Asymptotic Efficiency in High-Dimensional Covariance Estimation

Vladimir Koltchinskii

School of Mathematics
Georgia Institute of Technology

Lübbenau, March 2018

Problems

Let $X, X_1, \dots, X_n$ be i.i.d. Gaussian vectors with values in $\mathbb{R}^d$, with $\mathbb{E}X = 0$ and with covariance operator $\Sigma = \mathbb{E}(X \otimes X) \in \mathcal{C}_+^d$.

- Given a smooth function $f : \mathbb{R} \mapsto \mathbb{R}$ and a linear operator $B : \mathbb{R}^d \mapsto \mathbb{R}^d$ with $\|B\|_1 \leq 1$, estimate $\langle f(\Sigma), B \rangle$ based on $X_1, \dots, X_n$.
- More precisely, we are interested in finding asymptotically efficient estimators of $\langle f(\Sigma), B \rangle$ with $\sqrt{n}$-convergence rate in the case when $d = d_n \to \infty$.
- Suppose $d_n \leq n^\alpha$ for some $\alpha > 0$. Is there $s(\alpha)$ such that for all $s > s(\alpha)$ and for all functions f of smoothness s, asymptotically efficient estimation is possible?

Some Related Results

- Efficient estimation of smooth functionals in nonparametric models: Levit (1975, 1978), Ibragimov and Khasminskii (1981);
- In particular, in Gaussian shift model: Ibragimov, Nemirovski and Khasminskii (1987), Nemirovski (1990, 2000)
- Girko (1987–): asymptotically normal estimators of a number of special functionals (such as $\log \det(\Sigma) = \mathrm{tr}(\log \Sigma)$, Stieltjes transform of spectral function of $\Sigma : \mathrm{tr}((I + t\Sigma)^{-1}))$, ... Based on martingale CLT
- Asymptotic normality of log-determinant $\log \det(\hat{\Sigma})$ has been studied by many authors (see, e.g., Cai, Liang and Zhou (2015) for a recent result)

Some Related Results

- Asymptotic normality of $\mathrm{tr}(f(\hat{\Sigma}))$ for a smooth function $f : \mathbb{R} \mapsto \mathbb{R}$: (linear spectral statistic). Common topic in random matrix theory (both for Wigner and for Wishart matrices): Bai and Silverstein (2004), Lytova and Pastur (2009), Sosoe and Wong (2015)
- Estimation of functionals of covariance matrices under sparsity: Fan, Rigollet and Wang (2015)
- Bernstein–von Mises theorems for functionals of covariance: Gao and Zhou (2016)
- Efficient estimation of linear functionals of principal components: Koltchinskii, Löffler and Nickl (2017)

Outline

- Part 1.
  Effective Rank and Sample Covariance
- Part 2.
  Taylor Expansions of Operator Functions and Normal Approximation of Plug-In Estimators of Smooth Functionals of Covariance
- Part 3.
  Wishart Operators, Bootstrap Chains, Invariant Functions and Bias Reduction
- Part 4.
  Asymptotic Efficiency

Part 1.
Effective Rank and Sample Covariance

Covariance Operator

- $(E, \|\cdot\|)$ a separable Banach space, E^* its dual space
- X a centered random variable in E, $\mathbb{E}|\langle X, u\rangle|^2 < +\infty, u \in E^*$
- The covariance operator:

$$\Sigma u := \mathbb{E}\langle X, u\rangle X, \ u \in E^*.$$

- $\Sigma : E^* \mapsto E$ a bounded symmetric nonnegatively definite operator. If $\mathbb{E}\|X\|^2 < +\infty$, then Σ is nuclear

Sample Covariance Operator

- $X_1, \dots, X_n$ i.i.d. copies of X.
- The sample (empirical) covariance operator $\hat{\Sigma} : E^* \mapsto E$,

$$\hat{\Sigma} u := n^{-1} \sum_{j=1}^{n} \langle X_j, u \rangle X_j, \ u \in E^*.$$

- Problems:
  - What is the size of $\mathbb{E}\|\hat{\Sigma} - \Sigma\|$, where $\|\cdot\|$ is the operator norm?
  - Concentration inequalities for $\|\hat{\Sigma} - \Sigma\|$ around its expectation or median.

Subgaussian Random Variables

Definition

A centered random variable X in E will be called *subgaussian* iff, for all $u \in E^*$,

$$\|\langle X, u\rangle\|_{\psi_2} \lesssim \|\langle X, u\rangle\|_{L_2(\mathbb{P})}.$$

Notations: Given a convex nondecreasing function $\psi : \mathbb{R}_+ \mapsto \mathbb{R}_+$, $\psi(0) = 0$, η a r.v. on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$,

$$\|\eta\|_\psi := \inf\left\{C > 0 : \mathbb{E}\psi\left(\frac{|\eta|}{C}\right) \le 1\right\}.$$

$$\psi_2(u) := e^{u^2} - 1, u \ge 0$$

$$\psi_1(u) := e^u - 1.$$

A Bound in the Finite Dimensional Case

$E = \mathbb{R}^d, d \geq 1$ (Euclidean space)

Theorem

Suppose that X is subgaussian. Then, there exists an absolute constant $C > 0$ such that, for all $t \geq 1$, with probability at least $1 - e^{-t}$

$$\|\hat{\Sigma} - \Sigma\| \leq C\|\Sigma\| \left(\sqrt{\frac{d}{n}} \bigvee \frac{d}{n} \bigvee \sqrt{\frac{t}{n}} \bigvee \frac{t}{n} \right).$$

It implies that

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \leq C\|\Sigma\| \left(\sqrt{\frac{d}{n}} \bigvee \frac{d}{n} \right).$$

Sketch of the proof

- $M \subset S^{d-1}$ is a $1/4$-net of the unit sphere S^{d-1}, $\mathrm{card}(M) \le 9^d$
-
$$\|\hat{\Sigma} - \Sigma\| \lesssim \max_{u,v \in M} |\langle (\hat{\Sigma} - \Sigma)u, v\rangle|$$
$$= \max_{u,v \in M} \left| n^{-1} \sum_{j=1}^{n} \langle X_j, u\rangle \langle X_j, v\rangle - \mathbb{E}\langle X, u\rangle\langle X, v\rangle \right|$$
- Use the union bound and Bernstein inequality

$$\mathbb{P}\left\{ \left| \frac{\xi_1 + \cdots + \xi_n}{n} \right| \gtrsim \|\xi\|_{\psi_1} \left(\sqrt{\frac{t + (2\log 9)d}{n}} \bigvee \frac{t + (2\log 9)d}{n} \right) \right\}$$
$$\le \exp\{-t - (2\log 9)d\}$$

for independent ψ_1 random variables $\xi_j := \langle X_j, u\rangle \langle X_j, v\rangle$.

Effective Rank

Definition

Assuming that X is a centered Gaussian random variable in E with covariance operator Σ, define

$$\mathbf{r}(\Sigma) := \frac{\mathbb{E}\|X\|^2}{\|\Sigma\|} = \frac{\mathbb{E}\sup_{\|u\|,\|v\|\leq 1}\langle X,u\rangle\langle X,v\rangle}{\sup_{\|u\|,\|v\|\leq 1}\mathbb{E}\langle X,u\rangle\langle X,v\rangle}.$$

- If E is a Hilbert space, $\mathbb{E}\|X\|^2 = \mathrm{tr}(\Sigma)$ and $\mathbf{r}(\Sigma) = \frac{\mathrm{tr}(\Sigma)}{\|\Sigma\|}$.
- $\mathbf{r}(\Sigma)$ is called "effective rank" (Vershynin (2012)).
- $\mathbf{r}(\Sigma) \leq \mathrm{rank}(\Sigma)$.
- If $\dim(\mathbb{H}) = d < +\infty$ and Σ is of *isotropic type*, that is, for some constants $0 < c_1 \leq c_2 < \infty$, $c_1 I_d \preceq \Sigma \preceq c_2 I_d$, then $\mathbf{r}(\Sigma) \asymp d$.

- Vershynin (2012)

$$\mathbb{E}\|\hat{\Sigma}-\Sigma\| \lesssim$$

$$\max\left\{\|\Sigma\|^{1/2}\mathbb{E}^{1/2}\max_{1\le j\le n}\|X_j\|^2\sqrt{\frac{\log d}{n}}, \mathbb{E}\max_{1\le j\le n}\|X_j\|^2\frac{\log d}{n}\right\}.$$

The proof is based on the approach by Rudelson (1999) and relies on noncommutative Khintchine inequality due to Lust-Picard and Pisier (1991).

- Note that, in the subgaussian case, $\Big\|\|X\|^2\Big\|_{\psi_1} \lesssim \mathrm{tr}(\Sigma)$, which implies that

$$\mathbb{E}\max_{1\le j\le n}\|X_j\|^2 \lesssim \mathrm{tr}(\Sigma)\log n = \|\Sigma\|\mathbf{r}(\Sigma)\log n.$$

This implies

$$\mathbb{E}\|\hat{\Sigma}-\Sigma\| \lesssim \|\Sigma\|\max\left\{\sqrt{\frac{\mathbf{r}(\Sigma)\log d\log n}{n}}, \frac{\mathbf{r}(\Sigma)\log d\log n}{n}\right\}.$$

Bounds in Terms of Effective Rank

If X is subgaussian, then with some constant $C > 0$ and with probability at least $1 - e^{-t}$

$$\|\hat{\Sigma} - \Sigma\| \leq C\|\Sigma\| \max\left\{\sqrt{\frac{\mathbf{r}(\Sigma)\log d + t}{n}}, \frac{\mathbf{r}(\Sigma)\log d + t)\log n}{n}\right\}.$$

The proof is based on a version of noncommutative Bernstein type inequality (Ahlswede and Winter (2002), Tropp (2012)).

Bounds for subgaussian r.v. in a separable Banach space

- E a separable Banach space
- Recall that
$$\mathbf{r}(\Sigma) = \frac{\mathbb{E}\|Y\|^2}{\|\Sigma\|}, \quad Y \sim N(0;\Sigma)$$

Definition

A weakly square integrable centered random variable X in E with covariance operator Σ is called pregaussian iff there exists a centered Gaussian random variable Y in E with the same covariance operator Σ.

Bounds for subgaussian r.v. in a separable Banach space (Koltchinskii and Lounici (2014))

Theorem

Let $X, X_1, \dots, X_n$ be i.i.d. weakly square integrable centered random vectors in E with covariance operator Σ. If X is subgaussian and pregaussian, then

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \lesssim \|\Sigma\| \max\left\{\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}, \frac{\mathbf{r}(\Sigma)}{n}\right\}$$

Moreover, if X is Gaussian, then

$$\|\Sigma\| \max\left\{\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}, \frac{\mathbf{r}(\Sigma)}{n}\right\} \lesssim \mathbb{E}\|\hat{\Sigma} - \Sigma\| \lesssim \|\Sigma\| \max\left\{\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}, \frac{\mathbf{r}(\Sigma)}{n}\right\}.$$

Proof of the Upper Bound: Gaussian Case

Lemma (Decoupling)

Let $X_1, \dots, X_n, X_1', \dots, X_n'$ *be i.i.d.* $N(0;\Sigma)$. *Then*

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \leq 2\mathbb{E} \sup_{\|u\|,\|v\|\leq 1} \left| n^{-1} \sum_{j=1}^{n} \langle X_j, u\rangle \langle X_j', v\rangle \right|.$$

Proof of the Upper Bound: Gaussian Case

Proof.

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| = \mathbb{E} \sup_{\|u\|,\|v\|\leq 1} \left| n^{-1} \sum_{j=1}^{n} \langle X_j, u\rangle \langle X_j, v\rangle - \mathbb{E}\langle X, u\rangle \langle X, v\rangle \right|$$

$$= \mathbb{E} \sup_{\|u\|,\|v\|\leq 1} \left| \mathbb{E}' n^{-1} \sum_{j=1}^{n} \Bigl(\langle X_j, u\rangle \langle X_j, v\rangle + \langle X_j', u\rangle \langle X_j, v\rangle \right.$$

$$\left. - \langle X_j, u\rangle \langle X_j', v\rangle - \langle X_j', u\rangle \langle X_j', v\rangle \Bigr) \right|$$

$$\leq 2\mathbb{E} \sup_{\|u\|,\|v\|\leq 1} \left| n^{-1} \sum_{j=1}^{n} \left\langle \frac{X_j + X_j'}{\sqrt{2}}, u \right\rangle \left\langle \frac{X_j - X_j'}{\sqrt{2}}, v \right\rangle \right|$$

$$= 2\mathbb{E} \sup_{\|u\|,\|v\|\leq 1} \left| n^{-1} \sum_{j=1}^{n} \langle X_j, u\rangle \langle X_j', v\rangle \right|.$$

Proof of the Upper Bound: Gaussian Case

- $$Y(u,v) := n^{-1/2}\sum_{j=1}^{n}\langle X_j, u\rangle\langle X_j', v\rangle$$

- $$Z(u,v) := \sqrt{2}\|\hat{\Sigma}'\|^{1/2}\langle X, u\rangle + \sqrt{2}\|\Sigma\|^{1/2}\Big\langle \frac{1}{\sqrt{n}}\sum_{j=1}^{n} g_j X_j', v\Big\rangle,$$

  where $\hat{\Sigma}$ is the sample covariance based on $X_1', \dots, X_n'$ and $\{g_j\}$ are i.i.d, $N(0,1)$ r.v. independent of $\{X_j\}, \{X_j'\}$

- Conditionally on $X_j', j = 1, \dots, n$, $(u,v) \mapsto Y(u,v)$ and $(u,v) \mapsto Z(u,v)$ are mean zero Gaussian processes

Proof of the Upper Bound: Gaussian Case

Gaussian Comparison Inequality (Slepian-Fernique-Sudakov): conditionally on $X_1', \dots, X_n'$,

- $$\mathbb{E}_{X,g}(Y(u,v) - Y(u',v'))^2 \leq \mathbb{E}_{X,g}(Z(u,v) - Z(u',v'))^2.$$

- $$\mathbb{E}_{X,g} \sup_{\|u\|,\|v\|\leq 1} Y(u,v) \leq \mathbb{E}_{X,g} \sup_{\|u\|,\|v\|\leq 1} Z(u,v)$$
$$\leq \sqrt{2}\|\hat{\Sigma}'\|^{1/2}\mathbb{E}\|X\| + \sqrt{2}\|\Sigma\|^{1/2}\mathbb{E}_g\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^n g_j X_j'\right\|.$$

Proof of the Upper Bound: Gaussian Case

- Combining this with the decoupling inequality and using the fact that

$$\mathbb{E}\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} g_j X_j'\right\| = \mathbb{E}\left(n^{-1}\sum_{j=1}^{n} g_j^2\right)^{1/2}\mathbb{E}\|X\| \leq \mathbb{E}\|X\|,$$

- we get

$$\begin{aligned}
\Delta &:= \mathbb{E}\|\hat{\Sigma} - \Sigma\| \\
&\leq 2\sqrt{2}\Delta^{1/2}\frac{\mathbb{E}\|X\|}{\sqrt{n}} + 4\sqrt{2}\|\Sigma\|^{1/2}\frac{\mathbb{E}\|X\|}{\sqrt{n}} \\
&\leq 2\sqrt{2}\Delta^{1/2}\|\Sigma\|^{1/2}\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} + 4\sqrt{2}\|\Sigma\|\sqrt{\frac{\mathbf{r}(\Sigma)}{n}},
\end{aligned}$$

and the upper bound follows by solving the above inequality w.r.t. Δ.

Proof of the Upper Bound: Subgaussian Case

The proof is based on generic chaining bounds for empirical processes indexed by squares of functions (Mendelson (2012))

Preliminaries: Generic Chaining

- (T, d) a metric space
- $\{\Delta_n\}$ an increasing sequence of partitions of T
- $\{\Delta_n\}$ admissible iff $\mathrm{card}(\Delta_n) \leq N_n$, where $N_n := 2^{2^n}, n \geq 1$, $N_0 := 1$.
- For $t \in T$, $\Delta_n(t)$ denotes the unique set of the partition Δ_n that contains t.
- $A \subset T$, $D(A)$ denotes the diameter of set A.
- Generic Chaining Complexity

$$\gamma_2(T, d) = \inf \sup_{t \in T} \sum_{n=0}^{\infty} 2^{n/2} D(\Delta_n(t)),$$

where the infimum is taken over all admissible sequences.

Preliminaries: Talagrand Theorem

Theorem

Let $X(t), t \in T$ be a centered Gaussian process and suppose that

$$d(t,s) := \mathbb{E}^{1/2}(X(t) - X(s))^2, t, s \in T.$$

Then, there exists an absolute constant $K > 0$ such that

$$K^{-1}\gamma_2(T; d) \leq \mathbb{E} \sup_{t \in T} X(t) \leq K\gamma_2(T; d)$$

Preliminaries: Mendelson (2010)

Theorem

Let $X, X_1, \dots, X_n$ be i.i.d. random variables in S with common distribution P and let $\mathcal{F}$ be a class of measurable functions on $(S, \mathcal{A})$ such that $f \in \mathcal{F}$ implies $-f \in \mathcal{F}$ and $\mathbb{E}f(X) = 0$. Then

$$\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f^2(X_i) - \mathbb{E}f^2(X)\right| \lesssim \max\left\{\sup_{f\in\mathcal{F}}\|f\|_{\psi_1}\frac{\gamma_2(\mathcal{F};\psi_2)}{\sqrt{n}}, \frac{\gamma_2^2(\mathcal{F};\psi_2)}{n}\right\}.$$

Proof of the Upper Bound: Subgaussian Case

- $$\mathbb{E}\|\hat{\Sigma} - \Sigma\| = \mathbb{E} \sup_{\|u\|\leq 1} \left| \frac{1}{n} \sum_{i=1}^{n} \langle X_i, u\rangle^2 - \langle \Sigma u, u\rangle \right|$$
  $$= \mathbb{E} \sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f^2(X_i) - \mathbb{E} f^2(X) \right|,$$
  where $\mathcal{F} := \Big\{\langle \cdot, u\rangle : u \in U_{E^*}\Big\}$, $U_{E^*} := \{u \in E^* : \|u\| \leq 1\}$ and P is the distribution of random variable X.
- Since X is subgaussian,
  $$\Big\|\langle X, u\rangle\Big\|_{\psi_2} \asymp \Big\|\langle X, u\rangle\Big\|_{\psi_1} \asymp \Big\|\langle X, u\rangle\Big\|_{L_2(\mathbb{P})}.$$
- Therefore,
  $$\sup_{f\in\mathcal{F}} \|f\|_{\psi_1} \lesssim \sup_{u\in U_{E^*}} \mathbb{E}^{1/2}\langle X, u\rangle^2 \leq \|\Sigma\|^{1/2}.$$

Proof of the Upper Bound: Subgaussian Case

- Also, since X is pregaussian, there exists $Y \sim N(0, \Sigma)$
- $$d_Y(u,v) = \|\langle \cdot, u\rangle - \langle \cdot, v\rangle\|_{L_2(P)}, u, v \in U_{E^*}.$$
- Using Talagrand Theorem,

$$\gamma_2(\mathcal{F}, \psi_2) \lesssim \gamma_2(\mathcal{F}, L_2) = \gamma_2(U_{E^*}; d_Y) \lesssim \mathbb{E} \sup_{u \in U_{E^*}} \langle Y, u\rangle \leq \mathbb{E}\|Y\|.$$

Therefore,

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \lesssim \max\left\{\|\Sigma\|^{1/2}\frac{\mathbb{E}\|Y\|}{\sqrt{n}}, \frac{(\mathbb{E}\|Y\|)^2}{n}\right\}$$

$$\lesssim \|\Sigma\| \max\left\{\sqrt{\frac{\mathbf{r}(\Sigma)}{n}}, \frac{\mathbf{r}(\Sigma)}{n}\right\}.$$

Proof of the Lower Bound

- $$\mathbb{E}\|\hat{\Sigma}-\Sigma\| \geq \sup_{\|u\|\leq 1} \mathbb{E}\Bigl\|n^{-1}\sum_{j=1}^{n}\langle X_j,u\rangle X_j - \mathbb{E}\langle X,u\rangle X\Bigr\|.$$
- For a fixed $u \in E^*$ with $\|u\| \leq 1$ and $\langle \Sigma u, u\rangle > 0$, define
$$X' := X - \langle X,u\rangle \frac{\Sigma u}{\langle \Sigma u,u\rangle}, \quad X_j' := X_j - \langle X_j,u\rangle \frac{\Sigma u}{\langle \Sigma u,u\rangle}, j=1,\dots,n.$$
$\{X', X_j' : j=1,\dots,n\}$ and $\{\langle X,u\rangle, \langle X_j,u\rangle : j=1,\dots,n\}$ are independent.
- $$\mathbb{E}\Bigl\|n^{-1}\sum_{j=1}^{n}\langle X_j,u\rangle X_j - \mathbb{E}\langle X,u\rangle X\Bigr\| =$$
$$\mathbb{E}\Bigl\|n^{-1}\sum_{j=1}^{n}(\langle X_j,u\rangle^2 - \mathbb{E}\langle X,u\rangle^2)\frac{\Sigma u}{\langle \Sigma u,u\rangle} + n^{-1}\sum_{j=1}^{n}\langle X_j,u\rangle X_j'\Bigr\|,$$

Proof of the Lower Bound

- Conditionally on $\langle X_j, u\rangle, j = 1, \dots, n$, the distribution of r.v.

$$n^{-1}\sum_{j=1}^{n}\langle X_j, u\rangle X_j'$$

is Gaussian and it coincides with the distribution of r.v.

$$\left(n^{-1}\sum_{j=1}^{n}\langle X_j, u\rangle^2\right)^{1/2}\frac{X'}{\sqrt{n}}.$$

-

$$\mathbb{E}\left\|n^{-1}\sum_{j=1}^{n}(\langle X_j, u\rangle^2 - \mathbb{E}\langle X, u\rangle^2)\frac{\Sigma u}{\langle \Sigma u, u\rangle} + n^{-1}\sum_{j=1}^{n}\langle X_j, u\rangle X_j'\right\|$$

$$= \mathbb{E}\left\|n^{-1}\sum_{j=1}^{n}(\langle X_j, u\rangle^2 - \mathbb{E}\langle X, u\rangle^2)\frac{\Sigma u}{\langle \Sigma u, u\rangle} + \left(n^{-1}\sum_{j=1}^{n}\langle X_j, u\rangle^2\right)^{1/2}\frac{X'}{\sqrt{n}}\right\|.$$

Proof of the Lower Bound

Denote $\mathbb{E}'$ the conditional expectation given $X_1', \dots, X_n'$.

$$\mathbb{E}\left\| n^{-1}\sum_{j=1}^n (\langle X_j, u\rangle^2 - \mathbb{E}\langle X, u\rangle^2)\frac{\Sigma u}{\langle \Sigma u, u\rangle} + \left(n^{-1}\sum_{j=1}^n \langle X_j, u\rangle^2\right)^{1/2}\frac{X'}{\sqrt{n}}\right\|$$

$$= \mathbb{E}\mathbb{E}'\left\| n^{-1}\sum_{j=1}^n (\langle X_j, u\rangle^2 - \mathbb{E}\langle X, u\rangle^2)\frac{\Sigma u}{\langle \Sigma u, u\rangle} + \left(n^{-1}\sum_{j=1}^n \langle X_j, u\rangle^2\right)^{1/2}\frac{X'}{\sqrt{n}}\right\|$$

$$\geq \mathbb{E}\left\| \mathbb{E}' n^{-1}\sum_{j=1}^n (\langle X_j, u\rangle^2 - \mathbb{E}\langle X, u\rangle^2)\frac{\Sigma u}{\langle \Sigma u, u\rangle} + \mathbb{E}'\left(n^{-1}\sum_{j=1}^n \langle X_j, u\rangle^2\right)^{1/2}\frac{X'}{\sqrt{n}}\right\|$$

$$= \mathbb{E}\left(n^{-1}\sum_{j=1}^n \langle X_j, u\rangle^2\right)^{1/2}\frac{\mathbb{E}\|X'\|}{\sqrt{n}}.$$

Proof of the Lower Bound

$$\mathbb{E}\|X'\| \geq \mathbb{E}\|X\| - \mathbb{E}|\langle X, u\rangle| \frac{\|\Sigma u\|}{\langle \Sigma u, u\rangle} = \mathbb{E}\|X\| - \sqrt{\frac{2}{\pi}} \frac{\|\Sigma u\|}{\langle \Sigma u, u\rangle^{1/2}}$$

and

$$\mathbb{E}\left\| n^{-1}\sum_{j=1}^{n}(\langle X_j, u\rangle^2 - \mathbb{E}\langle X, u\rangle^2)\frac{\Sigma u}{\langle \Sigma u, u\rangle} + \left(n^{-1}\sum_{j=1}^{n}\langle X_j, u^2\rangle\right)^{1/2}\frac{X'}{\sqrt{n}}\right\|$$

$$\geq \langle \Sigma u, u\rangle^{1/2}\mathbb{E}\left(n^{-1}\sum_{j=1}^{n} Z_j^2\right)^{1/2}\frac{\mathbb{E}\|X\| - \sqrt{\frac{2}{\pi}}\frac{\|\Sigma u\|}{\langle \Sigma u, u\rangle^{1/2}}}{\sqrt{n}},$$

where

$$Z_j = \frac{\langle X_j, u\rangle}{\langle \Sigma u, u\rangle^{1/2}}, j = 1, \dots, n \text{ i.i.d.} \sim N(0, 1)$$

Proof of the Lower Bound

Since

$$\mathbb{E}\left(n^{-1}\sum_{j=1}^{n} Z_j^2\right)^{1/2} \geq c_2 > 0,$$

$$\mathbb{E}\left\|n^{-1}\sum_{j=1}^{n}(\langle X_j, u\rangle^2 - \mathbb{E}\langle X, u\rangle^2)\frac{\Sigma u}{\langle \Sigma u, u\rangle} + \left(n^{-1}\sum_{j=1}^{n}\langle X_j, u\rangle^2\right)^{1/2}\frac{X'}{\sqrt{n}}\right\|$$

$$\geq c_2 \frac{\langle \Sigma u, u\rangle^{1/2}\mathbb{E}\|X\| - \sqrt{\frac{2}{\pi}}\|\Sigma u\|}{\sqrt{n}}.$$

Therefore

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \geq c_2 \sup_{\|u\|\leq 1} \frac{\langle \Sigma u, u\rangle^{1/2}\mathbb{E}\|X\| - \sqrt{\frac{2}{\pi}}\|\Sigma u\|}{\sqrt{n}}$$

$$\geq c_2 \frac{\|\Sigma\|^{1/2}\mathbb{E}\|X\| - \sqrt{\frac{2}{\pi}}\|\Sigma\|}{\sqrt{n}} \geq c_2\|\Sigma\|\left(\frac{c_3\sqrt{\mathbf{r}(\Sigma)} - \sqrt{\frac{2}{\pi}}}{\sqrt{n}}\right).$$

Proof of the Lower Bound

Also,

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \geq \sup_{\|u\|\leq 1} \left| n^{-1} \sum_{j=1}^{n} \langle X_j, u\rangle^2 - \mathbb{E}\langle X, u\rangle^2 \right|$$

$$= \sup_{\|u\|\leq 1} \langle \Sigma u, u\rangle \mathbb{E} \left| n^{-1} \sum_{j=1}^{n} Z_j^2 - 1 \right| \geq c_4 \frac{\|\Sigma\|}{\sqrt{n}},$$

implying that, for small enough c_2,

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \geq c_2 \|\Sigma\| \left(\frac{c_3 \sqrt{\mathbf{r}(\Sigma)} - \sqrt{\frac{2}{\pi}}}{\sqrt{n}} \right) \bigvee c_4 \frac{\|\Sigma\|}{\sqrt{n}}$$

$$\geq \frac{1}{2} \left(c_2 \|\Sigma\| \left(\frac{c_3 \sqrt{\mathbf{r}(\Sigma)} - \sqrt{\frac{2}{\pi}}}{\sqrt{n}} \right) + c_4 \frac{\|\Sigma\|}{\sqrt{n}} \right) \geq \frac{c_2}{2} \|\Sigma\| \frac{c_3 \sqrt{\mathbf{r}(\Sigma)}}{\sqrt{n}}.$$

Proof of the Lower Bound

On the other hand, if $\mathbf{r}(\Sigma) \geq 2n$,

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \geq \mathbb{E}\|\hat{\Sigma}\| - \|\Sigma\| \geq \mathbb{E} \sup_{\|u\|\leq 1} n^{-1}\sum_{j=1}^{n}\langle X_j, u\rangle^2 - \|\Sigma\|$$

$$\geq \mathbb{E} \sup_{\|u\|\leq 1} \frac{\langle X_1, u\rangle^2}{n} - \|\Sigma\| \geq \frac{\mathbb{E}\|X\|^2}{n} - \|\Sigma\|$$

$$= \|\Sigma\|\left(\frac{\mathbf{r}(\Sigma)}{n} - 1\right) \geq \frac{1}{2}\|\Sigma\|\frac{\mathbf{r}(\Sigma)}{n}.$$

Concentration inequality (Koltchinskii and Lounici (2014))

Theorem

Let M be either the median, or the mean of $\|\hat{\Sigma} - \Sigma\|$. There exists a constant $C > 0$ such that, for all $t \geq 1$ with probability at least $1 - e^{-t}$, the following bound holds:

$$\Big|\|\hat{\Sigma} - \Sigma\| - M\Big| \leq C\left[\|\Sigma\|\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee 1\right)\sqrt{\frac{t}{n}} \bigvee \|\Sigma\|\frac{t}{n}\right].$$

Corrolaries

Corollary

There exists a constant $C > 0$ *such that, for all* $t \geq 1$*, with probability at least* $1 - e^{-t}$,

$$\|\hat{\Sigma} - \Sigma\| \leq C\|\Sigma\| \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee \frac{\mathbf{r}(\Sigma)}{n} \bigvee \sqrt{\frac{t}{n}} \bigvee \frac{t}{n} \right).$$

This implies that for all $p \geq 1$

$$\mathbb{E}^{1/p}\|\hat{\Sigma} - \Sigma\|^p \lesssim_p \|\Sigma\| \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee \frac{\mathbf{r}(\Sigma)}{n} \right).$$

Proof of concentration inequality:

- The proof is based on Gaussian concentration
- Another proof: based on a concentration inequality for sup-norms of Gaussian chaos, Adamczak (2014)

Proof of concentration Inequality

Theorem

Let $X, X_1, \dots, X_n$ be i.i.d. centered Gaussian random vectors in E with covariance Σ and let $M := \mathrm{Med}(\|\hat{\Sigma} - \Sigma\|)$. Then, there exist constants $C > 0$ such that for all $t \geq 1$ with probability at least $1 - e^{-t}$,

$$\Big| \|\hat{\Sigma} - \Sigma\| - M \Big| \leq C \left[\|\Sigma\| \left(\sqrt{\frac{t}{n}} \bigvee \frac{t}{n} \right) \bigvee \|\Sigma\|^{1/2} M^{1/2} \sqrt{\frac{t}{n}} \right].$$

Note that

$$M = \mathrm{Med}(\|\hat{\Sigma} - \Sigma\|) \leq 2\mathbb{E}\|\hat{\Sigma} - \Sigma\| \lesssim \|\Sigma\| \left[\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee \frac{\mathbf{r}(\Sigma)}{n} \right],$$

implying the concentration inequality in the explicit form.

Proof of Concentration Inequality, reduction to the finite-dimensional case

Theorem

Let X be a centered Gaussian random variable in a separable Banach space E. Then there exists a sequence $\{x_k : k \geq 1\}$ of vectors in E and a sequence $\{Z_k : k \geq 1\}$ of i.i.d. standard normal random variables such that $X = \sum_{k=1}^{\infty} Z_k x_k$, where the series in the right hand side converges in E a.s. and $\sum_{k=1}^{\infty} \|x_k\|^2 < +\infty$.

It easily follows from this result that it is enough to proof the concentration inequality when

$$X = \sum_{k=1}^{m} Z_k x_k, X_j = \sum_{k=1}^{m} Z_{k,j} x_k, Z := (Z_{k,j}, k = 1, \dots, m, j = 1, \dots, n).$$

Gaussian Concentration

Denote

$$f(Z) =: g(X_1, \dots, X_n) := \|W\|\varphi\Big(\frac{\|W\|}{\delta}\Big),$$

where

- $W = \hat{\Sigma} - \Sigma,$
- φ is a Lipschitz function with constant 1 on $\mathbb{R}_+$, $0 \le \varphi(s) \le 1$, $\varphi(s) = 1, s \le 1, \varphi(s) = 0, s > 2,$
- $\delta > 0$ is fixed

Lemma

There exists a numerical constant $D > 0$ such that, for all $Z, Z' \in \mathbb{R}^{mn}$,

$$|f(Z) - f(Z')| \le D\frac{\|\Sigma\| + \|\Sigma\|^{1/2}\sqrt{\delta}}{\sqrt{n}}\Big(\sum_{j=1}^{n}\sum_{k=1}^{m}|Z_{k,j} - Z'_{k,j}|^2\Big)^{1/2}.$$

By Gaussian concentration inequality, for all $t \geq 1$ with probability at least $1 - e^{-t}$,

$$\Big| g(X_1, \dots, X_n) - \mathrm{Med}(g(X_1, \dots, X_n)) \Big| \leq D_1 \left(\|\Sigma\| + \|\Sigma\|^{1/2} \sqrt{\delta} \right) \sqrt{\frac{t}{n}}.$$

where D_1 is a numerical constant. It follows that, on the event $\|W\| \leq \delta$,

$$\|W\| = g(X_1, \dots, X_n) \leq \mathrm{Med}(g(X_1, \dots, X_n)) + D_1 \left(\|\Sigma\| + \|\Sigma\|^{1/2} \sqrt{\delta} \right) \sqrt{\frac{t}{n}}$$

$$\leq \mathrm{Med}(\|W\|) + D_1 \left(\|\Sigma\| + \|\Sigma\|^{1/2} \sqrt{\delta} \right) \sqrt{\frac{t}{n}} =: A + B\sqrt{\delta},$$

where

$$A := \mathrm{Med}(\|W\|) + D_1 \|\Sigma\| \sqrt{\frac{t}{n}}, \quad B := D_1 \|\Sigma\|^{1/2} \sqrt{\frac{t}{n}}.$$

Then we have

$$\mathbb{P}\left\{ \delta \geq \|W\| \geq A + B\sqrt{\delta} \right\} \leq e^{-t}.$$

Proof of Concentration Inequality: Iterative Bounds

- Denote
$$\delta_0 := D_2\|\Sigma\|\Bigl[\mathbf{r}(\Sigma)\Bigl(\sqrt{\frac{t}{n}}\bigvee\frac{t}{n}\Bigr)+\mathbf{r}(\Sigma)+1\Bigr].$$
- It is easy to prove that
$$\mathbb{P}\Bigl\{\|W\|\geq\delta_0\Bigr\}\leq e^{-t}$$
- Define δ_k for $k\geq 1$ as follows:
$$\delta_k = A + B\sqrt{\delta_{k-1}}.$$
- It is easy to check that $\{\delta_k\}$ is decreasing with limit $\bar{\delta}$,
$$\bar{\delta} = A + B\sqrt{\bar{\delta}},\ \ \bar{\delta}\lesssim A\vee B^2.$$
- Moreover,
$$\delta_k - \bar{\delta} \leq u_k := B^2\Bigl(\frac{\delta_0}{B^2}\Bigr)^{2^{-k}}.$$

Proof of Concentration Inequality: Iterative Bounds

- Let $\bar{k} := \min\left\{k : \left(\frac{\delta_0}{B^2}\right)^{2^{-k}} \le 2\right\}$. Then

$$\delta_{\bar{k}} \lesssim A \vee B^2, \quad \bar{k} \lesssim \log\log(c_1 \mathbf{r}(\Sigma)) \bigvee \log\log(c_1 n)$$

- Since $\mathbb{P}\left\{\delta_{k-1} > \|W\| \ge \delta_k\right\} \le e^{-t}$, we get that with probability at least $1 - (\bar{k}+1)e^{-t}$,

$$\|W\| \le \delta_{\bar{k}} \lesssim A \vee B^2 \lesssim \mathrm{Med}(\|W\|) \bigvee \|\Sigma\| \left(\sqrt{\frac{t}{n}} \bigvee \frac{t}{n}\right)$$

- With a bit more effort, it follows that with probability at least $1 - e^{-t}$

$$\|W\| = \|\hat{\Sigma} - \Sigma\| \le C\left[\mathrm{Med}(\|W\|) \bigvee \|\Sigma\| \left(\sqrt{\frac{t}{n}} \bigvee \frac{t}{n}\right)\right].$$

Gaussian Concentration

- Take

$$\delta := C\left[\mathrm{Med}(\|W\|) \bigvee \|\Sigma\| \left(\sqrt{\frac{t+2n}{n}} \bigvee \frac{t+2n}{n}\right)\right].$$

- Since $\mathbb{P}\{\|W\| \geq \delta\} \leq 2e^{-t-2n} \leq 1/4$,

$$\mathbb{P}\{g(X_1, \dots, X_n) \geq \mathrm{Med}(\|W\|)\} \geq 1/4,$$
$$\mathbb{P}\{g(X_1, \dots, X_n) \leq \mathrm{Med}(\|W\|)\} \geq 1/2.$$

- To complete the proof, note that, by Gaussian isoperimetric inequality, on the event where $\|W\| \leq \delta$,

$$\left|\|\hat{\Sigma} - \Sigma\| - \mathrm{Med}(\|W\|)\right| = \left|g(X_1, \dots, X_n) - \mathrm{Med}(\|W\|)\right|$$
$$\leq D_1\left(\|\Sigma\| + \|\Sigma\|^{1/2}\sqrt{\delta}\right)\sqrt{\frac{t}{n}}$$

with probability at least $1 - e^{-t}$.

Generic Chaining Tail Bound: Dirksen (2014), Bednorz (2014), Mendelson (2013, 2015)

Theorem

Let $X, X_1, \dots, X_n$ be i.i.d. random variables in S with common distribution P and let $\mathcal{F}$ be a class of measurable functions on $(S, \mathcal{A})$. Then, there exists a constant $C > 0$ such that for all $t \geq 1$ with probability at least $1 - e^{-t}$

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n f^2(X_i) - \mathbb{E}f^2(X) \right|$$
$$\leq C \max\left\{ \sup_{f\in\mathcal{F}} \|f\|_{\psi_2} \frac{\gamma_2(\mathcal{F};\psi_2)}{\sqrt{n}}, \frac{\gamma_2^2(\mathcal{F};\psi_2)}{n}, \sup_{f\in\mathcal{F}} \|f\|_{\psi_2}^2 \sqrt{\frac{t}{n}}, \sup_{f\in\mathcal{F}} \|f\|_{\psi_2}^2 \frac{t}{n} \right\}.$$

Part 2.

Taylor Expansions of Operator Functions and Normal Approximation of Plug-In Estimators of Smooth Functionals of Covariance

Problems

- Let $\mathbb{H}$ be a separable Hilbert space
- Let $X, X_1, \dots, X_n$ be i.i.d. Gaussian vectors with values in $\mathbb{H}$ with $\mathbb{E}X = 0$ and with covariance operator $\Sigma = \mathbb{E}(X \otimes X)$.
- Problems
  - Given a smooth function $f : \mathbb{R} \mapsto \mathbb{R}$ and a nuclear operator $B : \mathbb{H} \mapsto \mathbb{H}$, estimate $\langle f(\Sigma), B\rangle$ based on $X_1, \dots, X_n$.

Sample Covariance Operator and Effective Rank

- Let
$$\hat{\Sigma} := n^{-1} \sum_{j=1}^{n} X_j \otimes X_j$$
be the sample covariance based on $(X_1, \dots, X_n)$.
- Effective Rank:
$$\mathbf{r}(\Sigma) = \frac{\mathrm{tr}(\Sigma)}{\|\Sigma\|}$$
- $\mathbf{r}(\Sigma) \leq \mathrm{rank}(\Sigma) \leq \dim(\mathbb{H})$

Expectation bounds in terms of effective rank

Theorem (Koltchinskii and Lounici (2014))

Let $X, X_1, \dots, X_n$ be i.i.d centered Gaussian random vectors in $\mathbb{H}$ with covariance operator Σ. Then

$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \asymp \|\Sigma\| \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee \frac{\mathbf{r}(\Sigma)}{n} \right).$$

Concentration Inequality

Theorem (Koltchinskii and Lounici (2014))

There exists a constant $C > 0$ such that, for all $t \geq 1$ with probability at least $1 - e^{-t}$, the following bound holds:

$$\Big| \|\hat{\Sigma} - \Sigma\| - M \Big| \leq C\|\Sigma\| \left[\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee 1 \right) \sqrt{\frac{t}{n}} \bigvee \frac{t}{n} \right].$$

where M is either the mean, or the median of $\|\hat{\Sigma} - \Sigma\|$.

Remark

The results are also true for Gaussian random variables in separable Banach spaces with $\mathbf{r}(\Sigma) := \frac{\mathbb{E}\|X\|^2}{\|\Sigma\|}$.

Further exponential and moment bounds

Corollary

There exists a constant $C > 0$ *such that, for all* $t \geq 1$, *with probability at least* $1 - e^{-t}$,

$$\|\hat{\Sigma} - \Sigma\| \leq C\|\Sigma\|\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee \frac{\mathbf{r}(\Sigma)}{n} \bigvee \sqrt{\frac{t}{n}} \bigvee \frac{t}{n}\right).$$

This implies that for all $p \geq 1$

$$\mathbb{E}^{1/p}\|\hat{\Sigma} - \Sigma\|^p \lesssim_p \|\Sigma\|\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee \frac{\mathbf{r}(\Sigma)}{n}\right).$$

Normal Approximation Bounds for Smooth Functions of Sample Covariance

Problems

Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a given smooth function and B be a given operator with $\|B\|_1 < \infty$

- For given f and B, show that the distribution of r.v. $\frac{n^{1/2}\langle f(\hat{\Sigma}) - \mathbb{E}f(\hat{\Sigma}), B\rangle}{\sigma_f(\Sigma; B)}$ is close to standard normal for a proper $\sigma_f(\Sigma; B)$ when $n \to \infty$ and $\mathbf{r}(\Sigma) = o(n)$
- Is the plug-in estimator $\langle f(\hat{\Sigma}), B\rangle$ asymptotically efficient (is $\sqrt{n}\langle f(\hat{\Sigma}) - f(\Sigma), B\rangle$ asymptotically normal with limit variance $\sigma_f(\Sigma; B)$ as small as possible)?

Entire Functions of Exponential Type

- $f : \mathbb{C} \mapsto \mathbb{C}$ be an entire function
- For $\sigma > 0$, f is of exponential type σ if $\forall \varepsilon > 0\ \exists C = C(\varepsilon, \sigma, f) > 0$ such that
$$|f(z)| \leq Ce^{(\sigma+\varepsilon)|z|}, z \in \mathbb{C}.$$
- $\mathcal{E}_\sigma = \mathcal{E}_\sigma(\mathbb{C})$ denotes the space of all entire functions of exponential type σ.
- According to *Paley-Wiener theorem*,
$$\mathcal{E}_\sigma \bigcap L_\infty(\mathbb{R}) = \{f \in L_\infty(\mathbb{R}) : \mathrm{supp}(\mathcal{F}f) \subset [-\sigma, \sigma]\}.$$
- *Bernstein inequality:* $\forall f \in \mathcal{E}_\sigma \bigcap L_\infty(\mathbb{R})$
$$\|f'\|_{L_\infty(\mathbb{R})} \leq \sigma \|f\|_{L_\infty(\mathbb{R})}.$$

Littlewood-Paley Decomposition

- Let $w \in C^\infty(\mathbb{R})$, $w \geq 0$, $\mathrm{supp}(w) \subset [-2, 2]$, $w(t) = 1, t \in [-1, 1]$ and $w(-t) = w(t), t \in \mathbb{R}$.
- $w_0(t) := w(t/2) - w(t), t \in \mathbb{R}, \mathrm{supp}(w_0) \subset \{t : 1 \leq |t| \leq 4\}$
- $w_j(t) := w_0(2^{-j}t), t \in \mathbb{R}, \mathrm{supp}(w_j) \subset \{t : 2^j \leq |t| \leq 2^{j+2}\}, j \geq 0.$
- Then $w(t) + \sum_{j \geq 0} w_j(t) = 1, t \in \mathbb{R}.$
- Let $W, W_j \in \mathcal{S}(\mathbb{R})$,

$$w(t) = (\mathcal{F}W)(t),\ w_j(t) = (\mathcal{F}W_j)(t), t \in \mathbb{R}, j \geq 0.$$

- For $f \in \mathcal{S}'(\mathbb{R})$, define its Littlewood-Paley dyadic decomposition:

$$f_0 := f * W,\ f_n := f * W_{n-1}, n \geq 1$$

- Note that $f_n \in \mathcal{E}_{2^{n+1}} \bigcap L_\infty(\mathbb{R})$ and

$$\sum_{n \geq 0} f_n = f$$

with convergence of the series in the space $\mathcal{S}'(\mathbb{R})$.

Besov Spaces

- *Besov norms:*

$$\|f\|_{B^s_{\infty,1}} := \sum_{n\geq 0} 2^{ns}\|f_n\|_{L_\infty(\mathbb{R})}, s \in \mathbb{R}$$

- *Besov spaces:*

$$B^s_{\infty,1}(\mathbb{R}) := \Big\{ f \in \mathcal{S}'(\mathbb{R}) : \|f\|_{B^s_{\infty,1}} < \infty \Big\}.$$

- If $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \geq 0$, then $\sum_{n\geq 0} f_n$ converges uniformly to f in $\mathbb{R}$, which easily implies that $f \in C_u(\mathbb{R})$ and

$$\|f\|_{L_\infty} \leq \|f\|_{B^s_{\infty,1}}.$$

Perturbation Theory: Operator Lipschitz and Operator Differentiable Functions

- $\mathcal{B}_{sa}(\mathbb{H})$ the space of self-adjoint bounded operators in $\mathbb{H}$
- A continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ is called operator Lipschitz with respect to the operator norm iff there exists a constant $L_f > 0$ such that for all $A, B \in \mathcal{B}_{sa}(\mathbb{H})$

$$\|f(A) - f(B)\| \leq L_f \|A - B\|.$$

- If f is operator Lipschitz, then it is Lipschitz; however, $f(t) = |t|$ is not operator Lipschitz (Kato (1972)).
- A continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ is called operator differentiable iff $\mathcal{B}_{sa}(\mathbb{H}) \ni A \mapsto f(A) \in \mathcal{B}_{sa}(\mathbb{H})$ is Fréchet differentiable at any $A \in L(\mathbb{H})$, i.e., there exists a bounded linear mapping $\mathcal{B}_{sa}(\mathbb{H}) \ni E \mapsto Df(A; E) = Df(A)E \in \mathcal{B}_{sa}(\mathbb{H})$ such that

$$f(A+E) - f(A) = Df(A; E) + o(\|E\|) \text{ as } \|E\| \to 0.$$

Perturbation Theory: Operator Lipschitz and Operator Differentiable Functions

Theorem (Peller)

If $f \in B^1_{\infty,1}(\mathbb{R})$ then f is operator Lipschitz with Lipschitz constant $L_f = \|f\|_{B^1_{\infty,1}}$ and operator differentiable.

Moreover, if $A \in \mathcal{B}_{sa}(\mathbb{H})$ is an operator with spectral decomposition

$$A = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda,$$

then (Loewner, Daletsky-Krein)

$$Df(A; E) = \sum_{\lambda, \mu \in \sigma(A)} f^{[1]}(\lambda, \mu) P_\lambda E P_\mu,$$

where

$$f^{[1]}(\lambda, \mu) := \frac{f(\lambda) - f(\mu)}{\lambda - \mu}, \ \lambda \neq \mu; \ \ f^{[1]}(\lambda, \lambda) := f'(\lambda).$$

Perturbation Theory: Bounds on the Remainder of Differentiation

Lemma

Let $S_f(A;E) = f(A+E) - f(A) - (Df)(A;E)$ *be the remainder of differentiation. If, for some* $s \in [1,2]$, $f \in B^s_{\infty,1}(\mathbb{R})$, *then the following bounds hold:*

$$\|S_f(A;E)\| \leq 2^{3-s}\|f\|_{B^s_{\infty,1}}\|E\|^s$$

and

$$\|S_f(A;E) - S_f(A;E')\| \leq 2^{1+s}\|f\|_{B^s_{\infty,1}}(\|E\| \vee \|E'\|)^{s-1}\|E' - E\|.$$

The proof is based on Littlewood-Paley decomposition of f and on operator versions of Bernstein inequalities for entire functions of exponential type (as in the work by Peller on operator Lipschitz functions).

Bounds for Entire Functions of Exponential Type

Lemma

Let $f \in \mathcal{E}_\sigma \bigcap L_\infty(\mathbb{R})$. *Then, for all* $A, H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,

$$\|f(A+H) - f(A)\| \le \sigma \|f\|_{L_\infty(\mathbb{R})} \|H\|,$$

$$\|Df(A;H)\| \le \sigma \|f\|_{L_\infty(\mathbb{R})} \|H\|,$$

$$\|S_f(A;H)\| \le \frac{\sigma^2}{2} \|f\|_{L_\infty(\mathbb{R})} \|H\|^2$$

and

$$\|S_f(A;H') - S_f(A;H)\| \le \sigma^2 \|f\|_{L_\infty(\mathbb{R})} \delta(H,H') \|H' - H\|,$$

where

$$\delta(H,H') := (\|H\| \wedge \|H'\|) + \frac{\|H' - H\|}{2}.$$

Proof of Operator Lipschitz Property

- E a complex Banach space
- $\mathcal{E}_\sigma(E)$ the space of entire functions $F : \mathbb{C} \mapsto E$ of exponential type σ : $F \in \mathcal{E}_\sigma(E)$ iff $\forall \varepsilon > 0\ \exists C = C(\varepsilon, \sigma, F) > 0$:

$$\|F(z)\| \leq Ce^{(\sigma+\varepsilon)|z|}, z \in \mathbb{C}.$$

- If $F \in \mathcal{E}_\sigma(E)$ and $\sup_{x\in\mathbb{R}} \|F(x)\| < +\infty$, then Bernstein inequality holds for F :

$$\sup_{x\in\mathbb{R}} \|F'(x)\| \leq \sigma \sup_{x\in\mathbb{R}} \|F(x)|$$

and

$$\|F(x+h) - F(x)\| \leq \sigma \sup_{x\in\mathbb{R}} \|F(x)\||h|.$$

Proof of Operator Lipschitz Property

- Given $A, H \in \mathcal{B}_{sa}(\mathbb{H})$ and $f \in \mathcal{E}_\sigma \bigcap L_\infty(\mathbb{R})$, define

$$F(z) := f(A + zH), z \in \mathbb{C}.$$

- Then $F \in \mathcal{E}_{\sigma\|H\|}(\mathcal{B}(\mathbb{H}))$. Indeed, F is complex-differentiable at any point $z \in \mathbb{C}$ with derivative $F'(z) = Df(A + zH; H)$ and, by von Neumann theorem,

$$\|F(z)\| = \|f(A{+}zH)\| \leq \sup_{|\zeta|\leq\|A\|+|z|\|H\|} |f(\zeta)| \leq \|f\|_{L_\infty(\mathbb{R})} e^{\sigma\|A\|} e^{\sigma\|H\||z|},$$

implying that F is of exponential type $\sigma\|H\|$.

Proof of Operator Lipschitz Property

- Note also that

$$\sup_{x\in\mathbb{R}} \|F(x)\| = \sup_{x\in\mathbb{R}} \|f(A+xH)\| \leq \sup_{x\in\mathbb{R}} |f(x)| = \|f\|_{L_\infty(\mathbb{R})}.$$

- Hence

$$\|f(A+H) - f(A)\| = \|F(1) - F(0)\| \leq \sup_{x\in\mathbb{R}} \|F'(x)\|$$

$$\leq \sigma\|H\| \sup_{x\in\mathbb{R}} \|F(x)\| \leq \sigma\|f\|_{L_\infty(\mathbb{R})}\|H\|.$$

Proof Operator Lipschitz Property for $f \in B^1_{\infty,1}(\mathbb{R})$

- For $f \in B^1_{\infty,1}(\mathbb{R})$, the series $\sum_{n\geq 0} f_n$ converges uniformly in $\mathbb{R}$ to function f.
- Since $A, A+H, A+H'$ are bounded self-adjoint operators, we also get

$$\sum_{n\geq 0} f_n(A) = f(A), \ \sum_{n\geq 0} f_n(A+H) = f(A+H), \ \sum_{n\geq 0} f_n(A+H') = f(A+H')$$

with convergence of the series in the operator norm.

-

$$\|f(A+H) - f(A)\| = \Big\|\sum_{n\geq 0}[f_n(A+H) - f_n(A)]\Big\|$$

$$\leq \sum_{n\geq 0} \|f_n(A+H) - f_n(A)\| \leq \sum_{n\geq 0} 2^{n+1}\|f_n\|_{L_\infty(\mathbb{R})}\|H\| = 2\|f\|_{B^1_{\infty,1}}\|H\|.$$

Higher Order Operator Differentiability

- If $g : \mathcal{B}_{sa}(\mathbb{H}) \mapsto \mathcal{B}_{sa}(\mathbb{H})$ is a k times Fréchet differentiable function, its k-th derivative $D^k g(A)$, $A \in \mathcal{B}_{sa}(\mathbb{H})$ can be viewed as a symmetric multilinear operator valued form

$$D^k g(A)(H_1, \dots, H_k) = D^k g(A; H_1, \dots, H_k), H_1, \dots, H_k \in \mathcal{B}_{sa}(\mathbb{H}).$$

- For a k-linear form $M : \mathcal{B}_{sa}(\mathbb{H}) \times \cdots \times \mathcal{B}_{sa}(\mathbb{H}) \mapsto \mathcal{B}_{sa}(\mathbb{H})$, define its operator norm as

$$\|M\| := \sup_{\|H_1\|, \dots, \|H_k\| \leq 1} \|M(H_1, \dots, H_k)\|.$$

- The derivatives $D^k g(A)$ are defined iteratively:

$$D^k g(A)(H_1, \dots, H_{k-1}, H_k) = D(D^{k-1} g(A)(H_1, \dots, H_{k-1}))(H_k).$$

Higher Order Operator Differentiability

Lemma

Let $f \in \mathcal{E}_\sigma \bigcap L_\infty(\mathbb{R})$. *Then*

$$\|D^k f(A)\| \leq \sigma^k \|f\|_{L_\infty(\mathbb{R})}, A \in \mathcal{B}_{sa}(\mathbb{H}),$$

$$\|D^k f(A+H; H_1, \dots, H_k) - D^k f(A; H_1, \dots, H_k)\|$$
$$\leq \sigma^{k+1} \|f\|_{L_\infty(\mathbb{R})} \|H_1\| \dots \|H_k\| \|H\|$$

and

$$\|S_{D^k f(\cdot; H_1, \dots, H_k)}(A; H)\| \leq \frac{\sigma^{k+2}}{2} \|f\|_{L_\infty(\mathbb{R})} \|H_1\| \dots \|H_k\| \|H\|^2.$$

Lemma

Suppose $f \in B^k_{\infty,1}(\mathbb{R})$. Then the function $\mathcal{B}_{sa}(\mathbb{H}) \ni A \mapsto f(A) \in \mathcal{B}_{sa}(\mathbb{H})$ is k times Fréchet differentiable and

$$\|D^j f(A)\| \leq 2^j \|f\|_{B^j_{\infty,1}}, A \in \mathcal{B}_{sa}(\mathbb{H}), j = 1, \dots, k.$$

Moreover, if for some $s \in (k, k+1]$, $f \in B^s_{\infty,1}(\mathbb{R})$, then

$$\|D^k f(A+H) - D^k f(A)\| \leq 2^{k+1} \|f\|_{B^s_{\infty,1}} \|H\|^{s-k}, A, H \in \mathcal{B}_{sa}(\mathbb{H}).$$

Higher Order Operator Differentiability

For an open set $G \subset \mathcal{B}_{sa}(\mathbb{H})$, a k-times Fréchet differentiable functions $g : G \mapsto \mathcal{B}_{sa}(\mathbb{H})$ and, for $s = k + \beta, \beta \in (0, 1]$, define

$$\|g\|_{C^s(G)} := \max_{0 \le j \le k} \sup_{A \in G} \|D^j g(A)\| \bigvee \sup_{A, A+H \in G, H \neq 0} \frac{\|D^k g(A+H) - D^k g(A)\|}{\|H\|^\beta}.$$

Corollary

Suppose that, for some $s > 0$, $s \in (k, k+1]$, we have $f \in B^s_{\infty,1}(\mathbb{R})$. Then

$$\|f\|_{C^s(\mathcal{B}_{sa}(\mathbb{H}))} \le 2^{k+1} \|f\|_{B^s_{\infty,1}}.$$

Normal Approximation for Smooth Functions of Sample Covariance

- Let $\Sigma := \sum_{\lambda\in\sigma(\Sigma)} \lambda P_\lambda$ be the spectral decomposition of Σ, $\sigma(\Sigma)$ being the spectrum of Σ and P_λ being the spectral projection corresponding to the eigenvalue λ
- $f \in B^1_{\infty,1}(\mathbb{R})$
- $\|B\|_1 < \infty$
- $\sigma_f(\Sigma; B) := \sqrt{2}\|\Sigma^{1/2} Df(\Sigma; B)\Sigma^{1/2}\|_2$

Perturbation Theory: Application to Functions of Sample Covariance (Delta Method)

- $$\langle f(\hat{\Sigma}) - f(\Sigma), B\rangle = \langle Df(\Sigma; \hat{\Sigma} - \Sigma), B\rangle + \langle S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$$
- The linear term $\langle Df(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$ is of the order $O(n^{-1/2})$ and $n^{1/2}\langle Df(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$ is close in distribution to $N(0; \sigma_f^2(\Sigma; B))$.
- For $s \in (1, 2]$, $\|S_f(\Sigma; \hat{\Sigma} - \Sigma)\| \lesssim \|f\|_{B^s_{\infty,1}} \|\hat{\Sigma} - \Sigma\|^s$, implying that

$$|\langle S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle| \leq \|B\|_1 \|S_f(\Sigma; \hat{\Sigma} - \Sigma)\|$$
$$= O\Big(\Big(\frac{\mathbf{r}(\Sigma)}{n}\Big)^{s/2}\Big) = o(n^{-1/2}),$$

$$|\langle \mathbb{E}f(\hat{\Sigma}) - f(\Sigma), B\rangle| = |\langle \mathbb{E}S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle|$$
$$= O\Big(\Big(\frac{\mathbf{r}(\Sigma)}{n}\Big)^{s/2}\Big) = o(n^{-1/2})$$

provided that $\mathbf{r}(\Sigma) = o(n^{1/s-1})$.

Perturbation Theory: Application to Functions of Sample Covariance (Delta Method)

- The bounds are sharp, for instance, for $f(x) = x^2, B = u \otimes u, s = 2$:

$$\sup_{\|u\| \leq 1} |\langle \mathbb{E} f(\hat{\Sigma}) - f(\Sigma), u \otimes u \rangle| = \sup_{\|u\| \leq 1} |\langle \mathbb{E} S_f(\Sigma; \hat{\Sigma} - \Sigma), u \otimes u \rangle| =$$

$$= \frac{\|\mathrm{tr}(\Sigma)\Sigma + \Sigma^2\|}{n} \asymp \|\Sigma\|^2 \frac{\mathbf{r}(\Sigma)}{n}$$

- For $s = 2$, the Delta Method works if $\mathbf{r}(\Sigma) = o(n^{1/2})$. What if $\mathbf{r}(\Sigma) \geq n^{1/2}, \mathbf{r}(\Sigma) = o(n)$?

Normal Approximation for Smooth Functions of Sample Covariance

Let $\mathcal{G}(r;a) = \Big\{\Sigma : \|\Sigma\| \leq a, \mathbf{r}(\Sigma) \leq r\Big\}.$

Theorem

Let $f \in B^s_{\infty,1}(\mathbb{R})$ for some $s \in (1,2]$ and let B be a linear operator with $\|B\|_1 < \infty$. Suppose $a > 0, \sigma_0^2 > 0$ and

$$r_n = o(n) \text{ as } n \to \infty.$$

Then

$$\sup_{\Sigma \in \mathcal{G}(r_n;a), \sigma_f(\Sigma;B) \geq \sigma_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{n^{1/2}\Big\langle f(\hat{\Sigma}) - \mathbb{E}_\Sigma f(\hat{\Sigma}), B\Big\rangle}{\sigma_f(\Sigma, B)} \leq x \right\} - \Phi(x) \right| \to 0.$$

Normal Approximation Bounds for Smooth Functions of Sample Covariance

- Let $\Sigma := \sum_{\lambda \in \sigma(\Sigma)} \lambda P_\lambda$ be the spectral decomposition of Σ
- $f \in B^1_{\infty,1}(\mathbb{R})$
- $\|B\|_1 < \infty$
- $\sigma_f(\Sigma; B) := \sqrt{2}\|\Sigma^{1/2} Df(\Sigma; B)\Sigma^{1/2}\|_2$
- $\mu_f(\Sigma; B) := \|\Sigma^{1/2} Df(\Sigma; B)\Sigma^{1/2}\|_3$
- $\gamma_s(f; \Sigma) := \frac{\|f\|_{B^s_{\infty,1}} \|B\|_1 \|\Sigma\|^s}{\sigma_f(\Sigma; B)}$
- $t_n(\Sigma) := t_{n,s}(f; \Sigma) := \left[-\log \gamma_s(f; \Sigma) + \frac{s-1}{2} \log\left(\frac{n}{\mathbf{r}(\Sigma)}\right)\right] \bigvee 1.$

Normal Approximation Bounds for Smooth Functions of Sample Covariance

Theorem

Let $f \in B^{s}_{\infty,1}(\mathbb{R})$ *for some* $s \in (1,2]$. *Then*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{n^{1/2}\left\langle f(\hat\Sigma)-\mathbb{E}f(\hat\Sigma),B\right\rangle}{\sigma_f(\Sigma,B)}\leq x\right\}-\Phi(x)\right| \lesssim_s \left(\frac{\mu_f(\Sigma;B)}{\sigma_f(\Sigma;B)}\right)^3\frac{1}{\sqrt{n}}$$

$$+\gamma_s(f;\Sigma)\left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{(s-1)/2}\bigvee\left(\frac{t_n(\Sigma)}{n}\right)^{(s-1)/2}\bigvee\left(\frac{t_n(\Sigma)}{n}\right)^{s-1/2}\right)\sqrt{t_n(\Sigma)}.$$

Perturbation Theory for Functions of Sample Covariance

- $$\langle f(\hat{\Sigma}) - f(\Sigma), B\rangle = \langle Df(\Sigma; \hat{\Sigma} - \Sigma), B\rangle + \langle S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$$

  implies that

  $$\langle f(\hat{\Sigma}) - \mathbb{E}f(\hat{\Sigma}), B\rangle =$$
  $$= \langle Df(\Sigma)(\hat{\Sigma} - \Sigma), B\rangle + \langle S_f(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$$
  $$= \langle Df(\Sigma)(B), \hat{\Sigma} - \Sigma\rangle + \langle S_f(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_f(\Sigma; \hat{\Sigma} - \Sigma), B\rangle$$

Perturbation Theory for Functions of Sample Covariance

- The linear term

$$\langle Df(\Sigma)B, \hat{\Sigma} - \Sigma\rangle$$

$$= n^{-1}\sum_{j=1}^{n}\langle Df(\Sigma; B)X_j, X_j\rangle - \mathbb{E}\langle Df(\Sigma, B)X, X\rangle$$

is of the order $O(n^{-1/2})$ and it is approximated by a normal distribution using Berry-Esseen bound.

- The centered remainder

$$\left\langle S_f(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_f(\Sigma; \hat{\Sigma} - \Sigma), B\right\rangle$$

is of the order $o(n^{-1/2})$ when $\mathbf{r}(\Sigma) = o(n)$ and it is controlled using Gaussian concentration inequalities.

Normal Approximation for the Linear Term

- Let $A \in \mathcal{B}_{sa}(\mathbb{H})$, $\|A\|_1 < \infty$.
- Denote by $\lambda_j, j \geq 1$ the eigenvalues of $\Sigma^{1/2} A \Sigma^{1/2}$
- Then

$$\langle AX, X\rangle \stackrel{d}{=} \sum_{k\geq 1} \lambda_k Z_k^2,$$

where $Z_1, Z_2, \dots$ are i.i.d. standard normal random variables.

Normal Approximation for the Linear Term

Lemma

The following bound holds:

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{n^{1/2}\langle Df(\Sigma;B),\hat{\Sigma}-\Sigma\rangle}{\sqrt{2}\|\Sigma^{1/2}Df(\Sigma;B)\Sigma^{1/2}\|_2}\le x\right\}-\Phi(x)\right|$$

$$\lesssim\left(\frac{\|\Sigma^{1/2}Df(\Sigma;B)\Sigma^{1/2}\|_3}{\|\Sigma^{1/2}Df(\Sigma;B)\Sigma^{1/2}\|_2}\right)^3\frac{1}{\sqrt{n}}.$$

Proof.

$$\frac{n^{1/2}\langle Df(\Sigma;B),\hat{\Sigma}-\Sigma\rangle}{\sqrt{2}\|\Sigma^{1/2}Df(\Sigma)\Sigma^{1/2}\|_2}\stackrel{d}{=}\frac{\sum_{j=1}^n\sum_{k\ge 1}\lambda_k(Z_{k,j}^2-1)}{\mathrm{Var}^{1/2}\left(\sum_{j=1}^n\sum_{k\ge 1}\lambda_k(Z_{k,j}^2-1)\right)},$$

where $\{Z_{k,j}\}$ are i.i.d. standard normal random variables and λ_k the eigenvalues of $Df(\Sigma;B)$. It remains to use Berry-Esseen bound. □

Concentration of the Remainder

Theorem

Suppose that, for some $s \in (1, 2]$, $f \in B^s_{\infty,1}(\mathbb{R})$ *and also that* $\mathbf{r}(\Sigma) \lesssim n$. *Then, there exists a constant* $C = C_s > 0$ *such that, for all* $t \geq 1$, *with probability at least* $1 - e^{-t}$

$$\left|\left\langle S_f(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E} S_f(\Sigma; \hat{\Sigma} - \Sigma), B \right\rangle\right|$$

$$\leq C \|f\|_{B^s_{\infty,1}} \|B\|_1 \|\Sigma\|^s \left(\left(\frac{\mathbf{r}(\Sigma)}{n}\right)^{(s-1)/2} \bigvee \left(\frac{t}{n}\right)^{(s-1)/2} \bigvee \left(\frac{t}{n}\right)^{s-1/2} \right) \sqrt{\frac{t}{n}}.$$

Note: the centered remainder is $o_{\mathbb{P}}(n^{-1/2})$ provided that $\mathbf{r}(\Sigma) = o(n)$.

Concentration of the Remainder

- $g : \mathcal{B}_{sa}(\mathbb{H}) \mapsto \mathbb{R}$ Fréchet differentiable function with respect to the operator norm with derivative $Dg(A; H), H \in \mathcal{B}_{sa}(\mathbb{H})$.
- $S_g(A; H)$ the remainder of the first order Taylor expansion of g :

$$S_g(A; H) := g(A + H) - g(A) - Dg(A; H), A, H \in \mathcal{B}_{sa}(\mathbb{H}).$$

- Let $s \in [1, 2]$. Assume there exists a constant $L_{g,s} > 0$ such that, for all $\Sigma \in \mathcal{C}_+(\mathbb{H}), H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,

$$|S_g(\Sigma; H') - S_g(\Sigma; H)| \leq L_{g,s}(\|H\| \vee \|H'\|)^{s-1}\|H' - H\|.$$

Concentration of the Remainder

Theorem

There exists a constant $K_s > 0$ such that for all $t \geq 1$ with probability at least $1 - e^{-t}$

$$|S_g(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E} S_g(\Sigma; \hat{\Sigma} - \Sigma)|$$

$$\leq K_s L_{g,s} \|\Sigma\|^s \biggl(\Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{(s-1)/2} \bigvee \Bigl(\frac{\mathbf{r}(\Sigma)}{n}\Bigr)^{s-1/2} \bigvee \Bigl(\frac{t}{n}\Bigr)^{(s-1)/2} \bigvee \Bigl(\frac{t}{n}\Bigr)^{s-1/2} \biggr)$$

Sketch of the Proof

- $\varphi : \mathbb{R} \mapsto \mathbb{R}$, $\varphi(u) = 1, u \leq 1$, $\varphi(u) = 0, u \geq 2$, $\varphi(u) = 2 - u, u \in (1, 2)$
- $E := \hat{\Sigma} - \Sigma$
- For $\delta > 0$, define

$$h(X_1, \dots, X_n) := S_g(\Sigma; E)\varphi\left(\frac{\|E\|}{\delta}\right).$$

Sketch of the Proof

Lemma

The following bound holds with some constant $C_s > 0$ *for all* $X_1, \dots, X_n, X_1', \dots, X_n' \in \mathbb{H}$:

$$|h(X_1, \dots, X_n) - h(X_1', \dots, X_n')| \leq \frac{C_s L_{g,s}(\|\Sigma\|^{1/2} + \sqrt{\delta})\delta^{s-1}}{\sqrt{n}} \biggl(\sum_{j=1}^{n} \|X_j - X_j'\|^2\biggr)^{1/2}.$$

Sketch of the Proof

- $$
\begin{aligned}
&|h(X_1,\dots,X_n) - h(X_1',\dots,X_n')| \\
&\leq |S_g(\Sigma,E) - S_g(\Sigma,E')| + \frac{1}{\delta}|S_g(\Sigma,E')|\|E-E'\| \\
&\leq L_{g,s}(\|E\|\vee\|E'\|)^{s-1}\|E'-E\| + L_{g,s}\frac{1}{\delta}\|E'\|^s\|E'-E\|.
\end{aligned}
$$
- If both $\|E\| \leq 2\delta$ and $\|E'\| \leq 2\delta$, then

$$|h(X_1,\dots,X_n) - h(X_1',\dots,X_n')| \leq (2^{s-1}+2^s)L_{g,s}\delta^{s-1}\|E'-E\|$$

with similar bounds holding in other cases.

Sketch of the Proof

$$\|E' - E\| = \left\| n^{-1}\sum_{j=1}^{n} X_j \otimes X_j - n^{-1}\sum_{j=1}^{n} X_j' \otimes X_j' \right\|$$

$$\leq \left\| n^{-1}\sum_{j=1}^{n} (X_j - X_j') \otimes X_j \right\| + \left\| n^{-1}\sum_{j=1}^{n} X_j' \otimes (X_j - X_j') \right\|$$

$$= \sup_{\|u\|,\|v\|\leq 1} \left| n^{-1}\sum_{j=1}^{n} \langle X_j - X_j', u\rangle \langle X_j, v\rangle \right| + \sup_{\|u\|,\|v\|\leq 1} \left| n^{-1}\sum_{j=1}^{n} \langle X_j', u\rangle \langle X_j - X_j', v\rangle \right|$$

$$
\begin{aligned}
&\le \sup_{\|u\|\le 1}\left(n^{-1}\sum_{j=1}^{n}\langle X_j - X_j', u\rangle^2\right)^{1/2} \sup_{\|v\|\le 1}\left(n^{-1}\sum_{j=1}^{n}\langle X_j, v\rangle^2\right)^{1/2} \\
&+ \sup_{\|u\|\le 1}\left(n^{-1}\sum_{j=1}^{n}\langle X_j', u\rangle^2\right)^{1/2} \sup_{\|v\|\le 1}\left(n^{-1}\sum_{j=1}^{n}\langle X_j - X_j', v\rangle^2\right)^{1/2} \\
&\le \frac{\|\hat{\Sigma}\|^{1/2} + \|\hat{\Sigma}'\|^{1/2}}{\sqrt{n}}\left(\sum_{j=1}^{n}\|X_j - X_j'\|^2\right)^{1/2} \\
&\le (2\|\Sigma\|^{1/2} + \|E\|^{1/2} + \|E'\|^{1/2})\Delta,
\end{aligned}
$$

where $\Delta := \frac{1}{\sqrt{n}}\left(\sum_{j=1}^{n}\|X_j - X_j'\|^2\right)^{1/2}.$

By a simple further algebra,

$$\|E' - E\| \wedge \delta \leq 4\|\Sigma\|^{1/2}\Delta \bigvee (4\sqrt{2} + 2)\sqrt{\delta}\Delta,$$

which together with the bound

$$|h(X_1, \ldots, X_n) - h(X_1', \ldots, X_n')| \lesssim_s L_{g,s}\delta^{s-1}\|E' - E\|$$

implies the claim of the lemma.

Sketch of the Proof

- For a given $t > 0$, let

$$\delta = \delta_n(t) := \mathbb{E}\|\hat{\Sigma} - \Sigma\| + C\|\Sigma\|\Big[\Big(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee 1\Big)\sqrt{\frac{t}{n}} \bigvee \frac{t}{n}\Big]$$

  with constant $C > 0$ such that $\mathbb{P}\{\|\hat{\Sigma} - \Sigma\| \geq \delta_n(t)\} \leq e^{-t}$.
- Let $M := \mathrm{Med}(S_g(\Sigma; \hat{\Sigma} - \Sigma))$
- By Gaussian concentration inequality, probability at least $1 - e^{-t}$

$$|h(X_1, \dots, X_n) - M| \lesssim_s L_{g,s}\delta^{s-1}(\|\Sigma\|^{1/2} + \delta^{1/2})\|\Sigma\|^{1/2}\sqrt{\frac{t}{n}}.$$

- On the event $\{\|\hat{\Sigma} - \Sigma\| \leq \delta\}$, replace $h(X_1, \dots, X_n)$ by $S_g(\Sigma; \hat{\Sigma} - \Sigma)$

Assumptions on the Loss Function

Let $\mathcal{L}$ be the class of functions $\ell : \mathbb{R} \mapsto \mathbb{R}_+$ such that

- $\ell(0) = 0$
- $\ell(u) = \ell(-u), u \in \mathbb{R}$
- ℓ is nondecreasing and convex on $\mathbb{R}_+$
- For some constants $c_1, c_2 > 0$

$$\ell(u) \leq c_1 e^{c_2 u}, u \geq 0.$$

When is Plug-In Estimator Asymptotically Efficient?

Theorem

Suppose, for some $s \in (1,2]$, $f \in B^s_{\infty,1}(\mathbb{R})$. *Let* B *be a nuclear operator and let* $a > 0, \sigma_0 > 0$. *Suppose that* $r_n > 1$ *and* $r_n = o(n^{1-\frac{1}{s}})$ *as* $n \to \infty$. *Then*

$$\sup_{\Sigma \in \mathcal{G}(r_n;a), \sigma_f(\Sigma;B) \geq \sigma_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{n^{1/2}(\langle f(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle)}{\sigma_f(\Sigma; B)} \leq x \right\} - \Phi(x) \right| \to 0$$

as $n \to \infty$. *Moreover, under the same assumptions on* f *and* r_n, *and for any loss function* $\ell \in \mathcal{L}$

$$\sup_{\Sigma \in \mathcal{G}(r_n;a), \sigma_f(\Sigma;B) \geq \sigma_0} \left| \mathbb{E}_\Sigma \ell \left(\frac{n^{1/2}\Big(\langle f(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle\Big)}{\sigma_f(\Sigma; B)} \right) - \mathbb{E}\ell(Z) \right| \to 0$$

as $n \to \infty$, *where* Z *is a standard normal random variable.*

Efficient Estimation of $\langle f(\Sigma), B\rangle$: A Lower Bound

Theorem

Suppose $f \in B^1_{\infty,1}(\mathbb{R})$. *Suppose* $r_n > 1, a > 1, \sigma_0 > 0$ *are such that, for some* $1 < a' < a$ *and* $\sigma_0' > \sigma_0$ *and for all large enough* n,

$$\mathcal{G}(r_n, a') \bigcap \Big\{\Sigma : \sigma_f(\Sigma; B) \geq \sigma_0'\Big\} \neq \emptyset.$$

Then the following bound holds:

$$\liminf_n \inf_{T_n} \sup_{\Sigma \in \mathcal{G}(r_n;a), \sigma_f(\Sigma;B) \geq \sigma_0} \frac{n\mathbb{E}_\Sigma \Big(T_n(X_1, \dots, X_n) - \langle f(\Sigma), B\rangle\Big)^2}{\sigma_f^2(\Sigma; B)} \geq 1,$$

where the infimum is taken over all sequences of estimators $T_n = T_n(X_1, \dots, X_n)$.

Part 3.

Wishart Operators, Bootstrap Chains, Invariant Functions and Bias Reduction

Wishart Operators and Bias Reduction

- Our next goal is to find an estimator $g(\hat{\Sigma})$ of $f(\Sigma)$ with a small bias $\mathbb{E}_{\Sigma} g(\hat{\Sigma}) - f(\Sigma)$ (of the order $o(n^{-1/2})$) and such that

$$n^{1/2}(\langle g(\hat{\Sigma}), B\rangle - \langle \mathbb{E}_{\Sigma} g(\hat{\Sigma}), B\rangle)$$

  is asymptotically normal.
- To this end, one has to find a sufficiently smooth approximate solution of the equation

$$\mathbb{E}_{\Sigma} g(\hat{\Sigma}) = f(\Sigma), \Sigma \in \mathcal{C}_+^d.$$

Wishart Operators

- $\mathcal{T}g(\Sigma) := \mathbb{E}_\Sigma g(\hat{\Sigma}) = \int_{\mathcal{C}_+^d} g(V)P(\Sigma; dV), \Sigma \in \mathcal{C}_+^d$, where Markov kernel $P(\Sigma;\cdot)$ is a rescaled Wishart distribution $\mathcal{W}_d(\Sigma; n)$:

$$P(\Sigma; A) := \mathbb{P}_\Sigma\{\hat{\Sigma} \in A\}, A \subset \mathcal{C}_+^d.$$

- For $d \leq n$, $P(\Sigma; dV) = np(\Sigma; nV)dV$,

$$p(\Sigma; V) :=$$

$$\frac{1}{2^{nd/2}(\mathrm{det}(\Sigma))^{n/2}\Gamma_d\left(\frac{n}{2}\right)}(\mathrm{det}(V))^{(n-d-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}V)\right\},$$

where Γ_d is the multivariate gamma function:

$$\Gamma_d\left(\frac{n}{2}\right) := \pi^{d(d-1)/4}\prod_{j=1}^{d}\Gamma\left(\frac{n}{2} - \frac{j-1}{2}\right).$$

Bias Reduction

- To find an estimator of $f(\Sigma)$ with a small bias, one needs to solve (approximately) the following integral equation
- the Wishart equation:

$$\mathcal{T}g(\Sigma) = f(\Sigma), \Sigma \in \mathcal{C}_+^d.$$

- Bias operator: $\mathcal{B} := \mathcal{T} - \mathcal{I}.$
- Formally, the solution of the Wishart equation is given by Neumann series:

$$g(\Sigma) = (\mathcal{I} + \mathcal{B})^{-1} f(\Sigma) = (\mathcal{I} - \mathcal{B} + \mathcal{B}^2 - \dots) f(\Sigma)$$

- Given a smooth function $f : \mathbb{R} \mapsto \mathbb{R}$, define

$$f_k(\Sigma) := \sum_{j=0}^{k} (-1)^j \mathcal{B}^j f(\Sigma) := f(\Sigma) + \sum_{j=1}^{k} (-1)^j \mathcal{B}^j f(\Sigma)$$

Bias Reduction

Proposition

The bias of estimator $f_k(\hat{\Sigma})$ of $f(\Sigma)$ is given by the following formula:

$$\mathbb{E}_\Sigma f_k(\hat{\Sigma}) - f(\Sigma) = (-1)^k \mathcal{B}^{k+1} f(\Sigma).$$

Proof.

$$\mathbb{E}_\Sigma f_k(\hat{\Sigma}) - g(\Sigma) = \mathcal{T} f_k(\Sigma) - f(\Sigma) = (\mathcal{I} + \mathcal{B}) f_k(\Sigma) - f(\Sigma)$$

$$= \sum_{j=0}^{k} (-1)^j \mathcal{B}^j f(\Sigma) - \sum_{j=1}^{k+1} (-1)^j \mathcal{B}^j f(\Sigma) - f(\Sigma) = (-1)^k \mathcal{B}^{k+1} f(\Sigma).$$

□

Bootstrap Chain

- $\mathcal{T}g(\Sigma) := \mathbb{E}_\Sigma g(\hat{\Sigma}) = \int_{\mathcal{C}_+^d} g(V)P(\Sigma; dV), \Sigma \in \mathcal{C}_+^d$
- $\mathcal{T}^k g(\Sigma) = \mathbb{E}_\Sigma g(\hat{\Sigma}^{(k)}), \Sigma \in \mathcal{C}_+^d$, where

$$\hat{\Sigma}^{(0)} = \Sigma \to \hat{\Sigma}^{(1)} = \hat{\Sigma} \to \hat{\Sigma}^{(2)} \to \dots$$

  is a Markov chain in $\mathcal{C}_+^d$ with transition probability kernel $P(\cdot;\cdot)$.
- Note that $\hat{\Sigma}^{(j+1)}$ is the sample covariance based on n i.i.d. observations $\sim N(0; \hat{\Sigma}^{(j)})$ (conditionally on $\hat{\Sigma}^{(j)}$)
- Conditionally on $\hat{\Sigma}^{(j)}$, with a “high probability",

$$\|\hat{\Sigma}^{(j+1)} - \hat{\Sigma}^{(j)}\| \lesssim \|\hat{\Sigma}^{(j)}\|\sqrt{\frac{d}{n}}$$

- The Markov Chain $\{\hat{\Sigma}^{(j)}, j = 0, 1, \dots\}$ will be called *the Bootstrap Chain*.

k-th order difference

- k-th order difference along the Markov chain: by Newton binomial formula,

$$\mathcal{B}^k f(\Sigma) = (\mathcal{T} - \mathcal{I})^k f(\Sigma) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \mathcal{T}^j f(\Sigma)$$

$$= \mathbb{E}_\Sigma \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(\hat{\Sigma}^{(j)})$$

- $\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(\hat{\Sigma}^{(j)})$ is the k-th order difference of f along the trajectory of the Bootstrap Chain.
- Question. Suppose f is of smoothness k. Is $\mathcal{B}^k f(\Sigma)$ of the order $\left(\sqrt{\frac{d}{n}}\right)^k$?

Orthogonally Invariant Functions

- A function $g \in L_\infty(\mathcal{C}_+^d)$ is called *orthogonally invariant* iff, for all orthogonal transformations U of $\mathbb{R}^d$,

$$g(U\Sigma U^{-1}) = g(\Sigma), \Sigma \in \mathcal{C}_+^d.$$

- Any orthogonally invariant function g could be represented as $g(\Sigma) = \varphi(\lambda_1(\Sigma), \dots, \lambda_d(\Sigma))$, where $\lambda_1(\Sigma) \geq \dots \lambda_d(\Sigma)$ are the eigenvalues of Σ and φ is a symmetric function of d variables.
- A typical example: $g(\Sigma) = \mathrm{tr}(\psi(\Sigma))$ for a function of real variable ψ.
- Let $L_\infty^O(\mathcal{C}_+^d)$ be the space of all orthogonally invariant functions from $L_\infty(\mathcal{C}_+^d)$.

Orthogonally Invariant Functions

Proposition

If $g \in L_\infty^O(\mathcal{C}_+^d)$, then $\mathcal{T}g \in L_\infty^O(\mathcal{C}_+^d)$ and $\mathcal{B}g \in L_\infty^O(\mathcal{C}_+^d)$.

Proof.

Indeed, the transformation $\Sigma \mapsto U\Sigma U^{-1}$ is a bijection of $\mathcal{C}_+^d$,

$$\mathcal{T}g(U\Sigma U^{-1}) = \mathbb{E}_{U\Sigma U^{-1}} g(\hat{\Sigma}) = \mathbb{E}_\Sigma g(U\hat{\Sigma}U^{-1}) = \mathbb{E}_\Sigma g(\hat{\Sigma}) = \mathcal{T}g(\Sigma)$$

and the function Tg is uniformly bounded. □ □

Orthogonally Equivariant Functions

$g : \mathcal{C}_+^d \mapsto \mathcal{B}_{sa}(\mathbb{R}^d)$ is called *orthogonally equivariant* iff for all orthogonal transformations U

$$g(U\Sigma U^{-1}) = Ug(\Sigma)U^{-1}, \Sigma \in \mathcal{C}_+^d.$$

$g : \mathcal{C}_+^d \mapsto \mathcal{B}_{sa}(\mathbb{R}^d)$ is continuously differentiable in $\mathcal{C}_+^d$ iff there exists a uniformly bounded, Lipschitz with respect to the operator norm and continuously differentiable extension of g to an open set G,

$$\mathcal{C}_+^d \subset G \subset \mathcal{B}_{sa}(\mathbb{R}^d).$$

Proposition

If $g : \mathcal{C}_+^d \mapsto \mathbb{R}$ is orthogonally invariant and continuously differentiable in $\mathcal{C}_+^d$ with derivative Dg, then Dg is orthogonally equivariant.

Orthogonally Equivariant Functions

Proof.

First suppose that Σ is positively definite (then extend to $\mathcal{C}_+^d$ by continuity). For all $H \in \mathcal{B}_{sa}(\mathbb{R}^d)$,

$$\langle Dg(U\Sigma U^{-1}), H\rangle = \lim_{t\to 0} \frac{g(U\Sigma U^{-1} + tH) - g(U\Sigma U^{-1})}{t}$$

$$= \lim_{t\to 0} \frac{g(U(\Sigma + tU^{-1}HU)U^{-1}) - g(U\Sigma U^{-1})}{t}$$

$$= \lim_{t\to 0} \frac{g(\Sigma + tU^{-1}HU)) - g(\Sigma)}{t}$$

$$= \langle Dg(\Sigma), U^{-1}HU\rangle = \langle UDg(\Sigma)U^{-1}, H\rangle.$$

□

“Lifting" Operator $\mathcal{D}$ and Reduction to Orthogonally Invariant Functions

- Define the following differential operator (“lifting" operator):

$$\mathcal{D}g(\Sigma) := \Sigma^{1/2} Dg(\Sigma) \Sigma^{1/2}$$

  acting on continuously differentiable functions in $\mathcal{C}_+^d$.
- Suppose $f(x) = x\psi'(x)$
- Let $g(\Sigma) := \mathrm{tr}(\psi(\Sigma))$
- g is orthogonally invariant function on $\mathcal{C}_+^d$
- $Dg(\Sigma) = \psi'(\Sigma)$
- $\mathcal{D}g(\Sigma) = f(\Sigma)$

“Lifting" Operator $\mathcal{D}$ and Reduction to Invariant Functions

- Suppose, for some $\delta > 0$, $\sigma(\Sigma) \subset [2\delta, \infty)$.
- Let $\gamma_\delta(x) = \gamma(x/\delta)$, $\gamma : \mathbb{R} \mapsto [0,1]$ be a nondecreasing C^∞ function, $\gamma(x) = 0, x \leq 1/2$, $\gamma(x) = 1, x \geq 1$.
- Define $f_\delta(x) = f(x)\gamma_\delta(x), x \in \mathbb{R}$.
- Then, $f(\Sigma) = f_\delta(\Sigma)$ and, for all Σ with $\sigma(\Sigma) \subset [2\delta, \infty)$, $Df(\Sigma) = Df_\delta(\Sigma)$.
- Let $\varphi(x) := \int_0^x \frac{f_\delta(t)}{t} dt, x \geq 0$, $\varphi(x) = 0, x < 0$.
- Clearly, $f_\delta(x) = x\varphi'(x), x \in \mathbb{R}$.
- Let $g(C) := \mathrm{tr}(\varphi(C)), C \in \mathcal{B}_{sa}(\mathbb{R}^d)$. Then

$$\mathcal{D}g(C) = C^{1/2}\varphi'(C)C^{1/2} = f_\delta(C), C \in \mathcal{C}_+^d.$$

- Moreover,

$$\|Dg\|_{C^s} \lesssim 2^{k+1}(\delta^{-1-s} \vee \delta^{-1})\|f\|_{B^s_{\infty,1}}.$$

Operator $\mathcal{D}$ and its Commutativity Properties

Proposition

Suppose $d \lesssim n$. For all functions $g \in L_\infty^O(\mathcal{C}_+^d)$ that are continuously differentiable in $\mathcal{C}_+^d$ with a uniformly bounded derivative Dg and for all $\Sigma \in \mathcal{C}_+^d$

$$\mathcal{D}\mathcal{T}g(\Sigma) = \mathcal{T}\mathcal{D}g(\Sigma) \text{ and } \mathcal{D}\mathcal{B}g(\Sigma) = \mathcal{B}\mathcal{D}g(\Sigma).$$

- Note that $\hat{\Sigma} \overset{d}{=} \Sigma^{1/2} W \Sigma^{1/2}$, where W is the sample covariance based on i.i.d. standard normal random variables $Z_1, \dots, Z_n$ in $\mathbb{R}^d$.
- Let $\Sigma^{1/2} W^{1/2} = RU$ be the polar decomposition of $\Sigma^{1/2} W^{1/2}$ with positively semidefinite R and orthogonal U.
- Then,

$$\hat{\Sigma} = \Sigma^{1/2} W \Sigma^{1/2} = \Sigma^{1/2} W^{1/2} W^{1/2} \Sigma^{1/2} = RUU^{-1} R = R^2$$

and

$$W^{1/2} \Sigma W^{1/2} = W^{1/2} \Sigma^{1/2} \Sigma^{1/2} W^{1/2}$$
$$= U^{-1} RRU = U^{-1} R^2 U = U^{-1} \Sigma^{1/2} W \Sigma^{1/2} U = U^{-1} \hat{\Sigma} U.$$

Operator $\mathcal{D}$ and its Commutativity Properties (Proof)

- Since g is orthogonally invariant, we have

$$\mathcal{T}g(\Sigma) = \mathbb{E}_{\Sigma} g(\hat{\Sigma}) = \mathbb{E}g(\Sigma^{1/2} W \Sigma^{1/2}) = \mathbb{E}g(W^{1/2}\Sigma W^{1/2}), \Sigma \in \mathcal{C}_+^d.$$

- For simplicity, aassume that Σ is positively definite.
- Let $H \in \mathcal{B}_{sa}(\mathbb{R}^d)$ and $\Sigma_t := \Sigma + tH, t > 0$. Note that

$$\begin{aligned}
&D\mathcal{T}g(\Sigma)\\
&= \lim_{t\to 0} \frac{\mathcal{T}g(\Sigma_t) - \mathcal{T}g(\Sigma)}{t}\\
&= \lim_{t\to 0} \mathbb{E}\frac{g(W^{1/2}\Sigma_t W^{1/2}) - g(W^{1/2}\Sigma W^{1/2})}{t}\\
&= \mathbb{E}\langle W^{1/2} Dg(W^{1/2}\Sigma W^{1/2}) W^{1/2}, H\rangle\\
&= \langle \mathbb{E} W^{1/2} Dg(W^{1/2}\Sigma W^{1/2}) W^{1/2}, H\rangle.
\end{aligned}$$

Operator $\mathcal{D}$ and its Commutativity Properties (Proof)

- It follows that

$$DTg(\Sigma) = \mathbb{E} W^{1/2} Dg(W^{1/2}\Sigma W^{1/2}) W^{1/2}.$$

- Since $W^{1/2}\Sigma W^{1/2} = U^{-1}\hat{\Sigma} U$ and Dg is an orthogonally equivariant function, we get

$$Dg(W^{1/2}\Sigma W^{1/2}) = U^{-1} Dg(\hat{\Sigma}) U.$$

- Therefore,

$$\begin{aligned}
&\mathcal{D}\mathcal{T}g(\Sigma) = \Sigma^{1/2} D\mathcal{T}g(\Sigma)\Sigma^{1/2} \\
&= \Sigma^{1/2}\mathbb{E}(W^{1/2} Dg(W^{1/2}\Sigma W^{1/2}) W^{1/2})\Sigma^{1/2} \\
&= \mathbb{E}(\Sigma^{1/2} W^{1/2} Dg(W^{1/2}\Sigma W^{1/2}) W^{1/2}\Sigma^{1/2}) \\
&= \mathbb{E}(\Sigma^{1/2} W^{1/2} U^{-1} Dg(\hat{\Sigma}) U W^{1/2}\Sigma^{1/2}) \\
&= \mathbb{E}(RUU^{-1} Dg(\hat{\Sigma}) UU^{-1} R) = \mathbb{E}(R Dg(\hat{\Sigma}) R) = \mathbb{E}_{\Sigma}(\hat{\Sigma}^{1/2} Dg(\hat{\Sigma})\hat{\Sigma}^{1/2}) \\
&= \mathbb{E}_{\Sigma}\mathcal{D}g(\hat{\Sigma}) = \mathcal{T}\mathcal{D}g(\Sigma). \quad \square
\end{aligned}$$

Properties of Operators $\mathcal{T}^k$ and $\mathcal{B}^k$

Let $W_1, \dots, W_k, \dots$ be i.i.d. copies of W.

Proposition

Suppose $d \lesssim n$. Then, for all $g \in L_\infty^O(\mathcal{C}_+^d)$ and for all $k \geq 1$,

$$\mathcal{T}^k g(\Sigma) = \mathbb{E} g(W_k^{1/2} \dots W_1^{1/2} \Sigma W_1^{1/2} \dots W_k^{1/2})$$

and

$$\mathcal{B}^k g(\Sigma) = \mathbb{E} \sum_{I \subset \{1,\dots,k\}} (-1)^{k-|I|} g(A_I^* \Sigma A_I),$$

where $A_I := \prod_{i \in I} W_i^{1/2}$.

Properties of Operators $\mathcal{T}^k$ and $\mathcal{B}^k$ (Proof)

- Since $\hat\Sigma \overset{d}{=} \Sigma^{1/2} W \Sigma^{1/2}$, $W^{1/2}\Sigma W^{1/2} = U^{-1}\Sigma^{1/2} W \Sigma^{1/2} U$, where U is an orthogonal operator, and g is orthogonally invariant, we have

$$\mathcal{T} g(\Sigma) = \mathbb{E}_\Sigma g(\hat\Sigma) = \mathbb{E} g(W^{1/2}\Sigma W^{1/2}).$$

- Orthogonal invariance of g implies the same property of $\mathcal{T} g$ and, by induction, of $\mathcal{T}^k g$ for all $k \geq 1$. Then, also by induction,

$$\mathcal{T}^k g(\Sigma) = \mathbb{E} g(W_k^{1/2} \dots W_1^{1/2} \Sigma W_1^{1/2} \dots W_k^{1/2}).$$

- If $I \subset \{1, \dots, k\}$ with $|I| = \mathrm{card}(I) = j$ and $A_I = \prod_{i\in I} W_i^{1/2}$, it implies that

$$\mathcal{T}^j g(\Sigma) = \mathbb{E} g(A_I^* \Sigma A_I).$$

Properties of Operators $\mathcal{T}^k$ and $\mathcal{B}^k$ (Proof)

- Recall that

$$\mathcal{B}^k g(\Sigma) = (\mathcal{T} - \mathcal{I})^k g(\Sigma) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \mathcal{T}^j g(\Sigma)$$

- Therefore,

$$\mathcal{B}^k g(\Sigma) = \mathbb{E} \sum_{I \subset \{1,\dots,k\}} (-1)^{k-|I|} g(A_I^* \Sigma A_I).$$

□

Properties of Operators $\mathcal{T}^k$ and $\mathcal{B}^k$

Proposition

Suppose that $d \lesssim n$ and that g is continuously differentiable in $\mathcal{C}_+^d$ with a uniformly bounded derivative Dg. Then

1. $D\mathcal{B}^k g(\Sigma) = \mathbb{E}\sum_{I\subset\{1,\dots,k\}}(-1)^{k-|I|}A_I Dg(A_I^*\Sigma A_I)A_I^*.$
2. For all $\Sigma \in \mathcal{C}_+^d$,

$$\mathcal{D}\mathcal{T}^k g(\Sigma) = \mathcal{T}^k\mathcal{D}g(\Sigma) \text{ and } \mathcal{D}\mathcal{B}^k g(\Sigma) = \mathcal{B}^k\mathcal{D}g(\Sigma).$$

3.

$$\begin{aligned}\mathcal{B}^k\mathcal{D}g(\Sigma) &= \mathcal{D}\mathcal{B}^k g(\Sigma)\\ &= \mathbb{E}\Big(\sum_{I\subset\{1,\dots,k\}}(-1)^{k-|I|}\Sigma^{1/2}A_I Dg(A_I^*\Sigma A_I)A_I^*\Sigma^{1/2}\Big).\end{aligned}$$

An Integral Representation of Operator $\mathcal{B}^k\mathcal{D}g(\Sigma)$

- Linear interpolation between I and $W_1^{1/2}, \dots, W_k^{1/2}$:

$$V_j(t_j) := I + t_j(W_j^{1/2} - I), t_j \in [0,1], 1 \le j \le k.$$

- For all $j = 1, \dots, k, t_j \in [0,1]$, $V_j(t_j) \in \mathcal{C}_+^d$.
-

$$R := R(t_1, \dots, t_k) = V_1(t_1) \dots V_k(t_k),$$
$$L := L(t_1, \dots, t_k) = V_k(t_k) \dots V_1(t_1) = R^*,$$
$$S := S(t_1, \dots, t_k) = L(t_1, \dots, t_k)\Sigma R(t_1, \dots, t_k), (t_1, \dots, t_k) \in [0,1]^k.$$

- Let

$$\varphi(t_1, \dots, t_k) := \Sigma^{1/2} R(t_1, \dots, t_k) Dg(S(t_1, \dots, t_k)) L(t_1, \dots, t_k) \Sigma^{1/2}.$$

- Let $(t_1, \dots, t_k) \in \{0,1\}, I := \{j : 1 \le j \le k, t_j = 1\}$. Then

$$\varphi(t_1, \dots, t_k) = \Sigma^{1/2} A_I Dg(A_I^* \Sigma A_I) A_I^* \Sigma^{1/2}.$$

An Integral Representation of Operator $\mathcal{B}^k\mathcal{D}g(\Sigma)$

Proposition

Suppose $g \in L_{\infty}^{O}(\mathcal{C}_+^d)$ is $k+1$ times continuously differentiable function with uniformly bounded derivatives $D^j g, j = 1, \dots, k+1$. Then the function φ is k times continuously differentiable in $[0,1]^k$ and

$$\mathcal{B}^k\mathcal{D}g(\Sigma) = \mathbb{E}\int_0^1 \dots \int_0^1 \frac{\partial^k \varphi(t_1, \dots, t_k)}{\partial t_1 \dots \partial t_k} dt_1 \dots dt_k, \Sigma \in \mathcal{C}_+^d.$$

An Integral Representation of Operator $\mathcal{B}^k\mathcal{D}g(\Sigma)$ (Proof)

- Given a function $\phi : [0,1]^k \mapsto \mathbb{R}$, define for $1 \le i \le k$ finite difference operators

$$\Delta_i\phi(t_1,\dots,t_k)$$
$$:= \phi(t_1,\dots,t_{i-1},1,t_{i+1},\dots,t_k) - \phi(t_1,\dots,t_{i-1},0,t_{i+1},\dots,t_k)$$

- Then

$$\Delta_1\dots\Delta_k\phi = \sum_{(t_1,\dots,t_k)\in\{0,1\}^k} (-1)^{k-(t_1+\dots+t_k)}\phi(t_1,\dots,t_k).$$

- It is well known that if ϕ is k times continuously differentiable in $[0,1]^k$, then

$$\Delta_1\dots\Delta_k\phi = \int_0^1\dots\int_0^1 \frac{\partial^k\phi(t_1,\dots,t_k)}{\partial t_1\dots\partial t_k}dt_1\dots dt_k.$$

An Integral Representation of Operator $\mathcal{B}^k \mathcal{D} g(\Sigma)$ (Proof)

- $$\mathcal{B}^k \mathcal{D} g(\Sigma) = \mathcal{D}\mathcal{B}^k g(\Sigma)$$
$$= \mathbb{E}\Big(\sum_{I \subset \{1,\dots,k\}} (-1)^{k-|I|} \Sigma^{1/2} A_I Dg(A_I^* \Sigma A_I) A_I^* \Sigma^{1/2} \Big)$$
$$= \sum_{(t_1,\dots,t_k)\in\{0,1\}^k} (-1)^{k-(t_1+\cdots+t_k)} \varphi(t_1,\dots,t_k)$$
$$= \mathbb{E}\Delta_1 \dots \Delta_k \varphi.$$
- Since Dg is k times continuously differentiable and the functions $S(t_1,\dots,t_k)$, $R(t_1,\dots,t_k)$ are polynomials with respect to $t_1,\dots,t_k$, the function φ is k times continuously differentiable in $[0,1]^k$.
- This implies

$$\mathcal{B}^k \mathcal{D} g(\Sigma) = \mathbb{E} \int_0^1 \dots \int_0^1 \frac{\partial^k \varphi(t_1,\dots,t_k)}{\partial t_1 \dots \partial t_k} dt_1 \dots dt_k, \Sigma \in \mathcal{C}_+^d.$$

A bound on $\mathcal{B}^k\mathcal{D}g(\Sigma)$

Theorem

Suppose that $k \le d \le n$ *and that* $g \in L_\infty^O(\mathcal{C}_+^d)$ *is* $k+1$ *times continuously differentiable function with uniformly bounded derivatives* $D^j g, j = 1, \dots, k+1.$ *Then, for some* $C > 1$,

$$\|\mathcal{B}^k\mathcal{D}g(\Sigma)\| \le C^{k^2} \max_{1\le j\le k+1} \|D^j g\|_{L_\infty}(\|\Sigma\|^{k+1} \vee \|\Sigma\|)\left(\frac{d}{n}\right)^{k/2}.$$

Bounds on the bias of $\mathcal{D}g_k(\hat{\Sigma})$

Corollary

Suppose that $k+1 \leq d \leq n$ *and that* $g \in L_\infty^O(\mathcal{C}_+^d)$ *is* $k+2$ *times continuously differentiable function with uniformly bounded derivatives* $D^j g, j = 1, \dots, k+2$. *Then, for some* $C > 1$,

$$\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\|$$
$$\leq C^{(k+1)^2} \max_{1\leq j\leq k+2} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+2} \vee \|\Sigma\|) \Big(\frac{d}{n}\Big)^{(k+1)/2}.$$

If, for some $\alpha \in (1/2, 1)$, $2\log n \leq d \leq n^\alpha$ and $k > \frac{\alpha}{1-\alpha}$, then

$$\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\| = o(n^{-1/2}).$$

Further bound on the bias of $\mathcal{D}g_k(\hat{\Sigma})$

Theorem

Suppose $g \in L_\infty^O(\mathcal{C}_+^d)$ *is* $k+2$ *times continuously differentiable for some* $k \le d \le n$ *and, for some* $\beta \in (0,1]$, $\|Dg\|_{C^{k+1+\beta}} < \infty$. *In addition, suppose that for some* $\delta > 0$ $\sigma(\Sigma) \subset \left[\delta, \frac{1}{\delta}\right]$. *Then, for some constant* $C > 0$,

$$\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\|$$
$$\le C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} (\|\Sigma\| \vee 1)^{k+3/2} \|\Sigma\| \left(\frac{d}{n}\right)^{(k+1+\beta)/2}.$$

Sketch of the proof: bounding partial derivatives

- To compute $\frac{\partial^k \varphi}{\partial t_1 \dots \partial t_k}$, we derive formulas for partial derivatives of operator-valued function $h(S(t_1, \dots, t_k))$, $h = Dg$. Recall that

$$R := R(t_1, \dots, t_k) = V_1(t_1) \dots V_k(t_k),$$
$$L := L(t_1, \dots, t_k) = V_k(t_k) \dots V_1(t_1) = R^*,$$
$$S := S(t_1, \dots, t_k) = L(t_1, \dots, t_k) \Sigma R(t_1, \dots, t_k).$$

- Given $T = \{t_{i_1}, \dots, t_{i_m}\} \subset \{t_1, \dots, t_k\}$, let $\partial_T S := \frac{\partial^m S(t_1, \dots, t_k)}{\partial t_{i_1} \dots \partial t_{i_m}}$ (similar notations are used for partial derivatives of $h(S)$, etc.).
- Let $\mathcal{D}_{j,T}$ be the set of all partitions $(\Delta_1, \dots, \Delta_j)$ of $T \subset \{t_1, \dots, t_k\}$ with non-empty sets Δ_i, $i = 1, \dots, j$ (partitions with different order of $\Delta_1, \dots, \Delta_j$ being identical).
- For $\Delta = (\Delta_1, \dots, \Delta_j) \in \mathcal{D}_{j,T}$, set $\partial_\Delta S = (\partial_{\Delta_1} S, \dots, \partial_{\Delta_j} S)$.
- Denote $\mathcal{D}_T := \bigcup_{j=1}^{|T|} \mathcal{D}_{j,T}$.
- For $\Delta = (\Delta_1, \dots, \Delta_j) \in \mathcal{D}_T$, set $j_\Delta := j$.

Sketch of the proof: bounding partial derivatives

Lemma

Suppose, for some $m \leq k$, $h = Dg \in L_\infty(\mathcal{C}_+^d; \mathcal{B}_{sa}(\mathbb{R}^d))$ *is m times continuously differentiable with derivatives* $D^j h, j \leq m$. *Then the function* $[0,1]^k \ni (t_1, \dots, t_k) \mapsto h(S(t_1, \dots, t_k))$ *is m times continuously differentiable and for any* $T \subset \{t_1, \dots, t_k\}$ *with* $|T| = m$

$$\partial_T h(S) = \sum_{\Delta \in \mathcal{D}_T} D^{j_\Delta} h(S)(\partial_\Delta S) = \sum_{j=1}^{m} \sum_{\Delta \in \mathcal{D}_{j,T}} D^j h(S)(\partial_\Delta S).$$

Sketch of the proof: bounding partial derivatives

Denote

$$\delta_i := \|W_i - I\|, i = 1, \ldots, k.$$

Lemma

For all $T \subset \{t_1, \ldots, t_k\}$,

$$\|\partial_T R\| \leq \prod_{t_i \in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^{k} (1+\delta_i),$$

$$\|\partial_T L\| \leq \prod_{t_i \in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^{k} (1+\delta_i),$$

$$\|\partial_T S\| \leq 2^k \|\Sigma\| \prod_{t_i \in T} \frac{\delta_i}{1+\delta_i} \prod_{i=1}^{k} (1+\delta_i)^2.$$

Sketch of the proof: bounding partial derivatives

Lemma

Suppose that, for some $0 \leq m \leq k$, $h = Dg \in L_\infty(\mathcal{C}_+^d; \mathcal{B}_{sa}(\mathbb{R}^d))$ *is* m *times differentiable with uniformly bounded continuous derivatives* $D^j h, j = 1, \dots, m$. *Then for all* $T \subset \{t_1, \dots, t_k\}$ *with* $|T| = m$

$$\|\partial_T h(S)\| \leq 2^{m(k+m+1)} \max_{0 \leq j \leq m} \|D^j h\|_{L_\infty} (\|\Sigma\|^m \vee 1) \prod_{i=1}^{k} (1 + \delta_i)^{2m} \prod_{t_i \in T} \frac{\delta_i}{1 + \delta_i}.$$

Sketch of the proof: bounding partial derivatives

Lemma

$$\left\| \frac{\partial^k \varphi(t_1, \dots, t_k)}{\partial t_1 \dots \partial t_k} \right\|$$

$$\leq 3^k 2^{k(2k+1)} \max_{1 \leq j \leq k+1} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+1} \vee \|\Sigma\|) \prod_{i=1}^{k} \delta_i (1+\delta_i)^{2k+1},$$

where $\delta_i := \|W_i - I\|$.

Sketch of the proof: bound on $\|\mathcal{B}^k \mathcal{D} g(\Sigma)\|$

$$
\begin{aligned}
&\|\mathcal{B}^k \mathcal{D} g(\Sigma)\| = \|\mathcal{D}\mathcal{B}^k g(\Sigma)\| \\
&\leq \mathbb{E}\int_0^1 \dots \int_0^1 \left\|\frac{\partial^k \varphi(t_1,\dots,t_k)}{\partial t_1 \dots \partial t_k}\right\| dt_1 \dots dt_k \\
&\leq 3^k 2^{k(2k+1)} \max_{1\leq j\leq k+1} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+1} \vee \|\Sigma\|) \mathbb{E}\prod_{i=1}^k \delta_i (1+\delta_i)^{2k+1} \\
&\leq 3^k 2^{k(2k+1)} \max_{1\leq j\leq k+1} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+1} \vee \|\Sigma\|) \\
&\times \Bigl(\mathbb{E}\|W-I\|(1+\|W-I\|)^{2k+1}\Bigr)^k \\
&\leq C^{k^2} \max_{1\leq j\leq k+1} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+1} \vee \|\Sigma\|) \left(\frac{d}{n}\right)^{k/2}.
\end{aligned}
$$

Part 4.
Asymptotic Efficiency

Problems

Let $X, X_1, \dots, X_n$ be i.i.d. Gaussian vectors with values in $\mathbb{R}^d$, with $\mathbb{E}X = 0$ and with covariance operator $\Sigma = \mathbb{E}(X \otimes X) \in \mathcal{C}_+^d$.

- Given a smooth function $f : \mathbb{R} \mapsto \mathbb{R}$ and a linear operator $B : \mathbb{R}^d \mapsto \mathbb{R}^d$ with $\|B\|_1 \leq 1$, estimate $\langle f(\Sigma), B \rangle$ based on $X_1, \dots, X_n$.
- More precisely, we are interested in finding asymptotically efficient estimators of $\langle f(\Sigma), B \rangle$ with $\sqrt{n}$-convergence rate in the case when $d = d_n \to \infty$.
- Suppose $d_n \leq n^\alpha$ for some $\alpha > 0$. Is there $s(\alpha)$ such that for all $s > s(\alpha)$ and for all functions f of smoothness s, asymptotically efficient estimation is possible?

Sample Covariance Operator

- Let
$$\hat{\Sigma} := n^{-1}\sum_{j=1}^{n} X_j \otimes X_j$$
be the sample covariance based on $(X_1, \dots, X_n)$.
- Let
$$\mathcal{S}_{a,d} := \Big\{\Sigma \in \mathcal{C}_+^d : a^{-1} I_d \preceq \Sigma \preceq a I_d\Big\}, a > 1.$$
- If $\Sigma \in \mathcal{S}_{a,d}$, then
$$\mathbb{E}\|\hat{\Sigma} - \Sigma\| \asymp_a \|\Sigma\| \Bigg(\sqrt{\frac{d}{n}} \bigvee \frac{d}{n}\Bigg)$$
and, for all $t \geq 1$ with probability at least $1 - e^{-t}$,
$$\|\hat{\Sigma} - \Sigma\| \lesssim_a \|\Sigma\| \Bigg(\sqrt{\frac{d}{n}} \bigvee \frac{d}{n} \bigvee \sqrt{\frac{t}{n}} \bigvee \frac{t}{n}\Bigg).$$

Operator Differentiability

- Let $f \in C^1(\mathbb{R})$ and let $f^{[1]}(\lambda, \mu)$ be the Loewner kernel:

$$f^{[1]}(\lambda, \mu) := \frac{f(\lambda) - f(\mu)}{\lambda - \mu}, \ \lambda \neq \mu; \ \ f^{[1]}(\lambda, \lambda) := f'(\lambda).$$

- $A \mapsto f(A)$ is Fréchet differentiable at $A = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda$ with derivative

$$Df(A; H) = \sum_{\lambda, \mu \in \sigma(A)} f^{[1]}(\lambda, \mu) P_\lambda H P_\mu.$$

- Let
$$\sigma_f^2(\Sigma; B) := 2\|\Sigma^{1/2} Df(\Sigma; B)\Sigma^{1/2}\|_2^2.$$

- Loss functions. Let $\mathcal{L}$ be the class of functions $\ell : \mathbb{R} \mapsto \mathbb{R}_+$ such that
  - $\ell(0) = 0$
  - $\ell(-t) = \ell(t), t \in \mathbb{R}$
  - ℓ is convex and nondecreasing on $\mathbb{R}_+$
  - For some $c > 0$, $\ell(t) = O(e^{ct})$ as $t \to \infty$
- Suppose that
  - A.1. $d_n \geq 3 \log n$
  - A.2. for some $\alpha \in (0, 1)$, $d_n \leq n^\alpha$
  - A.3. For some $s > \frac{1}{1-\alpha}$, $f \in B^s_{\infty,1}(\mathbb{R})$.

Efficient Estimation of $\langle f(\Sigma), B\rangle$

Theorem

Under assumptions A.1-A.3, there exists an estimator $h(\hat{\Sigma})$ *such that for all* $\sigma_0 > 0$

$$\sup_{\Sigma \in \mathcal{S}_{a,d_n}, \sigma_f(\Sigma;B) \geq \sigma_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{n^{1/2}\Big(\langle h(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle\Big)}{\sigma_f(\Sigma, B)} \leq x \right\} - \Phi(x) \right| \to 0$$

and, for all $\ell \in \mathcal{L}$,

$$\sup_{\Sigma \in \mathcal{S}_{a,d_n}, \sigma_f(\Sigma;B) \geq \sigma_0} \left| \mathbb{E}_\Sigma \ell \left(\frac{n^{1/2}\Big(\langle h(\hat{\Sigma}), B\rangle - \langle f(\Sigma), B\rangle\Big)}{\sigma_f(\Sigma, B)} \right) - \mathbb{E}\ell(Z) \right| \to 0.$$

Efficient Estimation of $\langle f(\Sigma), B\rangle$: A Lower Bound

Theorem

Let, for some $s \in (1,2]$, $f \in B^s_{\infty,1}(\mathbb{R})$. *Suppose* $d_n \geq 1, a > 1, \sigma_0 > 0$ *are such that, for some* $1 < a' < a$ *and* $\sigma_0' > \sigma_0$ *and for all large enough* n,

$$\left\{\Sigma \in \mathcal{S}_{a,d_n}, \sigma_f(\Sigma; B) \geq \sigma_0\right\} \neq \emptyset.$$

Then, the following bound holds:

$$\liminf_n \inf_{T_n} \sup_{\Sigma \in \mathcal{S}_{a,d_n}, \sigma_f(\Sigma;B)\geq\sigma_0} \frac{n\mathbb{E}_\Sigma\Big(T_n(X_1,\dots,X_n) - \langle f(\Sigma), B\rangle\Big)^2}{\sigma_f^2(\Sigma; B)} \geq 1,$$

where the infimum is taken over all sequences of estimators $T_n = T_n(X_1,\dots,X_n)$.

Lower bound: sketch of the proof

- Let

$$\Sigma_t := \Sigma_0 + \frac{tH}{\sqrt{n}} \ \text{and} \ \mathcal{S}_{c,n}(\Sigma_0, H) := \{\Sigma_t : t \in [-c, c]\},$$

  where

$$H := \Sigma_0 Df(\Sigma_0; B)\Sigma_0.$$

- For all large enough n,

$$\mathcal{S}_{c,n}(\Sigma_0, H) \subset \mathcal{S}_{a,d_n} \cap \{\Sigma : \sigma_f(\Sigma, B) \geq \sigma_0\}.$$

Lower bound: sketch of the proof

- Consider the following parametric model:

$$X_1, \dots, X_n \ \text{i.i.d.} \ \sim N(0; \Sigma_t), t \in [-c, c].$$

- The goal is to estimate the function

$$\varphi(t) := \langle f(\Sigma_t), B \rangle.$$

which is continuously differentiable with derivative

$$\varphi'(t) = \frac{1}{\sqrt{n}} \langle Df(\Sigma_t; H), B \rangle, t \in [-c, c].$$

- The Fisher information:

$$I_n(t) = nI(t) = \frac{1}{2} \langle I(\Sigma_t) H, H \rangle = \frac{1}{2} \langle (\Sigma_t^{-1} \otimes \Sigma_t^{-1}) H, H \rangle$$

$$= \frac{1}{2} \langle \Sigma_t^{-1} H \Sigma_t^{-1}, H \rangle = \frac{1}{2} \mathrm{tr}(\Sigma_t^{-1} H \Sigma_t^{-1} H) = \frac{1}{2} \|\Sigma_t^{-1/2} H \Sigma_t^{-1/2}\|_2^2.$$

Lower bound: sketch of the proof

- Let $\pi_c(t) := \frac{1}{c}\pi\left(\frac{t}{c}\right)$ for a smooth density π on $[-1,1]$ with $\pi(1) = \pi(-1) = 0$ and $J_\pi := \int_{-1}^{1} \frac{(\pi'(t))^2}{\pi(t)} dt < \infty.$
- Van Trees Inequality: for any estimator $T(X_1, \dots, X_n)$,

$$\sup_{t\in[-c,c]} \mathbb{E}_t(T_n(X_1, \dots, X_n) - \varphi(t))^2$$

$$\geq \int_{-c}^{c} \mathbb{E}_t(T_n(X_1, \dots, X_n) - \varphi(t))^2 \pi_c(t)dt \geq \frac{\left(\int_{-c}^{c} \varphi'(t)\pi_c(t)dt\right)^2}{\int_{-c}^{c} I_n(t)\pi_c(t)dt + J_{\pi_c}}.$$

Lower bound: sketch of the proof

Denote $\sigma^2(t) := \sigma_f^2(\Sigma_t; B), t \in [-c, c]$. Then

$$\sup_{\Sigma \in \mathcal{S}_{a,d_n}, \sigma_f(\Sigma;B) \geq \sigma_0} \frac{n\mathbb{E}_\Sigma(T_n(X_1, \dots, X_n) - \langle f(\Sigma), B\rangle)^2}{\sigma_f^2(\Sigma; B)}$$

$$\geq \sup_{t \in [-c,c]} \frac{n\mathbb{E}_t(T_n(X_1, \dots, X_n) - \varphi(t))^2}{\sigma^2(t)} \geq \frac{1 - \gamma_{n,c}(f; B; a; \sigma_0)}{1 + \frac{\lambda_{n,c}(f;B;a)}{\sigma_0^2}},$$

where

$$\gamma_{n,c}(f; B; a; \sigma_0) := \frac{\frac{3a^3\|f'\|_{L_\infty}^3\|B\|_1^3 c}{\sqrt{n}} + \frac{4a^{2s}\|f\|_{B^s_{\infty,1}}\|f'\|_{L_\infty}^s\|B\|_1^{1+s}c^{s-1}}{n^{(s-1)/2}} + \frac{J_\pi}{c^2}}{\frac{1}{4}\sigma_0^2 + \frac{3a^3\|f'\|_{L_\infty}^3\|B\|_1^3 c}{\sqrt{n}} + \frac{J_\pi}{c^2}}$$

and

$$\lambda_{n,c}(f; B; a) := \frac{6ca^3\|f'\|_{L_\infty}^3\|B\|_1^3}{n^{1/2}} + \frac{24c^{s-1}a^{2s}\|f'\|_{L_\infty}^s\|f\|_{B^s_{\infty,1}}\|B\|_1^{s+1}}{n^{(s-1)/2}}.$$

Construction of an Asymptotically Efficient Estimator

- If $d_n \leq n^{\alpha}$ for some $\alpha \in (0, 1/2)$ and $s > \frac{1}{1-\alpha}$, then the plug-in estimator $\langle f(\hat{\Sigma}), B\rangle$ is asymptotically efficient with convergence rate $\sqrt{n}$.
- If $d_n \geq n^{\alpha}$ for some $\alpha \geq 1/2$, then the plug-in estimator $\langle f(\hat{\Sigma}), B\rangle$, for “generic" smooth functions f has a large bias (larger than $n^{-1/2}$) and it is not even $\sqrt{n}$ consistent.
- In the last case, the crucial problem is to construct an estimator with a reduced bias.

Bounds on the Remainder of Differentiation

Lemma

Let $S_f(A;H) = f(A+H) - f(A) - (Df)(A;H)$ *be the remainder of differentiation. If, for some* $s \in [1,2]$, $f \in B^s_{\infty,1}(\mathbb{R})$, *then the following bounds hold:*

$$\|S_f(A;H)\| \leq 2^{3-s}\|f\|_{B^s_{\infty,1}}\|H\|^s$$

and

$$\|S_f(A;H) - S_f(A;H')\| \leq 2^{1+s}\|f\|_{B^s_{\infty,1}}(\|H\| \vee \|H'\|)^{s-1}\|H' - H\|.$$

The proof is based on Littlewood-Paley decomposition of f and on operator versions of Bernstein inequalities for entire functions of exponential type (as in the work by Peller (1985, 2006), Aleksandrov and Peller (2016) on operator Lipschitz and operator differentiable functions).

Perturbation Theory: Application to Functions of Sample Covariance (The Delta Method)

- $$\langle f(\hat\Sigma) - f(\Sigma), B\rangle = \langle Df(\Sigma;\hat\Sigma-\Sigma), B\rangle + \langle S_f(\Sigma;\hat\Sigma-\Sigma), B\rangle$$
- The linear term $\langle Df(\Sigma;\hat\Sigma-\Sigma), B\rangle$ is of the order $O(n^{-1/2})$ and $n^{1/2}\langle Df(\Sigma;\hat\Sigma-\Sigma), B\rangle$ is close in distribution to $N(0;\sigma_f^2(\Sigma;B))$.
- For $s\in(1,2]$, $\|S_f(\Sigma;\hat\Sigma-\Sigma)\| \lesssim \|f\|_{B^s_{\infty,1}}\|\hat\Sigma-\Sigma\|^s$, implying that

$$|\langle S_f(\Sigma;\hat\Sigma-\Sigma), B\rangle| \le \|B\|_1\|S_f(\Sigma;\hat\Sigma-\Sigma)\|$$

$$= O\Big(\Big(\frac{d}{n}\Big)^{s/2}\Big) = O(n^{(1-\alpha)s/2}) = o(n^{-1/2})$$

and, similarly,

$$|\langle \mathbb{E}f(\hat\Sigma) - f(\Sigma), B\rangle| = |\langle \mathbb{E}S_f(\Sigma;\hat\Sigma-\Sigma), B\rangle| = o(n^{-1/2})$$

provided that $s > \frac{1}{1-\alpha}, \alpha\in(0,1/2)$. In this case, $h(\hat\Sigma) = f(\hat\Sigma)$.

Wishart Operators and Bias Reduction

- Our next goal is to find an estimator $g(\hat{\Sigma})$ of $f(\Sigma)$ with a small bias $\mathbb{E}_{\Sigma} g(\hat{\Sigma}) - f(\Sigma)$ (of the order $o(n^{-1/2})$) and such that

$$n^{1/2}(\langle g(\hat{\Sigma}), B\rangle - \langle \mathbb{E}_{\Sigma} g(\hat{\Sigma}), B\rangle)$$

  is asymptotically normal.
- To this end, one has to find a sufficiently smooth approximate solution of the equation

$$\mathbb{E}_{\Sigma} g(\hat{\Sigma}) = f(\Sigma), \Sigma \in \mathcal{C}_{+}^{d}.$$

Wishart Operators and Bias Reduction

- $$\mathcal{T}g(\Sigma) := \mathbb{E}_\Sigma g(\hat{\Sigma}) = \int_{\mathcal{C}_+^d} g(V)P(\Sigma; dV), \Sigma \in \mathcal{C}_+^d$$

- To find an estimator of $f(\Sigma)$ with a small bias, one needs to solve (approximately) the following integral equation (“the Wishart equation")
$$\mathcal{T}g(\Sigma) = f(\Sigma), \Sigma \in \mathcal{C}_+^d.$$

- Bias operator: $\mathcal{B} := \mathcal{T} - \mathcal{I}$.
- Formally, the solution of the Wishart equation is given by Neumann series:
$$g(\Sigma) = (\mathcal{I} + \mathcal{B})^{-1} f(\Sigma) = (\mathcal{I} - \mathcal{B} + \mathcal{B}^2 - \dots) f(\Sigma)$$

Wishart Operators and Bias Reduction

- Given a smooth function $f : \mathbb{R} \mapsto \mathbb{R}$, define

$$f_k(\Sigma) := \sum_{j=0}^{k} (-1)^j \mathcal{B}^j f(\Sigma) := f(\Sigma) + \sum_{j=1}^{k} (-1)^j \mathcal{B}^j f(\Sigma)$$

- Then

$$\mathbb{E}_\Sigma f_k(\hat{\Sigma}) - f(\Sigma) = (\mathcal{I} + \mathcal{B}) f_k(\Sigma) - f(\Sigma) = (-1)^k \mathcal{B}^{k+1} f(\Sigma).$$

- Asymptotically efficient estimator is $h(\hat{\Sigma}) = f_k(\hat{\Sigma})$, where k is an integer such that, for some $\beta \in (0, 1]$, $\frac{1}{1-\alpha} < k + 1 + \beta \leq s$.

Bootstrap Chain

- $$\mathcal{T}g(\Sigma) := \mathbb{E}_{\Sigma} g(\hat{\Sigma}) = \int_{\mathcal{C}_+^d} g(V) P(\Sigma; dV), \Sigma \in \mathcal{C}_+^d$$

- $$\mathcal{T}^k g(\Sigma) = \mathbb{E}_{\Sigma} g(\hat{\Sigma}^{(k)}), \Sigma \in \mathcal{C}_+^d,$$
  where
  $$\hat{\Sigma}^{(0)} = \Sigma \to \hat{\Sigma}^{(1)} = \hat{\Sigma} \to \hat{\Sigma}^{(2)} \to \dots$$
  is a Markov chain in $\mathcal{C}_+^d$ with transition probability kernel P.
- Note that $\hat{\Sigma}^{(j+1)}$ is the sample covariance based on n i.i.d. observations $\sim N(0; \hat{\Sigma}^{(j)})$ (conditionally on $\hat{\Sigma}^{(j)}$)
- Conditionally on $\hat{\Sigma}^{(j)}$, with a “high probability",
  $$\|\hat{\Sigma}^{(j+1)} - \hat{\Sigma}^{(j)}\| \lesssim \|\hat{\Sigma}^{(j)}\| \sqrt{\frac{d}{n}}$$

k-th order difference

- k-th order difference along the Markov chain:

$$\mathcal{B}^k f(\Sigma) = (\mathcal{T} - \mathcal{I})^k f(\Sigma) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \mathcal{T}^j f(\Sigma)$$

$$= \mathbb{E}_{\Sigma} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(\hat{\Sigma}^{(j)})$$

- $\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(\hat{\Sigma}^{(j)})$ is the k-th order difference of f along the trajectory of the Bootstrap Chain. For f of smoothness k, it should be of the order $\left(\sqrt{\frac{d}{n}}\right)^k$.

A bound on $\mathcal{B}^k f(\Sigma)$

Theorem

Suppose that $f \in B^k_{\infty,1}(\mathbb{R})$ *and that* $k \leq d \leq n$. *Then, for some* $C > 1$,

$$\|\mathcal{B}^k f(\Sigma)\| \leq C^{k^2} \|f\|_{B^k_{\infty,1}} (\|\Sigma\|^{k+1} \vee \|\Sigma\|) \left(\frac{d}{n}\right)^{k/2}.$$

Bounds on the bias of $f_k(\hat{\Sigma})$

Corollary

Suppose $f \in B^{k+1}_{\infty,1}(\mathbb{R})$ and $k+1 \le d \le n$. Then, for some $C > 1$,

$$\|\mathbb{E}_\Sigma f_k(\hat{\Sigma}) - f(\Sigma)\| \le C^{(k+1)^2} \|f\|_{B^{k+1}_{\infty,1}} (\|\Sigma\|^{k+2} \vee \|\Sigma\|) \left(\frac{d}{n}\right)^{(k+1)/2}.$$

If, for some $\alpha \in (1/2, 1)$, $2 \log n \le d \le n^\alpha$ and $k > \frac{\alpha}{1-\alpha}$, then

$$\|\mathbb{E}_\Sigma f_k(\hat{\Sigma}) - f(\Sigma)\| = o(n^{-1/2}).$$

“Lifting" Operator $\mathcal{D}$ and Reduction to Orthogonally Invariant Functions

- Define the following differential operator (“lifting" operator):

$$\mathcal{D}g(\Sigma) := \Sigma^{1/2} Dg(\Sigma) \Sigma^{1/2}$$

acting on continuously differentiable functions in $\mathcal{C}_+^d$.

- Suppose $f(x) = x\psi'(x)$
- Let $g(\Sigma) := \mathrm{tr}(\psi(\Sigma))$
- g is orthogonally invariant function on $\mathcal{C}_+^d$
- $Dg(\Sigma) = \psi'(\Sigma)$
- $\mathcal{D}g(\Sigma) = f(\Sigma)$

An Integral Representation for Operator $\mathcal{B}^k$

- Linear interpolation between I and $W_1^{1/2}, \dots, W_k^{1/2}$ (i.i.d. copies of $W = n^{-1}\sum_{j=1}^n Z_j \otimes Z_j$):

$$V_j(t_j) := I + t_j(W_j^{1/2} - I), t_j \in [0,1], 1 \le j \le k.$$

- For all $j = 1, \dots, k, t_j \in [0,1]$, $V_j(t_j) \in \mathcal{C}_+^d$.
-

$$R := R(t_1, \dots, t_k) = V_1(t_1) \dots V_k(t_k),$$
$$L := L(t_1, \dots, t_k) = V_k(t_k) \dots V_1(t_1) = R^*,$$
$$S := S(t_1, \dots, t_k) = L(t_1, \dots, t_k)\Sigma R(t_1, \dots, t_k), (t_1, \dots, t_k) \in [0,1]^k.$$

- Let

$$\varphi(t_1, \dots, t_k) := \Sigma^{1/2} R(t_1, \dots, t_k) Dg(S(t_1, \dots, t_k)) L(t_1, \dots, t_k) \Sigma^{1/2}.$$

An Integral Representation for Operator $\mathcal{B}^k$

Proposition

Suppose $g \in L_\infty^O(\mathcal{C}_+^d)$ is $k+1$ times continuously differentiable function with uniformly bounded derivatives $D^j g, j = 1, \dots, k+1$. Then the function φ is k times continuously differentiable in $[0,1]^k$ and

$$\mathcal{B}^k \mathcal{D} g(\Sigma) = \mathbb{E} \int_0^1 \dots \int_0^1 \frac{\partial^k \varphi(t_1, \dots, t_k)}{\partial t_1 \dots \partial t_k} dt_1 \dots dt_k, \Sigma \in \mathcal{C}_+^d.$$

A bound on $\mathcal{B}^k\mathcal{D}g(\Sigma)$

Theorem

Suppose that $k \leq d \leq n$ and that $g \in L^O_\infty(\mathcal{C}^d_+)$ is $k+1$ times continuously differentiable function with uniformly bounded derivatives $D^j g, j = 1, \dots, k+1$. Then, for some $C > 1$,

$$\|\mathcal{B}^k\mathcal{D}g(\Sigma)\| \leq C^{k^2} \max_{1\leq j\leq k+1} \|D^j g\|_{L_\infty}(\|\Sigma\|^{k+1} \vee \|\Sigma\|)\left(\frac{d}{n}\right)^{k/2}.$$

Bounds on the bias of $\mathcal{D}g_k(\hat{\Sigma})$

Corollary

Suppose that $k+1 \leq d \leq n$ *and that* $g \in L_\infty^O(\mathcal{C}_+^d)$ *is* $k+2$ *times continuously differentiable function with uniformly bounded derivatives* $D^j g, j = 1, \ldots, k+2$. *Then, for some* $C > 1$,

$$\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\| \leq C^{(k+1)^2} \max_{1 \leq j \leq k+2} \|D^j g\|_{L_\infty} (\|\Sigma\|^{k+2} \vee \|\Sigma\|) \Big(\frac{d}{n}\Big)^{(k+1)/2}.$$

If, for some $\alpha \in (1/2, 1)$, $2 \log n \leq d \leq n^\alpha$ and $k > \frac{\alpha}{1-\alpha}$, then

$$\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\| = o(n^{-1/2}).$$

Further bounds on the bias of $\mathcal{D}g_k(\hat{\Sigma})$

Theorem

Suppose $g \in L_\infty^O(\mathcal{C}_+^d)$ *is* $k+2$ *times continuously differentiable for some* $k \le d \le n$ *and, for some* $\beta \in (0,1]$, $\|Dg\|_{C^{k+1+\beta}} < \infty$. *In addition, suppose that for some* $\delta > 0$ $\sigma(\Sigma) \subset \left[\delta, \frac{1}{\delta}\right]$. *Then, for some constant* $C > 0$,

$$\|\mathbb{E}_\Sigma \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma)\| \le C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} (\|\Sigma\| \vee 1)^{k+3/2} \|\Sigma\| \left(\frac{d}{n}\right)^{(k+1+\beta)/2}.$$

Further details of the proof

- We use the representation

$$\langle \mathcal{D}g_k(\hat\Sigma) - \mathcal{D}g(\Sigma), B\rangle = \langle D\mathcal{D}g_k(\Sigma)(\hat\Sigma - \Sigma), B\rangle$$
$$+ S_{\mathfrak{d}_k}(\Sigma; \hat\Sigma - \Sigma) - \mathbb{E}S_{\mathfrak{d}_k}(\Sigma; \hat\Sigma - \Sigma) + \langle \mathbb{E}\mathcal{D}g_k(\hat\Sigma) - \mathcal{D}g(\Sigma), B\rangle,$$

  where $\mathfrak{d}_k(\Sigma) := \langle \mathcal{D}g_k(\Sigma), B\rangle$.
- We control the bias $\langle \mathbb{E}\mathcal{D}g_k(\hat\Sigma) - \mathcal{D}g(\Sigma), B\rangle$ as follows:

$$|\langle \mathbb{E}\mathcal{D}g_k(\hat\Sigma) - \mathcal{D}g(\Sigma), B\rangle| = O\left(\frac{d}{n}\right)^{(k+1+\beta)/2} = o(n^{-1/2})$$

  provided that $d \leq n^{\alpha}, \alpha \in (0,1)$ and $s \geq k + 1 + \beta > \frac{1}{1-\alpha}$.
- We also need to prove concentration of

$$S_{\mathfrak{d}_k}(\Sigma; \hat\Sigma - \Sigma) - \mathbb{E}S_{\mathfrak{d}_k}(\Sigma; \hat\Sigma - \Sigma).$$

Lipschitz Condition for the Remainder of Taylor Expansion of $\mathcal{D}g_k(\Sigma)$

Lemma

Suppose that, for some $k \leq d$, g *is* $k+2$ *times continuously differentiable and, for some* $\beta \in (0,1]$, $\|Dg\|_{C^{k+1+\beta}} < \infty$. *Suppose also that* $g \in L_\infty^O(\mathcal{C}_+^d)$, $d \leq n/2$ *and that, for some* $\delta \in (0,1)$, $\sigma(\Sigma) \subset \left[\delta, \frac{1}{\delta}\right]$. *Denote*

$$\gamma_{\beta,k}(\Sigma; u) := (\|\Sigma\| \vee u \vee 1)^{k+1/2}(u \vee u^{\beta}), u > 0, \beta \in [0,1], k \geq 1.$$

Then, for some constant $C \geq 1$ *and for all* $H, H' \in \mathcal{B}_{sa}(\mathbb{R}^d)$

$$\|S_{\mathcal{D}g_k}(\Sigma; H') - S_{\mathcal{D}g_k}(\Sigma; H)\|$$
$$\leq C^{k^2} \frac{\log^2(2/\delta)}{\delta} \|Dg\|_{C^{k+1+\beta}} \gamma_{\beta,k}(\Sigma; \|H\| \vee \|H'\|)\|H' - H\|.$$

Sketch of the proof of the Lemma

- Let $\gamma : \mathbb{R} \mapsto \mathbb{R}$ be a C^∞ function such that:

$$0 \leq \gamma(u) \leq \sqrt{u}, u \geq 0, \gamma(u) = \sqrt{u}, u \in \left[\delta, \frac{1}{\delta}\right],$$
$$\mathrm{supp}(\gamma) \subset \left[\frac{\delta}{2}, \frac{2}{\delta}\right] \text{ and } \|\gamma\|_{B^1_{\infty,1}} \lesssim \frac{\log(2/\delta)}{\sqrt{\delta}}.$$

- For instance, one can take $\gamma(u) := \lambda(u/\delta)\sqrt{u}(1 - \lambda(\delta u/2))$, where λ is a C^∞ non-decreasing function with values in $[0, 1]$, $\lambda(u) = 0, u \leq 1/2$ and $\lambda(u) = 1, u \geq 1$.

Further Integral Representations

- $$\phi(t_1,\dots,t_k;s_1,s_2) :=$$
$$\gamma(\bar\Sigma(s_1,s_2))R(t_1,\dots,t_k)Dg\Bigl(L(t_1,\dots,t_k)\bar\Sigma(s_1,s_2)R(t_1,\dots,t_k)\Bigr)$$
$$L(t_1,\dots,t_k)\gamma(\bar\Sigma(s_1,s_2)),$$

  where
$$\bar\Sigma(s_1,s_2) = \Sigma + s_1 H + s_2(H'-H), s_1,s_2\in\mathbb{R}.$$

- Note that
$$\varphi(t_1,\dots,t_k) = \phi(t_1,\dots,t_k,0,0).$$

Further Integral Representations

- $$B_k(\Sigma) := \mathcal{B}^k \mathcal{D} g(\Sigma), \quad D_k(\Sigma) := \mathcal{D} g_k(\Sigma).$$

- $$D_k(\Sigma) = \sum_{j=0}^{k} (-1)^j B_j(\Sigma)$$

- $$B_k(\Sigma) := \mathbb{E} \int_0^1 \dots \int_0^1 \frac{\partial^k \phi(t_1, \dots, t_k, 0, 0)}{\partial t_1 \dots \partial t_k} dt_1 \dots dt_k$$

Further Integral Representations

- $$S_{B_k}(\Sigma; H') - S_{B_k}(\Sigma; H)$$
  $$= DB_k(\Sigma + H; H' - H) - DB_k(\Sigma; H' - H) + S_{B_k}(\Sigma + H; H' - H)$$

- $$DB_k(\Sigma + H; H' - H) - DB_k(\Sigma; H' - H)$$
  $$= \mathbb{E}\int_0^1 \dots \int_0^1 \Big[\frac{\partial^{k+1}\phi(t_1, \dots, t_k, 1, 0)}{\partial t_1 \dots \partial t_k \partial s_2}$$
  $$- \frac{\partial^{k+1}\phi(t_1, \dots, t_k, 0, 0)}{\partial t_1 \dots \partial t_k \partial s_2}\Big] dt_1 \dots dt_k$$

Further Integral Representations

- $$\begin{aligned} & S_{B_k}(\Sigma + H; H' - H) \\ & = \mathbb{E} \int_0^1 \dots \int_0^1 \int_0^1 \Big[\frac{\partial^{k+1} \phi(t_1, \dots, t_k, 1, s_2)}{\partial t_1 \dots \partial t_k \partial s_2} \\ & - \frac{\partial^{k+1} \phi(t_1, \dots, t_k, 1, 0)}{\partial t_1 \dots \partial t_k \partial s_2} \Big] ds_2 dt_1 \dots dt_k. \end{aligned}$$

- The rest of the proof is based on bounding the partial derivatives involved in the integral representations, such as, for instance,

$$\left\|\frac{\partial^{k+1}\phi(t_1,\dots,t_k,1,0)}{\partial t_1\dots\partial t_k\partial s_2}-\frac{\partial^{k+1}\phi(t_1,\dots,t_k,0,0)}{\partial t_1\dots\partial t_k\partial s_2}\right\|$$

$$\leq C^{k^2}\frac{\log^2(2/\delta)}{\delta}\|Dg\|_{C^{k+1+\beta}}((\|\Sigma\|+\|H\|)^{k+1/2}\vee 1)$$

$$\prod_{i=1}^{k}\delta_i(1+\delta_i)^{2k+5}(\|H\|\vee\|H\|^{\beta})\|H'-H\|,$$

where $\delta_i := \|W_i - I\|$.

Concentration of the Remainder: A More General Version

Assumption

Assume that, for all $\Sigma \in \mathcal{C}_+(\mathbb{H}), H, H' \in \mathcal{B}_{sa}(\mathbb{H})$,

$$|S_h(\Sigma; H') - S_h(\Sigma; H)| \leq \eta(\Sigma; \|H\| \vee \|H'\|)\|H' - H\|,$$

where $0 < \delta \mapsto \eta(\Sigma; \delta)$ is a nondecreasing function of the following form:

$$\eta(\Sigma; \delta) := \eta_1(\Sigma)\delta^{\alpha_1} \bigvee \cdots \bigvee \eta_m(\Sigma)\delta^{\alpha_m},$$

for given nonnegative functions $\eta_1, \dots, \eta_m$ on $\mathcal{C}_+(\mathbb{H})$ and given positive numbers $\alpha_1, \dots \alpha_m$.

Concentration of the Remainder: A More General Version

Theorem

For all $t \geq 1$ *with probability at least* $1 - e^{-t}$,

$$|S_h(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E} S_h(\Sigma; \hat{\Sigma} - \Sigma)|$$

$$\lesssim_\eta \eta(\Sigma; \delta_n(\Sigma; t))\Big(\sqrt{\|\Sigma\|} + \sqrt{\delta_n(\Sigma; t)}\Big)\sqrt{\|\Sigma\|}\sqrt{\frac{t}{n}},$$

where

$$\delta_n(\Sigma; t) := \|\Sigma\| \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{n}} \bigvee \frac{\mathbf{r}(\Sigma)}{n} \bigvee \sqrt{\frac{t}{n}} \bigvee \frac{t}{n}\right).$$

Concentration Inequality for the Remainder of Taylor Expansion of $\langle \mathcal{D}g_k(\Sigma), B\rangle$

Lemma

With probability at least $1 - e^{-t}$,

$$|S_{\mathfrak{d}_k}(\Sigma;\hat\Sigma-\Sigma) - \mathbb{E}S_{\mathfrak{d}_k}(\Sigma;\hat\Sigma-\Sigma)|$$

$$\le C^{k^2}\tilde\Lambda_{k,\beta}(g;\Sigma;B)\gamma_{\beta,k}(\Sigma;\bar\delta_n(\Sigma;t))\Big(\sqrt{\|\Sigma\|}+\sqrt{\bar\delta_n(\Sigma;t)}\Big)\sqrt{\|\Sigma\|}\sqrt{\frac{t}{n}},$$

where

$$\bar\delta_n(\Sigma;t) := \|\Sigma\|\left(\sqrt{\frac{d}{n}}\bigvee\sqrt{\frac{t}{n}}\bigvee\frac{t}{n}\right),$$

$$\tilde\Lambda_{k,\beta}(g;\Sigma;B) := \|B\|_1\|Dg\|_{C^{k+1+\beta}}(\|\Sigma\|\vee\|\Sigma^{-1}\|)\log^2(2(\|\Sigma\|\vee\|\Sigma^{-1}\|)).$$

Normal Approximation of $\langle \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma), B\rangle$

- Recall that

$$\langle \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma), B\rangle = \langle D\mathcal{D}g_k(\Sigma)(\hat{\Sigma} - \Sigma), B\rangle$$
$$+ S_{\mathfrak{d}_k}(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_{\mathfrak{d}_k}(\Sigma; \hat{\Sigma} - \Sigma) + \langle \mathbb{E}\mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma), B\rangle$$

- It follows from the concentration bound on the remainder that, for $d = o(n)$,

$$|S_{\mathfrak{d}_k}(\Sigma; \hat{\Sigma} - \Sigma) - \mathbb{E}S_{\mathfrak{d}_k}(\Sigma; \hat{\Sigma} - \Sigma)| = o_{\mathbb{P}}(n^{-1/2})$$

Normal Approximation of $\langle \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma), B\rangle$

- By the bound on the bias,

$$|\langle \mathbb{E}\mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma), B\rangle| = O\left(\frac{d}{n}\right)^{(k+1+\beta)/2} = o(n^{-1/2})$$

  provided that $d \leq n^{\alpha}, \alpha \in (0,1)$ and $s \geq k+1+\beta > \frac{1}{1-\alpha}$.
- Thus

$$\langle \mathcal{D}g_k(\hat{\Sigma}) - \mathcal{D}g(\Sigma), B\rangle = \langle D\mathcal{D}g_k(\hat{\Sigma})(\hat{\Sigma}-\Sigma), B\rangle + o_{\mathbb{P}}(n^{-1/2}),$$

  and asymptotic normality of $n^{1/2}\langle D\mathcal{D}g_k(\hat{\Sigma})(\hat{\Sigma}-\Sigma), B\rangle$ follows from Berry-Esseen bound.

Efficient Estimation of Linear Functionals of Principal Components (Koltchinskii, Löffler and Nickl (2017))

- $\lambda(\Sigma) = \|\Sigma\|$
- $\lambda(\Sigma)$ eigenvalue of multiplicity 1
- $g(\Sigma) := \mathrm{dist}(\lambda(\Sigma); \sigma(\Sigma) \setminus \{\lambda(\Sigma)\})$
- $\theta(\Sigma)$ eigenvector of Σ corresponding to $\lambda(\Sigma)$, $\|\theta(\Sigma)\| = 1$ (the top principal component)
- Problem: given $u \in \mathbb{H}$, estimate $\langle \theta(\Sigma), u \rangle$ based on i.i.d. observations $X_1, \dots, X_n \sim N(0; \Sigma)$

Efficient Estimation of Linear Functionals of Principal Components

- $\Sigma = \sum_{\lambda \in \sigma(\Sigma)} \lambda P_\lambda$
- $C(\Sigma) := \sum_{\lambda \in \sigma(\Sigma), \lambda \neq \lambda(\Sigma)} \frac{1}{\lambda - \lambda(\Sigma)} P_\lambda$
- $\sigma_u^2(\Sigma) := \|\Sigma\| \left\langle \Sigma C(\Sigma) u, C(\Sigma) u \right\rangle$
- For $r > 1, a > 1, \sigma_0^2 > 0$, denote

$$\mathcal{S}(r; a) := \left\{ \Sigma : \mathbf{r}(\Sigma) \leq r, \frac{\|\Sigma\|}{g(\Sigma)} \leq a \right\}$$

Efficient Estimation of Linear Functionals: Lower Bound

Theorem

Let $r > 1, a > 1, \sigma_0^2 > 0$. *Suppose, for some* $r' < r, a' < a$ *and* $\sigma_0' > \sigma_0$

$$\mathcal{S}(r'; a') \cap \{\Sigma : \sigma_u(\Sigma) \geq \sigma_0'\} \neq \emptyset.$$

Then

$$\liminf_n \inf_{T_n} \sup_{\Sigma \in \mathcal{S}(r;a), \sigma_u(\Sigma) \geq \sigma_0} \frac{n\mathbb{E}_\Sigma \Big(T_n(X_1, \ldots, X_n) - \langle \theta(\Sigma), u \rangle\Big)^2}{\sigma_u^2(\Sigma)} \geq 1,$$

where the infimum is taken over all sequences of estimators $T_n = T_n(X_1, \ldots, X_n)$.

Efficient Estimation of Linear Functionals: Asymptotic Normality

Theorem

Suppose $a > 1, \sigma_0^2 > 0$ *and* $r_n = o(n)$ as $n \to \infty$. *There exists a sequence of estimators* $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n) \in \mathbb{H}$ *such that*

$$\sup_{\Sigma \in \mathcal{S}(r_n;a), \sigma_u(\Sigma) \geq 0} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{n^{1/2}\Big(\langle \hat{\theta}_n, u\rangle - \langle \theta(\Sigma), u\rangle\Big)}{\sigma_u(\Sigma)} \leq x \right\} - \Phi(x) \right| \to 0$$

and, for all $\ell \in \mathcal{L}$,

$$\sup_{\Sigma \in \mathcal{S}(r_n;a), \sigma_u(\Sigma) \geq \sigma_0} \left| \mathbb{E}_\Sigma \ell\left(\frac{n^{1/2}\Big(\langle \hat{\theta}_n, u\rangle - \langle \theta(\Sigma), u\rangle\Big)}{\sigma_u(\Sigma)} \right) - \mathbb{E}\ell(Z) \right| \to 0$$

as $n \to \infty$.

Plug-in estimator $\theta(\hat{\Sigma}_n)$

Theorem (Koltchinskii and Lounici (2016))

Suppose $r_n = o(n)$. *Then*

$$\sup_{\Sigma \in \mathcal{S}(r_n;a), \sigma_u(\Sigma) \geq \sigma_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \zeta_n(\Sigma; u) \leq x \right\} - \Phi(x) \right| \to 0 \text{ as } n \to \infty,$$

where

$$\zeta_n(\Sigma; u) := \frac{n^{1/2} \Big(\langle \theta(\hat{\Sigma}_n), u \rangle - d_n(\Sigma) \langle \theta(\Sigma), u \rangle \Big)}{\sigma_u(\Sigma)}$$

and

$$d_n(\Sigma) := \mathbb{E}_\Sigma^{1/2} \langle \theta(\hat{\Sigma}_n), \theta(\Sigma) \rangle^2.$$

Moreover, $d_n^2(\Sigma) = 1 + b_n(\Sigma)$ with $|b_n(\Sigma)| \asymp \frac{\|\Sigma\|^2}{g^2(\Sigma)} \frac{\mathbf{r}(\Sigma)}{n}$.

Plug-in estimator $\theta(\hat{\Sigma}_n)$

- $\langle \theta(\hat{\Sigma}_n), u\rangle$ “concentrates around" $d_n(\Sigma)\langle \theta(\Sigma), u\rangle$
- Its “bias" $(d_n(\Sigma)-1)\langle \theta(\Sigma), u\rangle$ could be as large as $\frac{r_n}{n}$
- If $\sqrt{n} \leq r_n = o(n)$, the bias is too large.

Bias Reduction (Koltchinskii and Lounici (2016))

- Suppose $n = 2n'$
- $\hat{\Sigma}^{(1)}, \hat{\Sigma}^{(2)}$ are sample covariances based on two sub-samples of size n'
- $\hat{b}_n := \langle \theta(\hat{\Sigma}^{(1)}), \theta(\hat{\Sigma}^{(2)}) \rangle - 1$
- For some $\delta_n \to 0$, $\sup_{\Sigma \in \mathcal{S}(r_n;a)} \mathbb{P}_\Sigma \left\{ |\hat{b}_n - b_{n/2}(\Sigma)| \geq \frac{\delta_n}{\sqrt{n}} \right\} \to 0$
- Let $\hat{d}_n := \sqrt{1 + \hat{b}_n}$ and $\tilde{\theta}_n := \theta(\hat{\Sigma}^{(1)})/\hat{d}_n$. Then, under the assumption $r_n = o(n)$,

$$\sup_{\Sigma \in \mathcal{S}(r_n;a), \sigma_u(\Sigma) \geq \sigma_0} \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left\{ \frac{\sqrt{\frac{n}{2}} \left(\langle \tilde{\theta}_n, u \rangle - \langle \theta(\Sigma), u \rangle \right)}{\sigma_u(\Sigma)} \leq x \right\} - \Phi(x) \right| \to 0.$$

Construction of Asymptotically Efficient Estimators $\hat{\theta}_n$.

- Recall that $r_n = o(n)$
- Let $m = m_n$ be such that $m_n = o(n)$ and $r_n = o(m_n)$.
- Split the sample $X_1, \dots, X_n$ into three sub-samples, the first one of size $n' = n - 2m$ and two others of size m each and construct three sample covariances, $\hat{\Sigma}^{(1)}, \hat{\Sigma}^{(2)}$ and $\hat{\Sigma}^{(3)}$, based on each of the sub-samples.
- Finally, define

$$\tilde{d}_n := \frac{\langle \theta(\hat{\Sigma}^{(1)}), \theta(\hat{\Sigma}^{(2)}) \rangle}{\langle \theta(\hat{\Sigma}^{(2)}), \theta(\hat{\Sigma}^{(3)}) \rangle^{1/2}}$$

and

$$\hat{\theta}_n := \frac{\theta(\hat{\Sigma}^{(1)})}{\tilde{d}_n}.$$