}_t\right)\right] - (1-t)\mathbb{E}\left[\Gamma_t^2\right]}{(1-t)^2} = \frac{\mathbb{E}\left[\Gamma_t\right] - \mathbb{E}\left[\Gamma_t^2\right]}{1-t}.$$

Since $\Gamma_t \leq \mathrm{I}_d$ almost surely

$$\frac{\mathbb{E}\left[\Gamma_t\right] - \mathbb{E}\left[\Gamma_t^2\right]}{1-t} \geq 0.$$

□

Thank you!