確率過程の高次元統計　講義ノート [*,†]

小池 祐太 [‡]

2025 年 4 月 30 日

目次

---

[*] 大阪大学大学院基礎工学研究科での 2022 年度の集中講義 (科目名: 数理特論 II) の講義ノートです. 校正をしっかりと行っていないので, 誤植・誤りがあると思います. 気付きましたらご連絡いただけるとありがたいです.

[†] 2022 年 9 月 27 日のバージョンはかなり誤りを含んでいたので修正しました. 特に, ブートストラップに関する結果の証明を修正しています.

[‡] 東京大学 数理情報・教育研究センター, 大学院数理科学研究科. Email: kyuta (at) ms.u-tokyo.ac.jp

■各節の参考文献

- 1 節: van Handel (2017)
- 2 節: Kuchibhotla & Chakrabortty (2022); van der Vaart & Wellner (1996); Vershynin (2018)
- 3 節: Boucheron, Lugosi & Massart (2013); Vershynin (2018); Wainwright (2019)
- 4 節
  - 4.1 節: Ikeda & Watanabe (1989); Jacod & Shiryaev (2003); Karatzas & Shreve (1998); 風巻 (2003); 長井 (1999); Revuz & Yor (1999)
  - 4.2 節: Aït-Sahalia & Jacod (2014); Jacod & Protter (2012)
- 5 節: Barlow & Yor (1982); Revuz & Yor (1999)
- 6 節: Janková & van de Geer (2018)
- 7 節: Vershynin (2018)
- 8 節
  - 8.1 節: Horn & Johnson (2013); Magnus & Neudecker (2019)
  - 8.2–8.3 節: Aït-Sahalia & Xiu (2017); Fan, Furger & Xiu (2016); Pelger (2019)
- 9 節
  - 高次元中心極限定理関連: Belloni et al. (2018); Chernozhukov, Chetverikov & Kato (2013, 2017); Chernozhukov et al. (2022, 2023); Koike (2019); Wasserman (2014)
  - Stein の方法関連: Chen, Shao & Goldstein (2011); Nourdin & Peccati (2012)
  - 最大不等式関連: Boucheron, Lugosi & Massart (2013); Bühlmann & van de Geer (2011); Ledoux & Talagrand (1991); van der Vaart & Wellner (1996)

■謝辞　2022 年 9 月 27 日のバージョンについて, 東京大学大学院経済学研究科統計学コースの院生有志の皆様から誤り・誤植をご指摘いただき修正しました. ここに感謝を申し上げます. (2023 年 4 月 11 日)

本講義ノートを通じて, 確率空間 $(\Omega, \mathcal{F}, \mathrm{P})$ を 1 つ固定して考える. 特に断りがない限り, 考えている確率変数はすべてこの確率空間 $(\Omega, \mathcal{F}, \mathrm{P})$ 上で定義されているものとする. また, 特に断らない限り, 確率変数の間の等式や不等式は, すべて確率 1 で成立しているものと解釈する. 例えば, 2 つの確率変数 X, Y に対して $X = Y$ と書いた場合には, P-a.s. でこの等式が成立していることを意味している.

1 導入: 独立正規標本の共分散行列の推定

平均 0, 共分散行列 Σ の d 次元正規分布に従う独立同分布な確率変数列 $X_1, \dots, X_n$ が観測データとして得られているとし, Σ を推定する問題を考える. Σ の推定量としては, 標本共分散行列

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top$$

を考えるのが自然であろう (ここでは平均 0 と仮定しているので中心化は行なっていない). 任意の $j, k = 1, \dots, d$ に対して,

$$\mathrm{E}\left[\left|\widehat{\Sigma}_{n,jk} - \Sigma_{jk}\right|^2\right] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^n X_{ij}X_{ik}\right] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left[X_{ij}X_{ik}\right] = \frac{1}{n}\mathrm{Var}\left[X_{1j}X_{1k}\right] \tag{1.1}$$

が成り立つから, $n \to \infty$ のとき $\widehat{\Sigma}_{n,jk}$ は Σ_{jk} に L^2-収束する. 特に, $\widehat{\Sigma}_{n,jk}$ は Σ_{jk} の (弱) 一致推定量である (実際には大数の強法則より強一致推定量になっている).

$p \times q$ 行列 $A = (A_{jk})_{1 \le j \le p, 1 \le k \le q}$ に対して, A の Frobenius ノルムを $\|A\|_F$ で表すことにする:

$$\|A\|_F^2 = \mathrm{tr}(AA^\top) = \sum_{j=1}^p \sum_{k=1}^q A_{jk}^2.$$

d が n によらず固定されているような低次元の設定では, $n \to \infty$ のとき $\|\widehat{\Sigma}_n - \Sigma\|_F$ が 0 に L^2-収束することは, 上の結果から直ちに従う. 特に, $\widehat{\Sigma}_n$ が Frobenius ノルムに関して Σ の一致推定量となることがわかる. しかし, $n \to \infty$ のとき $d = d_n \to \infty$ となるような高次元の設定では, このような議論は一般には正当化できない. 実際, (1.1) より

$$\mathrm{E}\left[\left\|\widehat{\Sigma}_n - \Sigma\right\|_F^2\right] = \frac{1}{n}\sum_{j,k=1}^d \mathrm{Var}\left[X_{1j}X_{1k}\right] = \frac{\mathrm{E}\,|X_1|^4 - \|\Sigma\|_F^2}{n}$$

が成り立つことがわかる. ただし, ベクトル $x \in \mathbb{R}^d$ に対して $|x|$ は Euclid ノルムを表す. いま, Σ が単位行列 I_d の場合を考えると, X_1 の成分は独立となるから,

$$\mathrm{Var}[|X_1|^2] = \mathrm{Var}\left[\sum_{j=1}^d X_{1j}^2\right] = \sum_{j=1}^d \mathrm{Var}[X_{1j}^2] = 2d$$

が成り立つ. 従って,

$$\mathrm{E}\left[\left\|\widehat{\Sigma}_n - \Sigma\right\|_F^2\right] = \frac{(\mathrm{E}\,|X_1|^2)^2 + \mathrm{Var}[|X_1|^2] - d}{n} = \frac{d(d+1)}{n}$$

を得る. 故に, $n \to \infty$ のとき $\|\widehat{\Sigma}_n - \Sigma\|_F$ が 0 に L^2-収束するためには, $d^2 = o(n)$ という条件が必要である. 特に, データの次元 d はサンプル数 n よりずっと小さくなければならない.

Σ は未知パラメータを $d(d+1)/2$ 個含むため,「$d^2 = o(n)$」という条件は「サンプル数が未知パラメータの数より十分大きい」という条件に対応することになり, 直感的には自然な要請に思える. しかし, 推定誤差 $\widehat{\Sigma}_n - \Sigma$ の計測方法を, Frobenius ノルムから別のノルムに変更すると, 以下で見るように, 一致性を得るために必要な次元 d に関する条件をはるかに緩められるケースがある.

命題 1.1. $p \times q$ 行列 $A = (A_{jk})_{1 \le j \le p, 1 \le k \le q}$ に対して $|A|_\infty := \max_{1 \le j \le p, 1 \le k \le q} |A_{jk}|$ と定める. このとき,

$$\mathrm{E}\left[\left|\widehat{\Sigma}_n - \Sigma\right|_\infty^2\right] \le \frac{64|\Sigma|_\infty^2 \log^2(d^2\sqrt{2})}{n} \tag{1.2}$$

が成り立つ.

証明. 3 ステップに分けて証明する.

Step 1. $d \times n$ ランダム行列 X を $X = (X_1, \dots, X_n)$ で定義する. $\widehat{\Sigma}_n = XX^\top/n$ と書けることに注意する. また, $\widetilde{X} = (\widetilde{X}_1, \dots, \widetilde{X}_n)$ を X の独立なコピー, すなわち X と同じ分布をもち X と独立であるような $d \times n$ ランダム行列とする. このとき, $\mathrm{E}[\widetilde{X} \mid X] = 0$ かつ $\mathrm{E}[\widetilde{X}\widetilde{X}^\top/n \mid X] = \Sigma$ となることに注意すれば,

$$\widehat{\Sigma}_n - \Sigma = \frac{1}{n}\mathrm{E}\left[XX^\top - \widetilde{X}\widetilde{X}^\top \mid X\right] = \frac{1}{n}\mathrm{E}\left[(X + \widetilde{X})(X - \widetilde{X})^\top \mid X\right]$$

が成り立つ. 従って, Jensen の不等式より,

$$\mathrm{E}\left[\left|\widehat{\Sigma}_n - \Sigma\right|_\infty^2\right] \le \frac{1}{n^2}\mathrm{E}\left[\left|(X + \widetilde{X})(X - \widetilde{X})^\top\right|_\infty^2\right]$$

を得る. ここで, $X_i \overset{i.i.d.}{\sim} N_d(0, \Sigma)$ であることに注意すると, $(X + \widetilde{X}, X - \widetilde{X})$ と $\sqrt{2}(X, \widetilde{X})$ は同じ分布に従うことがわかる. 故に

$$\mathrm{E}\left[\left|\widehat{\Sigma}_n - \Sigma\right|_\infty^2\right] \le \frac{4}{n^2}\mathrm{E}\left[\left|X\widetilde{X}^\top\right|_\infty^2\right] = \frac{4}{n^2}\mathrm{E}\left[\left|\sum_{i=1}^n X_i\widetilde{X}_i^\top\right|_\infty^2\right]$$

が成り立つ.

Step 2. $Z_1, \dots, Z_p$ がそれぞれ平均 0 の正規分布に従う確率変数ならば,

$$\mathrm{E}\left[\max_{1 \le j \le p} Z_j^2\right] \le 4\log(p\sqrt{2}) \max_{1 \le j \le p} \mathrm{E}[Z_j^2] \tag{1.3}$$

が成り立つことを示す. 各 $j = 1, \dots, p$ について $\sigma_j := \sqrt{\mathrm{E}[Z_j^2]}$ と定め, $\overline{\sigma} := \max_{1 \le j \le d} \sigma_j$ とおく. $\sigma_j = 0$ のとき $Z_j = 0$ となることに注意すると, すべての j について $\sigma_j > 0$ であると仮定して一般性を失わない. このとき,

$$\mathrm{E}\left[\max_{1 \le j \le p} Z_j^2\right] \le \overline{\sigma}^2 \,\mathrm{E}\left[\max_{1 \le j \le p} \frac{Z_j^2}{\sigma_j^2}\right]$$

が成り立つ. 従って, すべての j について $\sigma_j = 1$ であるような場合に (1.3) を示せば十分であることがわかる. この場合, Jensen の不等式より

$$\mathrm{E}\left[\max_{1\le j\le p} Z_j^2\right] = 4\log\left(\exp\left(\mathrm{E}\left[\max_{1\le j\le p} Z_j^2/4\right]\right)\right) \le 4\log \mathrm{E}\left[\exp\left(\max_{1\le j\le p} Z_j^2/4\right)\right]$$

が成り立つことと, 各 j について Z_j^2 が自由度 1 の χ^2 分布に従うことから

$$\mathrm{E}\left[\exp\left(\max_{1\le j\le p} Z_j^2/4\right)\right] = \mathrm{E}\left[\max_{1\le j\le p}\exp\left(Z_j^2/4\right)\right] \le \sum_{j=1}^p \mathrm{E}\left[\exp\left(Z_j^2/4\right)\right] = p\sqrt{2}$$

が成り立つことを組み合わせることで, (1.3) が得られる.

Step 3. $\widetilde{X}$ が与えられた下で, $\sum_{i=1}^n X_{ij}\widetilde{X}_{ik}$ $(j,k=1,\dots,d)$ はそれぞれ平均 0, 分散 $\Sigma_{jj}\sum_{i=1}^n \widetilde{X}_{ik}^2$ の正規分布に従うことに注意すると, (1.3) より

$$\mathrm{E}\left[\left|\sum_{i=1}^n X_i\widetilde{X}_i^\top\right|_\infty^2 \mid \widetilde{X}\right] \le 4\log(d^2\sqrt{2})|\Sigma|_\infty \max_{1\le k\le d}\sum_{i=1}^n \widetilde{X}_{ik}^2$$

が成り立つ. また,

$$\mathrm{E}\left[\max_{1\le k\le d}\sum_{i=1}^n \widetilde{X}_{ik}^2\right] \le \sum_{i=1}^n \mathrm{E}\left[\max_{1\le k\le d}\widetilde{X}_{ik}^2\right] = n\,\mathrm{E}\left[\max_{1\le k\le d}\widetilde{X}_{1k}^2\right]$$

であり, X_1 の各成分は平均 0 の正規分布に従うから, 再び (1.3) より

$$\mathrm{E}\left[\max_{1\le k\le d}\sum_{i=1}^n \widetilde{X}_{ik}^2\right] \le 4n\log(d\sqrt{2})|\Sigma|_\infty$$

が成り立つ. 以上より,

$$\mathrm{E}\left[\left|\sum_{i=1}^n X_i\widetilde{X}_i^\top\right|_\infty^2\right] \le 16n\log^2(d^2\sqrt{2})|\Sigma|_\infty^2$$

を得る. この評価と Step 1 をあわせて示すべき結論を得る. □

命題 1.1 より, 例えば $\Sigma = I_d$ の場合, $n\to\infty$ のとき $|\widehat{\Sigma}_n - \Sigma|_\infty$ が 0 に L^2-収束するためには, $\log^2 d = o(n)$ という条件があれば十分である. これは, d が n よりはるかに大きい状況を許容する. 後に見るように, いくつかの統計的応用においては, $|\widehat{\Sigma}_n - \Sigma|_\infty$ の収束さえあれば十分なため, この結果は応用上も意味のあるものである.

命題 1.1 の証明は, 観測データの独立性と正規性をフルに使っている. しかし, (1.2) のような評価は, 観測データが正規でなくても, 裾が十分軽ければ成立し, またデータが確率過程から得られているようなケースにも一般化できることが知られている. 本講義では, そのような結果を, 導出するための数学的技法も含めて解説することを目標とする. あわせて, 関連する統計的応用にもいくつか触れる.

2 Orlicz ノルム

定義 2.1 (Orlicz ノルム)**.** $\psi : [0,\infty) \to [0,\infty)$ を全単射増加関数とする. 確率変数 X に対して, X の **ψ-Orlicz ノルム**を

$$\|X\|_\psi := \inf\{C > 0 : \mathrm{E}[\psi(|X|/C)] \le 1\}$$

で定義する.

例 2.1. $p > 0$ として $\psi(x) = x^p$ と定めると, $\|X\|_\psi = (\mathrm{E}\,|X|^p)^{1/p}$ となることが容易に確認できる. 従って, Orlicz ノルムは L^p-ノルムの一般化となっている. また, この例からわかるように, Orlicz ノルムはノルムの条件を満たすとは限らない (三角不等式を満たさない場合がある). 次の命題で示すように, ψ が凸であれば, Orlicz ノルムは実際にノルムの条件を満たす.

命題 2.1. $\psi : [0,\infty) \to [0,\infty)$ を全単射増加関数, X を確率変数とするとき, 以下が成り立つ.

(a) $\|X\|_\psi = 0 \Leftrightarrow X = 0$ a.s.
(b) $0 < \|X\|_\psi < \infty$ ならば, $\mathrm{E}[\psi(|X|/\|X\|_\psi)] \le 1$.
(c) 任意の $c \in \mathbb{R}$ に対して, $\|cX\|_\psi = |c|\|X\|_\psi$.
(d) ψ が凸ならば, 2 つの確率変数 X, Y に対して $\|X+Y\|_\psi \le \|X\|_\psi + \|Y\|_\psi$.

証明. まず, 条件から特に, $\psi(0) = 0, \psi(x) \to \infty$ $(x \to \infty)$ であり, かつ ψ は連続となることに注意しておく.

(a) $\|X\|_\psi = 0$ ならば, 任意の $\varepsilon > 0$ に対して, ある $C_\varepsilon \in (0,\varepsilon)$ が存在して $\mathrm{E}[\psi(|X|/C_\varepsilon)] \le 1$ となる. ψ は増加関数であるから, これは $\mathrm{E}[\psi(|X|/\varepsilon)] \le 1$ を意味する. $X \ne 0$ ならば $\psi(|X|/\varepsilon) \uparrow \infty$ $(\varepsilon \downarrow 0)$ となることに注意すると, $\mathrm{P}(X \ne 0) > 0$ ならば, 単調収束定理より $\mathrm{E}[\psi(|X|/\varepsilon)] \uparrow \infty$ $(\varepsilon \downarrow 0)$ とならなければならず, これは矛盾である. 従って $X = 0$ a.s. である.

逆に $X = 0$ a.s. ならば, $\psi(0) = 0$ より任意の $C > 0$ に対して $\mathrm{E}[\psi(|X|/C)] \le 1$ が成り立つから, $\|X\|_\psi = 0$ である.

(b) 任意の $B > \|X\|_\psi$ に対して, 上と同様の議論によって $\mathrm{E}[\psi(|X|/B)] \le 1$ が成り立つことがわかる. ψ が連続であることに注意すると, $B \downarrow \|X\|_\psi$ とすれば, 単調収束定理より示すべき結論が得られる.

(c) $\|X\|_\psi = 0$ または $c = 0$ の場合は (a) から従う. $c \ne 0$ かつ $\|X\|_\psi = \infty$ の場合は $\mathrm{E}[\psi(|X|/C)] \le 1$ となる $C > 0$ が存在しないため, $\mathrm{E}[\psi(|cX|/C)] \le 1$ となる $C > 0$ も存在せず, 従って $\|cX\|_\psi = \infty = |c|\|X\|_\psi$ となって成り立つ. 従って $c \ne 0$ かつ $0 < \|X\|_\psi < \infty$ の場合に示せばよい. この場合, (b) より $\mathrm{E}[\psi(|cX|/|c|\|X\|_\psi)] \le 1$ および $\mathrm{E}[\psi(|X|/(\|cX\|_\psi/|c|)] \le 1$ が成り立つから, $\|cX\|_\psi = |c|\|X\|_\psi$ となる.

(d) $\|X\|_\psi + \|Y\|_\psi < \infty$ かつ $\|X\|_\psi \wedge \|Y\|_\psi > 0$ の場合のみ考えれば十分である. ψ が増加関数であることから

$$\mathrm{E}[\psi(|X+Y|/(\|X\|_\psi + \|Y\|_\psi))] \le \mathrm{E}[\psi((|X| + |Y|)/(\|X\|_\psi + \|Y\|_\psi))]$$

が成り立つ. ここで, $t = \|X\|_\psi/(\|X\|_\psi + \|Y\|_\psi)$ とおくと,

$$(|X| + |Y|)/(\|X\|_\psi + \|Y\|_\psi) = t|X|/\|X\|_\psi + (1-t)|Y|/\|Y\|_\psi$$

と書けるから, ψ の凸性と (b) より

$$\mathrm{E}[\psi(|X+Y|/(\|X\|_\psi + \|Y\|_\psi))] \le t\,\mathrm{E}[\psi(|X|/\|X\|_\psi)] + (1-t)\,\mathrm{E}[\psi(|Y|/\|Y\|_\psi)] \le 1$$

が成り立つ. 故に $\|X+Y\|_\psi \le \|X\|_\psi + \|Y\|_\psi$ である. □

高次元統計において重要なのは, $\alpha > 0$ に対して定まる関数 $\psi_\alpha(x) = \exp(x^\alpha) - 1$ $(x \ge 0)$ に対応する ψ_α-Orlicz ノルムである. これは次の命題による.

命題 2.2 (cf. van der Vaart & Wellner (1996), Lemma 2.2.2)**.** ある普遍定数 $C > 0$ が存在して次が成り立つ: $X_1, \dots, X_d$ を確率変数, $\alpha > 0$ とすると,

$$\left\| \max_{1 \le j \le d} |X_j| \right\|_{\psi_\alpha} \le \{C \log(3d)\}^{1/\alpha} \max_{1 \le j \le d} \|X_j\|_{\psi_\alpha}. \tag{2.1}$$

証明. 確率変数 ξ に対して $\|\xi\|_{\psi_\alpha} = \||\xi|^\alpha\|_{\psi_1}^{1/\alpha}$ が成り立つことに注意すると, $\alpha = 1$ の場合に (2.1) を満たすような普遍定数 C が存在することを示せば十分である. また, $0 < \max_{1 \le j \le d} \|X_j\|_{\psi_1} < \infty$ の場合のみ考えれば十分である.

$Y := \max_{1 \le j \le d} |X_j|$ とおく. 任意の定数 $K, c > 0$ に対して

$$\exp\left(\frac{Y}{K \log(3d)}\right) \le \exp\left(\frac{Y}{K\log(3d)} 1_{\left\{\frac{Y}{K\log(3d)} \ge c\right\}}\right) + e^c$$

が成り立つ. ここで, $a \ge 0$ と $b \ge 1$ が $a/b \ge c$ を満たすならば,

$$a + b^2 \le ab + \frac{ab}{c} = \left(1 + \frac{1}{c}\right) ab, \qquad \text{従って, } \frac{a}{b} \le \left(1 + \frac{1}{c}\right) a - b$$

が成り立つ. この不等式を $a = Y/K, b = \log(3d)$ として適用すれば,

$$\frac{Y}{K\log(3d)} 1_{\left\{\frac{Y}{K\log(3d)} \ge c\right\}} \le \left(1 + \frac{1}{c}\right) \frac{Y}{K} - \log(3d)$$

を得る. 従って,

$$\begin{aligned}
\mathrm{E}\left[\exp\left(\frac{Y}{K\log(3d)} 1_{\left\{\frac{Y}{K\log(3d)} \ge c\right\}}\right)\right] &\le \frac{1}{3d} \mathrm{E}\left[\exp\left(\left(1 + \frac{1}{c}\right)\frac{Y}{K}\right)\right] \\
&\le \frac{1}{3d} \sum_{j=1}^d \mathrm{E}\left[\exp\left(\left(1 + \frac{1}{c}\right)\frac{|X_j|}{K}\right)\right]
\end{aligned}$$

が成り立つ. 故に, $K = (1 + 1/c)\max_{1 \le j \le d} \|X_j\|_{\psi_1}, c = \log(4/3)$ とすれば, 命題 2.1(b) より

$$\mathrm{E}\left[\exp\left(\frac{Y}{K\log(3d)}\right)\right] \le \frac{2}{3} + \frac{4}{3} = 2$$

を得る. 故に, $C = 1 + 1/c$ ととれば (2.1) が成立する. □

確率変数 X がある $\alpha > 0$ に対して $0 < \|X\|_{\psi_\alpha} < \infty$ を満たすならば, 命題 2.1(b) と Markov の不等式より, 任意の $u \geq 0$ に対して

$$\mathrm{P}(|X| \geq u) \leq 2\exp\left(-\frac{u^\alpha}{\|X\|_{\psi_\alpha}^\alpha}\right)$$

が成り立つ. 逆に, 上の形の不等式を満たすような確率変数は有限な ψ_α-Orlicz ノルムを持つ.

命題 2.3 (cf. van der Vaart & Wellner (1996), Lemma 2.2.1)**.** X を確率変数, $\alpha > 0$ とし, ある定数 $C, K > 0$ が存在して任意の $u > 0$ に対して $\mathrm{P}(|X| > u) \leq Ke^{-Cu^\alpha}$ が成り立つとする. このとき,

$$\|X\|_{\psi_\alpha} \leq ((1+K)/C)^{1/\alpha}.$$

以下のよく知られた結果を証明で使うため, 思い出しておく.

補題 2.1. X を非負確率変数とする. また, $f : [0,\infty) \to [0,\infty)$ を以下の 3 つの条件を満たす関数とする:

(i) f は非減少である.

(ii) 任意の $T > 0$ に対して, f は $[0, T]$ 上絶対連続である.

(iii) $f(0) = 0$.

このとき,

$$\mathrm{E}[f(X)] = \int_0^\infty f'(u)\,\mathrm{P}(X > u)du.$$

証明. (Rudin, 1987, Theorem 8.16) 参照. □

命題 2.3 の証明. $\|X\|_{\psi_\alpha} = \||X|^\alpha\|_{\psi_1}^{1/\alpha}$ が成り立つから, $\alpha = 1$ の場合に示せば十分である. 補題 2.1 と仮定より, $B = (1+K)/C$ に対して,

$$\begin{aligned}
\mathrm{E}[\psi_1(|X|/B)] &= B^{-1}\int_0^\infty e^{u/B}\,\mathrm{P}(|X| > u)du \\
&\leq \frac{CK}{1+K}\int_0^\infty e^{-CKu/(1+K)}du = \left[-e^{-CKu/(1+K)}\right]_0^\infty = 1
\end{aligned}$$

となることから結論が従う. □

ψ_α-Orlicz ノルムと L^p-ノルムの p に関する増大度には密接な関係がある.

命題 2.4. X を確率変数, $\alpha > 0$ とする. α のみに依存する定数 $C_\alpha > 0$ が存在して, 任意の $p \geq 1$ に対して

$$\|X\|_p := (\mathrm{E}\,|X|^p)^{1/p} \leq C_\alpha p^{1/\alpha}\|X\|_{\psi_\alpha} \tag{2.2}$$

が成り立つ.

証明にはガンマ関数の Stirling 近似を用いるため, 思い出しておく.

補題 2.2 (Stirling 近似)**.** Γ をガンマ関数とすると, 任意の $p \geq 1$ に対して

$$\sqrt{2\pi} p^{p-\frac{1}{2}} e^{-p} < \Gamma(p) < \sqrt{2\pi} p^{p-\frac{1}{2}} e^{-p+\frac{1}{12}}$$

が成り立つ.

証明. 高木 (1983) の 69 節参照. □

命題 2.4 の証明. まず $\alpha = 1$ の場合に成り立つことを示す. $0 < \|X\|_{\psi_1} < \infty$ の場合に示せば十分である. 補題 2.1 と Markov の不等式より

$$\mathrm{E}[|X|^p] \leq 2p \int_0^\infty u^{p-1} e^{-u/\|X\|_{\psi_1}} du = 2p \|X\|_{\psi_1}^p \Gamma(p) \tag{2.3}$$

が成り立つから, Stirling 近似と不等式 $p < e^p$ より

$$\mathrm{E}[|X|^p] \leq 2\sqrt{2\pi} p^p e^{\frac{1}{12}} \|X\|_{\psi_1}^p$$

が成り立つ. 従って, $p \geq 1$ に注意すると,

$$\|X\|_p \leq 2\sqrt{2\pi} e^{\frac{1}{12}} p \|X\|_{\psi_1}$$

となるから, $C = 2\sqrt{2\pi} e^{\frac{1}{12}}$ ととれば (2.2) は成立する.

次に $\alpha \neq 1$ を考える. $\alpha < 1$ の場合, $p/\alpha > 1$ だから,

$$\|X\|_p = \||X|^\alpha\|_{p/\alpha}^{1/\alpha} \leq (C(p/\alpha)\||X|^\alpha\|_{\psi_1})^{1/\alpha} = (C/\alpha)^{1/\alpha} p^{1/\alpha} \|X\|_{\psi_\alpha}$$

となる. $\alpha > 1$ の場合は, $\alpha p > p$ だから, Lyapunov の不等式より

$$\|X\|_p \leq \|X\|_{\alpha p} = \||X|^\alpha\|_p^{1/\alpha} \leq (Cp\||X|^\alpha\|_{\psi_1})^{1/\alpha} = (Cp)^{1/\alpha} \|X\|_{\psi_\alpha}$$

となる. □

応用上特に重要なのは, $\alpha = 1, 2$ の場合であり, それぞれ名前がついている.

定義 2.2 (劣 Gauss 性・劣指数性)**.** X を確率変数とする.

(a) $\|X\|_{\psi_2} < \infty$ のとき, X は**劣 Gauss 型**であるという.
(b) $\|X\|_{\psi_1} < \infty$ のとき, X は**劣指数型**であるという.

例 2.2. X を平均 0, 標準偏差 $\sigma > 0$ の正規分布に従う確率変数とする. このとき, X^2/σ^2 が自由度 1 の χ^2 分布に従うことに注意すると, $\|X\|_{\psi_2} = \sigma\sqrt{8/3}$ であることがわかる. 特に, X は劣 Gauss 型である.

劣 Gauss 型確率変数の積は劣指数型になる. より一般に次の不等式が成立する.

命題 2.5 (ψ_α-Orlicz ノルムに対する Hölder の不等式)**.** X, Y を 2 つの確率変数とする．定数 $\alpha, \beta, \gamma > 0$ が $1/\alpha + 1/\beta = 1/\gamma$ を満たすならば，

$$\|XY\|_{\psi_\gamma} \le \|X\|_{\psi_\alpha} \|Y\|_{\psi_\beta}$$

が成り立つ．

証明には次の初等的な不等式を用いる．

補題 2.3 (積に対する Young の不等式)**.** $a, b \ge 0$ と $p, q > 1$ を定数とする. $1/p + 1/q = 1$ ならば，

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}$$

が成り立つ．

証明. a, b ともに正の場合のみ考えれば十分である. 対数関数が凹関数であることに注意すると，

$$\log\left(\frac{a^p}{p} + \frac{b^q}{q}\right) \ge \frac{1}{p}\log a^p + \frac{1}{q}\log b^q = \log(ab)$$

が成り立つ. 両辺の指数をとって示すべき不等式を得る. □

命題 2.5 の証明. $0 < \|X\|_{\psi_\alpha} < \infty$ かつ $0 < \|Y\|_{\psi_\beta} < \infty$ の場合のみ考えればよい. $p = \alpha/\gamma, q = \beta/\gamma$ とおく. 仮定より $p, q > 1$ および $1/p + 1/q = 1$ が成り立つ. 従って, 積に対する Young の不等式より

$$\left(\frac{|XY|}{\|X\|_{\psi_\alpha}\|Y\|_{\psi_\beta}}\right)^\gamma \le \frac{1}{p}\frac{|X|^{p\gamma}}{\|X\|_{\psi_\alpha}^{p\gamma}} + \frac{1}{q}\frac{|Y|^{q\gamma}}{\|Y\|_{\psi_\beta}^{q\gamma}} = \frac{1}{p}\frac{|X|^{\alpha}}{\|X\|_{\psi_\alpha}^{\alpha}} + \frac{1}{q}\frac{|Y|^{\beta}}{\|Y\|_{\psi_\beta}^{\beta}}$$

が成り立つ. よって, Hölder の不等式と命題 2.1(b) より

$$\begin{aligned} \mathrm{E}\left[\exp\left(\left(\frac{|XY|}{\|X\|_{\psi_\alpha}\|Y\|_{\psi_\beta}}\right)^\gamma\right)\right] &\le \left(\mathrm{E}\left[\exp\left(\frac{|X|^\alpha}{\|X\|_{\psi_\alpha}^\alpha}\right)\right]\right)^{1/p}\left(\mathrm{E}\left[\exp\left(\frac{|Y|^\beta}{\|Y\|_{\psi_\beta}^\beta}\right)\right]\right)^{1/q} \\ &\le 2^{1/p}2^{1/q} = 2 \end{aligned}$$

が成り立つ. これは示すべき不等式を意味する. □

命題 2.5 から特に, ψ_α-Orlicz ノルムの有限性は α が大きくなるほど強い条件となることがわかる. より正確には次の結果が成り立つ.

系 2.1. X を確率変数とする. $0 < \alpha \le \beta < \infty$ に対して,

$$\|X\|_{\psi_\alpha} \le (\log 2)^{1/\beta - 1/\alpha}\|X\|_{\psi_\beta}$$

が成り立つ.

証明. $\alpha = \beta$ の場合は明らかだから, $\alpha < \beta$ とする. $\gamma > 0$ を $1/\beta + 1/\gamma = 1/\alpha$ を満たすように定めると, 命題 2.5 より

$$\|X\|_{\psi_\alpha} \leq \|X\|_{\psi_\beta} \|1\|_{\psi_\gamma}$$

が成り立つ. $\|1\|_{\psi_\gamma} = (\log 2)^{-1/\gamma} = (\log 2)^{1/\beta - 1/\alpha}$ であることが容易に確認できるから, 示すべき不等式が得られる. □

3 Bernstein の不等式

この節の目標は, 劣 Gauss 型の独立標本に対して, 命題 1.1 に対応する結果を示すことである. 命題 2.5 より, 劣 Gauss 型の独立標本から計算された標本共分散行列は, 独立な劣指数型確率変数の和となることがわかる. 従って, 一般に, 成分が独立な劣指数型確率変数の和からなる確率ベクトルの最大値ノルムを評価する問題を考えればよいことになる. まずこの節では, この問題を解くのに有用な道具である Bernstein の不等式を示すことから始める.

後に確率過程の場合に拡張する際にマルチンゲールに対する Bernstein の不等式が必要となるため, その場合で証明しておく. 念のためにマルチンゲール差分列の定義を思い出しておく.

定義 3.1 (マルチンゲール差分列)**.** $(\mathcal{G}_i)_{i=0}^n$ を $\mathcal{F}$ のフィルトレーションとする. すなわち, $(\mathcal{G}_i)_{i=0}^n$ は $\mathcal{F}$ の部分 σ-加法族の列で $\mathcal{G}_0 \subset \mathcal{G}_1 \subset \cdots \mathcal{G}_n$ を満たすものとする. 確率変数列 $(X_i)_{i=1}^n$ が $(\mathcal{G}_i)_{i=0}^n$ に関する**マルチンゲール差分列**であるとは, 任意の $i = 1, \dots, n$ に対して, X_i は $\mathcal{G}_i$-可測かつ可積分であり, かつ $\mathrm{E}[X_i \mid \mathcal{G}_{i-1}] = 0$ が成り立つことをいう.

注意 3.1. $(X_i)_{i=1}^n$ がフィルトレーション $(\mathcal{G}_i)_{i=0}^n$ に関するマルチンゲール差分列であるとは, $X_0 := 0$ として $Y_j := \sum_{i=0}^j X_i$ $(j = 0, 1, \dots, n)$ と定めたとき, $(Y_j)_{j=0}^n$ が $(\mathcal{G}_i)_{i=0}^n$ に関するマルチンゲールとなることと同値である.

例 3.1. $X_1, \dots, X_n$ を平均 0 の独立確率変数列とする. $\mathcal{G}_0 := \{\emptyset, \Omega\}, \mathcal{G}_i := \sigma(X_1, \dots, X_i)$ $(i = 1, \dots, n)$ とおくと, $(X_i)_{i=1}^n$ はフィルトレーション $(\mathcal{G}_i)_{i=0}^n$ に関するマルチンゲール差分列となる.

定理 3.1 (マルチンゲールに対する Bernstein の不等式)**.** $(X_i)_{i=1}^n$ をフィルトレーション $(\mathcal{G}_i)_{i=0}^n$ に関するマルチンゲール差分列, $b > 0$ を定数とする. また, 各 $i = 1, \dots, n$ に対して, ある定数 $\sigma_i > 0$ が存在して, 任意の整数 $p \geq 2$ に対して

$$\mathrm{E}[|X_i|^p \mid \mathcal{G}_i] \leq \frac{p!}{2} \sigma_i^2 b^{p-2} \tag{3.1}$$

が成り立つとする. このとき, $\sigma^2 := \sum_{i=1}^n \sigma_i^2$ とおくと, 任意の $u \geq 0$ に対して

$$\mathrm{P}\left(\left| \sum_{i=1}^n X_i \right| \geq u \right) \leq 2 \exp\left(-\frac{u^2}{2(\sigma^2 + bu)} \right)$$

が成り立つ.

定理 3.1 の証明のために 2 つ補題を準備する.

補題 3.1. X を確率変数, $b \geq 0, v > 0$ を定数とする. 任意の $\lambda \in [0, 1/b)$ に対して[*1]

$$\mathrm{E}\left[e^{\lambda X}\right] \leq \exp\left(\frac{v\lambda^2}{2(1-b\lambda)}\right) \tag{3.2}$$

が成り立つならば, 任意の $u \geq 0$ に対して

$$\mathrm{P}(X \geq u) \leq \exp\left(-\frac{u^2}{2(v+bu)}\right)$$

が成り立つ.

証明. $\lambda := u/(v+bu) \in [0, 1/b)$ に注意すると, Markov の不等式と (3.2) より

$$\mathrm{P}(X \geq u) = \mathrm{P}(e^{\lambda X} \geq e^{\lambda u}) \leq e^{-\lambda u}\,\mathrm{E}[e^{\lambda X}] \leq \exp\left(-\lambda u + \frac{v\lambda^2}{2(1-b\lambda)}\right)$$

を得る.

$$\frac{v\lambda^2}{2(1-b\lambda)} = \frac{\lambda^2}{2}(v+bu) = \frac{\lambda}{2}u$$

であるから,

$$-\lambda u + \frac{v\lambda^2}{2(1-b\lambda)} = -\frac{\lambda}{2}u = -\frac{u^2}{2(v+bu)}$$

となるので, 示すべき不等式を得る. □

補題 3.2. X を可積分な確率変数, $\mathcal{G}$ を $\mathcal{F}$ の部分 σ-加法族とする. $\mathrm{E}[X \mid \mathcal{G}] = 0$ であり, かつ, ある定数 $b, \sigma > 0$ が存在して, 任意の整数 $p \geq 2$ に対して

$$\mathrm{E}[|X|^p \mid \mathcal{G}] \leq \frac{p!}{2}\sigma^2 b^{p-2} \tag{3.3}$$

が成り立つならば, $|\lambda| < 1/b$ なる任意の $\lambda \in \mathbb{R}$ に対して

$$\mathrm{E}\left[e^{\lambda X} \mid \mathcal{G}\right] \leq \exp\left(\frac{\sigma^2\lambda^2}{2(1-b|\lambda|)}\right)$$

が成り立つ.

証明. 条件 (3.3) と $|b\lambda| < 1$ より,

$$\mathrm{E}\left[\left|\sum_{p=2}^{\infty}\lambda^p\frac{X^p}{p!}\right| \mid \mathcal{G}\right] \leq \sum_{p=2}^{\infty}|\lambda|^p\frac{\mathrm{E}[|X|^p \mid \mathcal{G}]}{p!} \leq \frac{\sigma^2\lambda^2}{2}\sum_{p=2}^{\infty}|b\lambda|^{p-2} = \frac{\sigma^2\lambda^2}{2(1-b|\lambda|)} < \infty$$

が成り立つ. 従って, 指数関数の Taylor 展開と $\mathrm{E}[X \mid \mathcal{G}] = 0$ より,

$$\mathrm{E}[e^{\lambda X} \mid \mathcal{G}] \leq 1 + \frac{\sigma^2\lambda^2}{2(1-b|\lambda|)}$$

を得る. 任意の $u \geq 0$ に対して成り立つ不等式 $1 + u \leq e^u$ より示すべき不等式を得る. □

[*1] $b = 0$ の場合は $1/b = \infty$ と解釈する.

定理 3.1 の証明. $|\lambda| < 1/b$ なる $\lambda \in \mathbb{R}$ を任意にとる. 補題 3.2 より, 各 $i = 1, \dots, n$ について

$$\mathrm{E}\left[e^{\lambda X_i} \mid \mathcal{G}_{i-1}\right] \leq \exp\left(\frac{\sigma_i^2 \lambda^2}{2(1-b|\lambda|)}\right)$$

が成り立つ. 従って,

$$\begin{aligned}
\mathrm{E}\left[\exp\left(\lambda \sum_{i=1}^n X_i\right)\right] &= \mathrm{E}\left[\exp\left(\lambda \sum_{i=1}^{n-1} X_i\right) \mathrm{E}[e^{\lambda X_n} \mid \mathcal{G}_{n-1}]\right] \\
&\leq \mathrm{E}\left[\exp\left(\lambda \sum_{i=1}^{n-1} X_i\right)\right] \exp\left(\frac{\sigma_n^2 \lambda^2}{2(1-b|\lambda|)}\right) \\
&\leq \cdots \leq \prod_{i=1}^n \exp\left(\frac{\sigma_i^2 \lambda^2}{2(1-b|\lambda|)}\right) = \exp\left(\frac{\sigma^2 \lambda^2}{2(1-b|\lambda|)}\right)
\end{aligned}$$

を得る. 故に, 補題 3.1 を $\sum_{i=1}^n X_i$ と $-\sum_{i=1}^n X_i$ のそれぞれに適用することで, 示すべき不等式を得る. □

次の補題は Bernstein の不等式の最大値ノルム評価における有用性を示している.

補題 3.3 (van der Vaart & Wellner (1996), Lemma 2.2.10)**.** $X_1, \dots, X_d$ を確率変数列, $b, \sigma > 0$ を定数とし, 任意の $u > 0$ と $j = 1, \dots, d$ に対して

$$\mathrm{P}(|X_j| > u) \leq 2\exp\left(-\frac{u^2}{2(\sigma^2 + bu)}\right)$$

が成り立つとする. このとき, ある普遍定数 C が存在して,

$$\left\| \max_{1 \leq j \leq d} |X_j| \right\|_{\psi_1} \leq C\left(\sigma\sqrt{\log(3d)} + b\log(3d)\right).$$

証明. 各 $j = 1, \dots, d$ について, 仮定より, $0 < u \leq \sigma^2/b$ ならば $\mathrm{P}(|X_j| > u) \leq 2e^{-u^2/(4\sigma^2)}$ となり, $u > \sigma^2/b$ ならば $\mathrm{P}(|X_j| > u) \leq 2e^{-u/(4b)}$ となる. 従って, 任意の $u > 0$ に対して $\mathrm{P}(|X_j| 1_{\{|X_j| \leq \sigma^2/b\}} > u) \leq 2e^{-u^2/(4\sigma^2)}$ かつ $\mathrm{P}(|X_j| 1_{\{|X_j| > \sigma^2/b\}} > u) \leq 2e^{-u/(4b)}$ が成り立つ. よって, 命題 2.3 より

$$\left\| |X_j| 1_{\{|X_j| \leq \sigma^2/b\}} \right\|_{\psi_2} \leq \sqrt{12}\sigma \qquad \text{かつ} \qquad \left\| |X_j| 1_{\{|X_j| > \sigma^2/b\}} \right\|_{\psi_1} \leq 12b$$

が成り立つ. 従って, 命題 2.2 より, ある普遍定数 $C_1 > 0$ が存在して

$$\begin{aligned}
\left\| \max_{1 \leq j \leq d} |X_j| 1_{\{|X_j| \leq \sigma^2/b\}} \right\|_{\psi_2} &\leq C_1 \sigma \sqrt{\log(3d)}, \\
\left\| \max_{1 \leq j \leq d} |X_j| 1_{\{|X_j| > \sigma^2/b\}} \right\|_{\psi_1} &\leq C_1 b \log(3d)
\end{aligned}$$

が成り立つから, 系 2.1 と ψ_1-Orlicz ノルムに対する三角不等式より示すべき不等式を得る. □

Bernstein の不等式を独立な劣指数型確率変数の和に対して適用すると, 以下の評価を得る.

命題 3.1. $X_1, \dots, X_n$ を独立な確率変数列で $K := \max_{1 \le i \le n} \|X_i\|_{\psi_1} \in (0, \infty)$ を満たすものとする. このとき, $S := \sum_{i=1}^n \|X_i\|_{\psi_1}^2$ とおくと, 任意の $u \ge 0$ に対して,

$$\mathrm{P}\left(\left| \sum_{i=1}^n (X_i - \mathrm{E}[X_i]) \right| \ge u \right) \le 2 \exp\left(-\frac{u^2}{2(16S + 2Ku)} \right).$$

証明. 各 $i = 1, \dots, n$ と $p \ge 2$ に対して, (2.3) より $\mathrm{E}[|X_i|^p] \le 2p! \|X_i\|_{\psi_1}^p$ が成り立つ. 従って, Jensen の不等式より

$$\mathrm{E}[|X_i - \mathrm{E}[X_i]|^p] \le 2^p \, \mathrm{E}[|X_i|^p] \le 2^{p+1} p! \|X_i\|_{\psi_1}^p \le \frac{p!}{2} (4\|X_i\|_{\psi_1})^2 (2K)^{p-2}$$

が成り立つ. 故に, Bernstein の不等式から示すべき不等式が従う. □

3.1 劣 Gauss 型独立標本の共分散行列の推定

$X_1, \dots, X_n$ を独立同分布な d 次元確率変数列とし, 平均 0, 共分散行列 Σ をもち, かつ $K := \max_{1 \le j \le d} \|X_{1j}\|_{\psi_2} < \infty$ を満たすとする. 命題 1.1 と同様に, 中心化していない標本共分散行列

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top$$

に対して, $|\widehat{\Sigma}_n - \Sigma|_\infty$ を評価する問題について考察する.

命題 3.2. ある普遍定数 C が存在して次の不等式が成り立つ:

$$\left\| |\widehat{\Sigma}_n - \Sigma|_\infty \right\|_{\psi_1} \le CK^2 \left(\sqrt{\frac{\log(3d^2)}{n}} + \frac{\log(3d^2)}{n} \right).$$

証明. 各 $j, k \in \{1, \dots, d\}$ について,

$$\widehat{\Sigma}_{n,jk} - \Sigma_{jk} = \frac{1}{n} \sum_{i=1}^n (X_{ij} X_{ik} - \mathrm{E}[X_{ij} X_{ik}])$$

と書けることと, 命題 2.5 より $\|X_{ij} X_{ik}\|_{\psi_1} \le K^2$ が成り立つことに注意すると, 命題 3.1 より, 任意の $u \ge 0$ に対して

$$\mathrm{P}\left(\left| \widehat{\Sigma}_{n,jk} - \Sigma_{jk} \right| \ge u \right) \le 2 \exp\left(-\frac{u^2}{2(16K^4/n + (2K^2/n)u)} \right)$$

が成り立つ. 従って, 補題 3.3 より, ある普遍定数 C が存在して

$$\left\| |\widehat{\Sigma}_n - \Sigma|_\infty \right\|_{\psi_1} \le C \left(\frac{K^2 \sqrt{\log(3d^2)}}{\sqrt{n}} + \frac{K^2 \log(3d^2)}{n} \right)$$

が成り立つ. 従って示すべき不等式が得られた. □

命題 3.2 と 2.4 から特に, ある普遍定数 C が存在して

$$\mathrm{E}\left[\left|\widehat{\Sigma}_n-\Sigma\right|_\infty^2\right]\leq CK^4\left(\frac{\log(3d^2)}{n}+\left(\frac{\log(3d^2)}{n}\right)^2\right)$$

が成り立つことがわかる. 従って, K が n に依存しない定数で上から抑えられる場合, $n\to\infty$ のとき $|\widehat{\Sigma}_n-\Sigma|_\infty$ が 0 に L^2-収束するためには, $\log d=o(n)$ という条件があれば十分である. X_i が正規分布に従う場合は $K=\sqrt{(8/3)|\Sigma|_\infty}$ となることが例 2.2 より従うので, 上の結果は (定数 C の具体的な値を求めていないことを除けば) 命題 1.1 の評価を改善している.

4 実現共分散行列: 低次元の場合

4.1 連続時間マルチンゲールに関する基本事項の復習

この節では, 以下の議論で必要となる確率解析における用語や結果を列挙する. 詳細は専門書を参照のこと. 和書では例えば長井 (1999) を参照するとよい.

■ 連続時間確率過程に関する基本事項

以下この小節では, Ξ は位相空間を表す.

定義 4.1 (確率過程)**.** Ξ 値確率変数の族 $X=(X_t)_{t\in[0,\infty)}$ を ($[0,\infty)$ で添字づけられた) Ξ **値 (確率) 過程**と呼ぶ. $\Xi=\mathbb{R}^d$ の場合は d 次元 (確率) 過程と呼び, 1 次元確率過程は単に (確率) 過程と呼ぶことにする.

$X=(X_t)_{t\geq0}$ が d 次元確率過程のとき, 各 $i=1,\dots,d$ と $t\geq0$ について X_t の第 i 成分で与えられる確率変数を X_t^i と書くことにし, $X^i:=(X_t^i)_{t\geq0}$ と定める. すなわち, X^i は X の「第 i 成分」から定まる確率過程である. $X^1,\dots,X^d$ を総称して X の**成分過程**と呼ぶことにする.

$H=(H_s)_{s\geq0}$ が $\mathbb{R}^{d\times r}$ 値確率過程のとき, $j=1,\dots,d, k=1,\dots,d$ と $s\geq0$ に対して, H_s の (j,k) 成分で与えられる確率変数を H_s^{jk} で表すことにする.

2 つの Ξ 値確率過程 $X=(X_t)_{t\in[0,\infty)}$ および $Y=(Y_t)_{t\in[0,\infty)}$ が**区別できない**とは, ある P-ゼロ集合 N が存在して, 任意の $\omega\in\Omega\setminus N$ と $t\geq0$ に対して $X_t(\omega)=Y_t(\omega)$ が成り立つことをいう. 以下, 特に断らない限り, 区別できない Ξ 値確率過程どうしは同一視して考える.

定義 4.2 (フィルトレーション)**.** 族 $\mathbf{F}=(\mathcal{F}_t)_{t\geq0}$ が $\mathcal{F}$ の**フィルトレーション**であるとは, 各 $\mathcal{F}_t$ が $\mathcal{F}$ の部分 σ-加法族であり, かつ $0\leq s<t$ ならば $\mathcal{F}_s\subset\mathcal{F}_t$ を満たすことをいう. このとき, 順序対 $(\Omega,\mathcal{F},\mathbf{F},\mathrm{P})$ を**確率基底**または**フィルター付き確率空間**と呼ぶ.

以下特に断らない限り, 確率基底 $\mathcal{B}=(\Omega,\mathcal{F},\mathbf{F},\mathrm{P})$ を 1 つ固定して考える. さらに, $\mathcal{B}$ は以下の 2 条件 (いわゆる「通常の仮定」) を満たすと仮定する:

(i) (右連続性) 任意の $t\geq0$ に対して, $\bigcap_{s>t}\mathcal{F}_s=\mathcal{F}_t$.
(ii) (完備性) $N\subset\Omega$ がある P-ゼロ集合に含まれるならば, $N\in\mathcal{F}_0$ が成り立つ.

定義 4.3 (適合過程)**.** Ξ 値確率過程 $X = (X_t)_{t\geq 0}$ が $\mathbf{F}$ に**適合している**とは, 任意の $t \geq 0$ に対して X_t が $\mathcal{F}_t$-可測となることをいう.

定義 4.4 (連続過程)**.** Ξ 値確率過程 $X = (X_t)_{t\geq 0}$ が**連続**であるとは, ほとんどすべての $\omega \in \Omega$ について, 関数 $[0,\infty) \ni t \mapsto X_t(\omega) \in \Xi$ が連続となることをいう.

定義 4.5 (連続増加過程)**.** 確率過程 $A = (A_t)_{t\geq 0}$ が ($\mathcal{B}$ 上の) **連続増加過程**であるとは, A が以下の 2 条件を満たすことをいう:

(i) A は $\mathbf{F}$ に適合している.
(ii) ほとんどすべての $\omega \in \Omega$ について, 関数 $[0,\infty) \ni t \mapsto A_t(\omega) \in \mathbb{R}$ が連続かつ非減少である.

定義 4.6 (連続有界変動過程)**.** $A = (A_t)_{t\geq 0}$ を連続確率過程とする. 各 $\omega \in \Omega$ と $t \geq 0$ について, 関数 $[0,\infty) \ni s \mapsto A_s(\omega) \in \mathbb{R}$ の区間 $[0,t]$ における全変動を $\mathrm{TV}(A)_t(\omega)$ で表すことによって $[0,\infty]$ 値確率過程 $\mathrm{TV}(A) = (\mathrm{TV}(A)_t)_{t\geq 0}$ を定義する. これを A の**全変動過程**と呼ぶ.

A が ($\mathcal{B}$ 上の) **連続有界変動過程**であるとは, A が以下の 2 条件を満たすことをいう:

(i) A は連続かつ $\mathbf{F}$ に適合している.
(ii) 任意の $t > 0$ に対して, $\mathrm{TV}(A)_t < \infty$ a.s.

d 次元確率過程 $A = (A_t)_{t\geq 0}$ の成分過程がいずれも連続有界変動過程であるとき, A を d 次元連続有界変動過程と呼ぶことにする.

■ マルチンゲールと局所マルチンゲール

定義 4.7 (連続マルチンゲール)**.** 確率過程 $M = (M_t)_{t\geq 0}$ が ($\mathcal{B}$ 上の) **連続マルチンゲール**であるとは, M が以下の 3 条件を満たすことをいう:

(i) M は連続であり, かつ $\mathbf{F}$ に適合している
(ii) 任意の $t \geq 0$ に対して $\mathrm{E}[|M_t|] < \infty$.
(iii) 任意の $0 \leq s < t$ に対して $\mathrm{E}[M_t \mid \mathcal{F}_s] = M_s$ a.s.

d 次元確率過程 $M = (M_t)_{t\geq 0}$ の成分過程がいずれも連続マルチンゲールであるとき, M を d 次元連続マルチンゲールと呼ぶことにする.

定義 4.8 (停止時刻)**.** 関数 $\tau : \Omega \to [0,\infty]$ が ($\mathbf{F}$-) **停止時刻**であるとは, 任意の $t \geq 0$ に対して $\{\tau \leq t\} \in \mathcal{F}_t$ となることをいう.

τ が停止時刻のとき, $F \in \mathcal{F}$ で任意の $t \geq 0$ に対して $F \cap \{\tau \leq t\} \in \mathcal{F}_t$ を満たすようなもの全体の集合を $\mathcal{F}_\tau$ と書く. $\mathcal{F}_\tau$ は Ω 上の σ-加法族となることが知られている. また, 確率過程 $X = (X_t)_{t\geq 0}$ に対して $X^\tau := (X_{\tau\wedge t})$ と定める.

命題 4.1. M が連続マルチンゲール, τ が停止時刻ならば, M^τ も連続マルチンゲールである.

証明. (長井, 1999, 問 1.15) 参照. □

定義 4.9 (局所化列)**.** 停止時刻の列 $(\tau_k)_{k=1}^{\infty}$ で $\tau_1 \leq \tau_2 \leq \cdots$ かつ $\tau_k \to \infty$ $(k \to \infty)$ を満たすものを ($\mathcal{B}$ 上の) **局所化列**と呼ぶ.

定義 4.10 (連続局所マルチンゲール)**.** 確率過程 $M = (M_t)_{t\geq 0}$ が ($\mathcal{B}$ 上の) **連続局所マルチンゲール**であるとは, ある局所化列 $(\tau_k)_{k=1}^{\infty}$ が存在して, すべての $k = 1, 2, \ldots$ について M^{τ_k} が連続マルチンゲールとなることをいう.

d 次元確率過程 $M = (M_t)_{t\geq 0}$ の成分過程がいずれも連続局所マルチンゲールであるとき, M を d 次元連続局所マルチンゲールと呼ぶことにする.

命題 4.2. M が連続局所マルチンゲール, τ が停止時刻ならば, M^{τ} も連続局所マルチンゲールである.

証明. 命題 4.1 から直ちに従う. □

定理 4.1. $M = (M_t)_{t\geq 0}$ を連続局所マルチンゲールで $M_0 = 0$ a.s. を満たすものとする. このとき, 連続増加過程 $A = (A_t)_{t\geq 0}$ で以下の 2 条件を満たすものが (区別できないものは同一視して) ただ一つ存在する:

(i) $A_0 = 0$ a.s.

(ii) $(M_t^2 - A_t)_{t\geq 0}$ は連続局所マルチンゲールである.

この A を M の**可予測 2 次変動過程**と呼び, 記号 $\langle M, M\rangle = (\langle M, M\rangle_t)_{t\geq 0}$ で表す.

証明. (長井, 1999, 系 1.7.6) 参照. □

定義 4.11 (可予測 2 次共変動過程)**.** $M = (M_t)_{t\geq 0}$ および $N = (N_t)_{t\geq 0}$ を連続局所マルチンゲールで $M_0 = N_0 = 0$ a.s. を満たすものとする. M と N の**可予測 2 次共変動過程** $\langle M, N\rangle = (\langle M, N\rangle_t)_{t\geq 0}$ を

$$\langle M, N\rangle_t := \frac{1}{4}(\langle M+N, M+N\rangle_t - \langle M-N, M-N\rangle_t) \qquad (t \geq 0)$$

で定義する.

命題 4.3. L, M, N を 3 つの連続局所マルチンゲール, $\alpha, \beta \in \mathbb{R}$ とすると, $\langle \alpha L + \beta M, N\rangle = \alpha\langle L, N\rangle + \beta\langle M, N\rangle$ が成り立つ.

証明. (長井, 1999, 命題 1.7.8) 参照. □

命題 4.4. M, N を 2 つの連続局所マルチンゲール, τ を停止時刻とすると, $\langle M^{\tau}, N^{\tau}\rangle = \langle M^{\tau}, N\rangle = \langle M, N\rangle^{\tau}$ が成り立つ.

証明. (長井, 1999, 命題 1.7.8) 参照. □

■ 確率積分

定義 4.12 (発展的可測過程)**.** Ξ 値確率過程 $H=(H_t)_{t\geq 0}$ が **(F-) 発展的可測**であるとは, 任意の $t\geq 0$ に対して関数 $[0,t]\times\Omega\ni(s,\omega)\mapsto H_s(\omega)\in\Xi$ が $\mathcal{B}([0,t])\otimes\mathcal{F}_t$-可測となることをいう. ただし, $\mathcal{B}([0,t])$ は $[0,t]$ の Borel σ-加法族を表し, $\mathcal{B}([0,t])\otimes\mathcal{F}_t$ は $\mathcal{B}([0,t])$ と $\mathcal{F}_t$ の直積 σ-加法族を表す.

定理 4.2. $M=(M_t)_{t\geq 0}$ を連続局所マルチンゲール, $H=(H_t)_{t\geq 0}$ を発展的可測過程とし, 任意の $t>0$ に対して

$$\int_0^t H_s^2 d\langle M,M\rangle_s<\infty \qquad \text{a.s.}$$

を満たすとする. このとき, $L_0=0$ なる連続局所マルチンゲール $L=(L_t)_{t\geq 0}$ で, 任意の連続局所マルチンゲール $N=(N_t)_{t\geq 0}$ と $t>0$ に対して

$$\langle L,N\rangle_t=\int_0^t H_s d\langle M,N\rangle_s \qquad \text{a.s.}$$

を満たすようなものがただ一つ存在する.

この L を M に関する H の**確率積分**と呼び, $H\bullet M=(H\bullet M_t)_{t\geq 0}$ で表すことにする. また, 各 $t\geq 0$ について, $H\bullet M_t$ のことを

$$\int_0^t H_s dM_s$$

とも書く (こちらの方が標準的).

証明. (長井, 1999, 問 2.1) 参照. もしくは (風巻, 2003, 定理 3.6.3) を参照. □

命題 4.5. $M=(M_t)_{t\geq 0}, N=(N_t)_{t\geq 0}$ を連続局所マルチンゲール, $H=(H_t)_{t\geq 0}, K=(K_t)_{t\geq 0}$ を発展的可測過程とし, 任意の $t>0$ に対して

$$\int_0^t H_s^2 d\langle M,M\rangle_s+\int_0^t K_s^2 d\langle N,N\rangle_s<\infty \qquad \text{a.s.}$$

が成り立つと仮定する. このとき, 任意の $t\geq 0$ と $\alpha,\beta\in\mathbb{R}$ に対して,

$$\begin{aligned}\int_0^t(\alpha H_s+\beta K_s)dM_s&=\alpha\int_0^t H_s dM_s+\beta\int_0^t K_s dM_s,\\ \int_0^t H_s d(\alpha M+\beta N)_s&=\alpha\int_0^t H_s dM_s+\beta\int_0^t H_s dN_s,\end{aligned}$$

および

$$\langle H\bullet M,K\bullet N\rangle_t=\int_0^t H_sK_s d\langle M,N\rangle_s$$

が成り立つ.

証明. 確率積分の定義と命題 4.3 から直ちに従う. □

命題 4.6. $M = (M_t)_{t \geq 0}$ を連続マルチンゲール, $H = (H_t)_{t \geq 0}$ を発展的可測過程とし, 任意の $t \geq 0$ に対して

$$\mathrm{E}\left[\langle M, M\rangle_t + \int_0^t H_s^2 d\langle M, M\rangle_s\right] < \infty$$

が成り立つと仮定する. このとき, $H \bullet M$ は連続マルチンゲールであり, 任意の $t \geq 0$ に対して

$$\mathrm{E}\left[\left(\int_0^t H_s dM_s\right)^2\right] = \mathrm{E}\left[\int_0^t H_s^2 d\langle M, M\rangle_s\right] \tag{4.1}$$

が成り立つ.

証明. (長井, 1999, 定理 2.2.2) 参照. □

注意 4.1. 等式 (4.1) は**確率積分の等長性 (Itô isometry)** と呼ばれる.

命題 4.7. 定理 4.2 と同じ仮定の下で, 任意の停止時刻 τ に対して $(H \bullet M)^\tau = (H1_{(0,\tau]}) \bullet M$ が成り立つ. すなわち, 任意の $t \geq 0$ に対して

$$\int_0^{t\wedge\tau} H_s dM_s = \int_0^t H_s 1_{(0,\tau]}(s) dM_s.$$

証明. $(H \bullet M)^\tau$ が連続局所マルチンゲールで $(H \bullet M)_0^\tau = 0$ を満たすことは容易に確認できる. さらに, 任意の連続局所マルチンゲール $N = (N_t)_{t \geq 0}$ と $t \geq 0$ に対して, 命題 4.4 と $H \bullet M$ の定義より

$$\langle (H \bullet M)^\tau, N\rangle_t = \langle H \bullet M, N\rangle_{t\wedge\tau} = \int_0^{t\wedge\tau} H_s d\langle M, N\rangle_s = \int_0^t H_s 1_{(0,\tau]}(s) d\langle M, N\rangle_s$$

が成り立つ. 従って, 確率積分の定義より $(H \bullet M)^\tau = (H1_{(0,\tau]}) \bullet M$ である. □

命題 4.8. 定理 4.2 と同じ仮定の下で, 発展的可測過程 $K = (K_t)_{t \geq 0}$ が, 任意の $t > 0$ に対して

$$\int_0^t K_s^2 H_s^2 d\langle M, M\rangle_s < \infty \qquad \text{a.s.}$$

を満たすならば, $K \bullet (H \bullet M) = KH \bullet M$ が成り立つ. すなわち, 任意の $t \geq 0$ に対して,

$$\int_0^t K_s d(H \bullet M)_s = \int_0^t K_s H_s dM_s.$$

証明. 任意の連続局所マルチンゲール N と $t > 0$ に対して, 確率積分の定義より,

$$\langle K \bullet (H \bullet M), N\rangle_t = \int_0^t K_s d\langle H \bullet M, N\rangle_s = \int_0^t K_s H_s d\langle M, N\rangle_s$$

が成り立つことから従う. □

■ 連続セミマルチンゲールと伊藤の公式

定義 4.13 (連続セミマルチンゲール)**.** 確率過程 $X = (X_t)_{t\geq 0}$ が ($\mathcal{B}$ 上の) **連続セミマルチンゲール**であるとは,

$$\begin{cases} X_t = X_0 + A_t + M_t \quad \text{for any } t \geq 0, \\ A_0 = M_0 = 0 \quad \text{a.s.} \end{cases} \tag{4.2}$$

を満たす連続有界変動過程 $A = (A_t)_{t\geq 0}$ および連続局所マルチンゲール $M = (M_t)_{t\geq 0}$ が存在することをいう.

d 次元確率過程 $X = (X_t)_{t\geq 0}$ の成分過程がいずれも連続セミマルチンゲールであるとき, X を d 次元連続セミマルチンゲールと呼ぶことにする.

命題 4.9. $X = (X_t)_{t\geq 0}$ が連続セミマルチンゲールならば, (4.2) を満たす連続有界変動過程 $A = (A_t)_{t\geq 0}$ および連続局所マルチンゲール $M = (M_t)_{t\geq 0}$ の組が (区別できないものは同一視して) ただ一つ存在する. 分解 (4.2) を X の**標準分解**と呼び, M を X の**連続マルチンゲール部分**と呼ぶ.

証明. 存在は定義から従う. 一意性は (長井, 1999, 補題 1.7.1) から従う. □

定義 4.14 (2 次共変動)**.** $X = (X_t)_{t\geq 0}, Y = (Y_t)_{t\geq 0}$ を 2 つの連続セミマルチンゲールとし, その連続マルチンゲール部分をそれぞれ M^X, M^Y とする. このとき, X と Y の **2 次共変動過程** $[X,Y] = ([X,Y]_t)_{t\geq 0}$ を $[X,Y] := \langle M^X, M^Y \rangle$ で定義する.

X が d 次元連続セミマルチンゲール, Y が d' 次元連続セミマルチンゲールのとき, 各 $t \geq 0$ について $[X,Y]_t := ([X^j, Y^k]_t])_{1\leq j\leq d, 1\leq k\leq d'}$ と定義する.

「2 次共変動」という名称は次の結果に由来する.

命題 4.10. $X = (X_t)_{t\geq 0}, Y = (Y_t)_{t\geq 0}$ を 2 つの連続セミマルチンゲールとする. 各 $n \in \mathbb{N}$ について, 実数の増加列 $0 = t_0^n < t_1^n < \cdots$ で $t_i^n \to \infty$ $(i \to \infty)$ を満たすものが与えられているとし, 任意の $t > 0$ に対して

$$\sup_{i\in\mathbb{N}} (t_i^n \wedge t - t_{i-1}^n \wedge t) \to 0 \qquad (n \to \infty)$$

が成り立つと仮定する. このとき, 任意の $T > 0$ に対して, $n \to \infty$ のとき

$$\sup_{t\in[0,T]} \left| \sum_{i=1}^{\infty} (X_{t_i^n \wedge t} - X_{t_{i-1}^n \wedge t})(Y_{t_i^n \wedge t} - Y_{t_{i-1}^n \wedge t}) - [X,Y]_t \right| \to^p 0$$

が成り立つ.

証明. 2 つの確率過程 $U = (U_t)_{t\geq 0}, V = (V_t)_{t\geq 0}$ が与えられたとき, 各 $n \in \mathbb{N}$ と $t \geq 0$ に対して

$$Q_t^n(U,V) := \sum_{i=1}^{\infty} (U_{t_i^n \wedge t} - U_{t_{i-1}^n \wedge t})(V_{t_i^n \wedge t} - V_{t_{i-1}^n \wedge t})$$

と書くことにする. X, Y の標準分解をそれぞれ $X = X_0 + A^X + M^X, Y = Y_0 + A^Y + M^Y$ とすると,

$$Q_t^n(X,Y) = Q_t^n(A^X, A^Y) + Q_t^n(M^X, A^Y) + Q_t^n(A^X, M^Y) + Q_t^n(M^X, M^Y)$$

と分解できる. $[X,Y] = \langle M^X, M^Y \rangle$ であったから, (長井, 1999, 系 1.7.7) より

$$\sup_{0 \le t \le T} |Q_t^n(M^X, M^Y) - [X,Y]_t| \to^p 0 \qquad (n \to \infty)$$

が成り立つ. 一方で, Schwarz の不等式より

$$Q_t^n(M^X, A^Y) \le \sqrt{Q_t^n(M^X, M^X) Q_t^n(A^Y, A^Y)}$$

が成り立ち, また, (長井, 1999, 系 1.7.6) より $\sup_{0 \le t \le T} Q_t^n(M^X, M^X) = O_p(1)$ である. さらに,

$$\sup_{0 \le t \le T} Q_t^n(A^Y, A^Y) \le \sup_{0 \le t \le T} |A^Y_{t_i^n \wedge t} - A^Y_{t_{i-1}^n \wedge t}| \, \mathrm{TV}(A^Y)_T$$

が成り立つが, A^Y が連続有界変動過程であることからこれは $\sup_{0 \le t \le T} Q_t^n(A^Y, A^Y) = o_p(1)$ を意味する. 以上より $\sup_{0 \le t \le T} |Q_t^n(M^X, A^Y)| = o_p(1)$ である. 同様にして $\sup_{0 \le t \le T} |Q_t^n(A^X, A^Y)| = o_p(1)$ および $\sup_{0 \le t \le T} |Q_t^n(A^X, M^Y)| = o_p(1)$ も示せる. 以上をあわせて証明すべき主張を得る. □

注意 4.2. X, Y がともに連続局所マルチンゲールの場合, 定義より $[X,Y] = \langle X, Y \rangle$ が成り立つ. X, Y が連続でない局所マルチンゲールの場合にも $[X,Y]$ と $\langle X, Y \rangle$ は個別に定義されるが, この場合は $[X,Y] = \langle X, Y \rangle$ という関係は一般には成立しない. もう少し詳しく述べると, $\langle X, Y \rangle$ は $[X,Y]$ の「可予測共役射影 (dual predictable projection)」と呼ばれるものになっている. $\langle X, Y \rangle$ が「可予測 2 次共変動過程」と呼ばれるのはこの事実に由来する. この辺りの詳細は (風巻, 2003, 6 章) もしくは (Jacod & Shiryaev, 2003, Chapter I) を参照のこと. 一方で, 命題 4.10 に対応する結果は X, Y が連続でない場合にも成立する. (Jacod & Shiryaev, 2003, Chapter I, Theorem 4.47) 参照.

注意 4.3. 命題 4.10 の結論は $(t_i^n)_{i=0}^\infty$ が停止時刻の列であっても成立することが知られている. (Jacod & Shiryaev, 2003, Chapter I, Proposition 4.44) 参照.

定義 4.15 (連続セミマルチンゲールに関する確率積分)**.** $X = (X_t)_{t \ge 0}$ を標準分解 (4.2) を持つ連続セミマルチンゲールとする. $H = (H_t)_{t \ge 0}$ を発展的可測過程とし, 任意の $t > 0$ に対して

$$\int_0^t |H_s| d\,\mathrm{TV}(A)_s + \int_0^t H_s^2 d\langle M, M \rangle_s < \infty \qquad \text{a.s.}$$

が成り立つとすると, 積分 $\int_0^t H_s dA_s, \int_0^t H_s dM_s$ はそれぞれ Lebesgue–Stieltjes 積分, 確率積分として well-defined である. そこで, X に関する H の**確率積分**を

$$\int_0^t H_s dX_s = \int_0^t H_s dA_s + \int_0^t H_s dM_s \qquad (t \ge 0)$$

で定義する.

定理 4.3 (伊藤の公式)**.** $X = (X_t)_{t \geq 0}$ を d 次元連続セミマルチンゲール, $f : \mathbb{R}^d \to \mathbb{R}$ を C^2 級関数とする. このとき, 任意の $t \geq 0$ に対して

$$f(X_t) = f(X_0) + \sum_{j=1}^{d} \int_0^t \partial_j f(X_s) dX_s^j + \frac{1}{2} \sum_{j,k=1}^{d} \int_0^t \partial_{jk} f(X_s) d[X^j, X^k]_s \quad \text{a.s.}$$

が成り立つ.

証明. (長井, 1999, 定理 2.4.2) 参照. □

系 4.1 (部分積分公式)**.** $X = (X_t)_{t \geq 0}, Y = (Y_t)_{t \geq 0}$ を 2 つの連続セミマルチンゲールとする. 任意の $t \geq 0$ に対して, 確率 1 で

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X, Y]_t$$

が成り立つ.

証明. 2 次元連続セミマルチンゲール $((X_t, Y_t))_{t \geq 0}$ と関数 $(x, y) \mapsto xy$ に対して伊藤の公式を適用すればよい. □

■ Wiener 過程

定義 4.16 (標準 Wiener 過程)**.** r 次元確率過程 $W = (W_t)_{t \geq 0}$ が以下の 3 条件を満たすとき, W を $\mathcal{B}$ 上の r 次元**標準 Wiener 過程**と呼ぶ:

(i) $W_0 = 0$ a.s.

(ii) W は連続であり, かつ **F** に適合している

(iii) 任意の $0 \leq s < t$ に対して, $W_t - W_s$ は $\mathcal{F}_s$ と独立であり, かつ平均 0, 共分散行列 $(t-s)I_r$ の r 次元正規分布に従う.

1 次元標準 Wiener 過程は単に標準 Wiener 過程と呼ぶことにする.

注意 4.4. Wiener 過程は **Brown 運動**とも呼ばれる.

命題 4.11. W を $\mathcal{B}$ 上の r 次元標準 Wiener 過程とする. W は r 次元連続マルチンゲールであり, 任意の $j, k \in \{1, \ldots, r\}$ と $t \geq 0$ に対して $\langle W^j, W^k \rangle_t = t\delta_{jk}$ が成り立つ. ただし, δ_{jk} は Kronecker のデルタを表す.

証明. (長井, 1999, 問 1.13, 問 1.16) 参照. □

$W = (W_t)_{t \geq 0}$ を $\mathcal{B}$ 上の r 次元標準 Wiener 過程, $\sigma = (\sigma_s)_{s \geq 0}$ を $\mathbb{R}^{d \times r}$ 値発展的可測過程とし, 任意の $t > 0$ に対して

$$\int_0^t \|\sigma_s\|_F^2 ds < \infty \quad \text{a.s.}$$

が成り立つとする. このとき, 命題 4.11 に注意すると, 各 $j = 1, \ldots, d$ について, 確率積分

$$\left(\sum_{k=1}^{r} \int_0^t \sigma_s^{jk} dW_s^k \right)_{t \geq 0}$$

を定義することができる. この確率積分を

$$\left(\int_0^t \sigma_s^{j\cdot} dW_s \right)_{t \geq 0}$$

で表すことにする. さらに, d 次元確率過程

$$\left(\left(\int_0^t \sigma_s^{1\cdot} dW_s, \ldots, \int_0^t \sigma_s^{d\cdot} dW_s \right)^\top \right)_{t \geq 0}$$

を

$$\left(\int_0^t \sigma_s dW_s \right)_{t \geq 0}$$

で表すことにする.

■ マルチンゲール不等式

今後の議論で必要となるマルチンゲールのモーメントに関する不等式を述べておく.

定理 4.4 (Doob の不等式)**.** $M = (M_t)_{t \geq 0}$ を連続マルチンゲールとする. 任意の $T > 0$ と $p > 1$ に対して,

$$\left\| \sup_{0 \leq t \leq T} |M_t| \right\|_p \leq \frac{p}{p-1} \|M_T\|_p .$$

証明. (長井, 1999, 定理 1.6.1) 参照. □

定理 4.5 (Burkholder–Davis–Gundy の不等式)**.** $M = (M_t)_{t \geq 0}$ を連続局所マルチンゲールで $M_0 = 0$ を満たすものとする. 任意の $T, p > 0$ に対して, p のみに依存する定数 $C_1(p), C_2(p) > 0$ が存在して

$$C_1(p) \left\| \langle M, M \rangle_T^{1/2} \right\|_p \leq \left\| \sup_{0 \leq t \leq T} |M_t| \right\|_p \leq C_2(p) \left\| \langle M, M \rangle_T^{1/2} \right\|_p .$$

が成り立つ.

証明. (長井, 1999, 定理 2.8.1) 参照. □

4.2　実現共分散行列

$W = (W_t)_{t \geq 0}$ を $\mathcal{B}$ 上の r 次元標準 Wiener 過程とする. 以下の形で与えられる d 次元連続セミマルチンゲール $X = (X_t)_{t \geq 0}$ を考える:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \qquad t \geq 0. \tag{4.3}$$

ここに, $b=(b_s)_{s\geq 0}$ は d 次元発展的可測過程, $\sigma=(\sigma_s)_{s\geq 0}$ は $\mathbb{R}^{d\times r}$ 値発展的可測過程であり, 任意の $t\geq 0$ に対して

$$\int_0^t |b_s|ds+\int_0^t \|\sigma_s\|_F^2 ds<\infty \quad \text{a.s.}$$

を満たすものとする. この条件は (4.3) に現れる積分がすべて well-defined となることを保証するためのものである.

$T>0$ とする. 命題 4.11 より

$$[X,X]_T=\int_0^T c_t dt$$

が成り立つ. ただし, $c_t:=\sigma_t\sigma_t^\top$ と定める. c_t は時点 t における X の瞬間的な変動の共分散行列を表す量だと解釈できるため, $\int_0^T c_t dt$ は区間 $[0,T]$ の各時点における X の瞬間的な変動の共分散行列を累積した量だと解釈される. そのため, 計量ファイナンスの分野では, $\int_0^T c_t dt$ を X の区間 $[0,T]$ における**累積共分散行列**と呼び, X の区間 $[0,T]$ における (条件付き) 共分散行列にあたる量とみなして分析を行う, ということがしばしば行われる.

数理ファイナンスでは, X は (対数) 資産価格過程に対する典型的なモデルである. 資産価格の収益率の共分散行列は, 投資戦略のリスク評価や最適化をする際にしばしば重要な役割を果たすため, それをデータから推定することは計量ファイナンスにおける重要な問題の 1 つである. ここでは, X が区間 $[0,T]$ を n 等分する時点 $t_i=Ti/n$ $(i=0,1,\dots,n)$ において離散観測される場合に, 観測データ $(X_{t_i})_{i=0}^n$ を用いて $[X,X]_T$ を推定する問題を考える. 典型的には, 区間 $[0,T]$ は特定の 1 日を表し, その 1 日内に特定の頻度 (例えば 5 分ごと) で資産価格の対数値を記録した高頻度データが $(X_{t_i})_{i=0}^n$ に対応する.

$[X,X]_T$ の推定量として最も典型的なものは, **実現共分散行列**

$$\widehat{[X,X]}_T:=\sum_{i=1}^n (X_{t_i}-X_{t_{i-1}})(X_{t_i}-X_{t_{i-1}})^\top$$

である. d が固定されている場合, $\widehat{[X,X]}_T$ が $[X,X]_T$ の一致推定量となること, すなわち $n\to\infty$ のとき $\widehat{[X,X]}_T$ が $[X,X]_T$ に確率収束することは, 命題 4.10 から直ちに従う. いまの設定では, 係数過程に適当な可積分条件を課すことで, 推定誤差に対する精密な評価を得ることができる:

定理 4.6. ある定数 $K>0$ が存在して, すべての $j=1,\dots,d$ について

$$\mathrm{E}\left[\int_0^T |b_s^j|^2 ds\right]+\sqrt{\mathrm{E}\left[\int_0^T |c_s^{jj}|^2 ds\right]}\leq K$$

が成り立つと仮定する. このとき, ある普遍定数 $C>0$ が存在して

$$\mathrm{E}\left[\left\|\widehat{[X,X]}_T-[X,X]_T\right\|_F\right]\leq CKd\left(\frac{T}{n}+\frac{\sqrt{T}+T^{3/4}}{\sqrt{n}}\right) \tag{4.4}$$

が成り立つ.

以降の結果の証明における記法の簡略化のために, いくつか記号を導入する. 2 つの実数 a, b に対して, ある正の普遍定数 C が存在して $a \le Cb$ が成り立つことを, 記号 $a \lesssim b$ で表す. d 次元確率過程 $U = (U_t)_{t \ge 0}$ に対して,

$$\Delta_i^n U := U_{t_i} - U_{t_{i-1}} \qquad (i = 1, \ldots, n)$$

と定める. さらに, 2 つの d 次元確率過程 U, V に対して,

$$\widehat{[U, V]}_T := \sum_{i=1}^n \Delta_i^n U (\Delta_i^n V)^\top$$

とおく.

定理 4.6 の証明. 3 ステップに分けて証明する.

Step 1. 2 つの d 次元確率過程 $A = (A_t)_{t \ge 0}$ および $M = (M_t)_{t \ge 0}$ を

$$A_t = \int_0^t b_s ds, \qquad M_t = \int_0^t \sigma_s dW_s$$

で定め, 以下の分解を考える:

$$\widehat{[X, X]}_T - [X, X]_T = \widehat{[A, A]}_T + \widehat{[A, M]}_T + \widehat{[M, A]}_T + (\widehat{[M, M]}_T - [X, X]_T). \tag{4.5}$$

任意の d 次元確率過程 U, V に対して, Schwarz の不等式より

$$\left\| \widehat{[U, V]}_T \right\|_F \le \sqrt{\sum_{j,k=1}^d \sum_{i=1}^n (\Delta_i^n U^j)^2 \sum_{i=1}^n (\Delta_i^n V^k)^2} = \sqrt{\sum_{i=1}^n |\Delta_i^n U|^2 \sum_{i=1}^n |\Delta_i^n V|^2}$$

が成り立つ. 従って,

$$\begin{aligned} &\left\| \widehat{[X, X]}_T - [X, X]_T \right\|_F \\ &\le \sum_{i=1}^n |\Delta_i^n A|^2 + 2\sqrt{\sum_{i=1}^n |\Delta_i^n A|^2 \sum_{i=1}^n |\Delta_i^n M|^2} + \left\| \widehat{[M, M]}_T - [X, X]_T \right\|_F \end{aligned} \tag{4.6}$$

が成り立つ.

Step 2. 各 $i = 1, \ldots, n$ について

$$\mathrm{E}\,|\Delta_i^n A|^2 = \sum_{j=1}^d \mathrm{E}\left[\left(\int_{t_{i-1}}^{t_i} b_s^j ds \right)^2 \right] \le \frac{T}{n} \sum_{j=1}^d \mathrm{E}\left[\int_{t_{i-1}}^{t_i} |b_s^j|^2 ds \right]$$

が成り立つから,

$$\mathrm{E}\left[\sum_{i=1}^n |\Delta_i^n A|^2 \right] \le \frac{T}{n} \sum_{j=1}^d \mathrm{E}\left[\int_0^T |b_s^j|^2 ds \right] \le \frac{TKd}{n} \tag{4.7}$$

を得る. また, 確率積分の等長性より

$$\mathrm{E}\left[|\Delta_i^n M|^2\right] = \sum_{j=1}^d \mathrm{E}\left[\int_{t_{i-1}}^{t_i} c_s^{jj} ds\right]$$

が成り立つから,

$$\mathrm{E}\left[\sum_{i=1}^n |\Delta_i^n M|^2\right] = \sum_{j=1}^d \mathrm{E}\left[\int_0^T c_s^{jj} ds\right] \leq \sum_{j=1}^d \sqrt{T\,\mathrm{E}\left[\int_0^T |c_s^{jj}|^2 ds\right]} \leq d\sqrt{T}K$$

を得る. 従って,

$$\mathrm{E}\left[\sqrt{\sum_{i=1}^n |\Delta_i^n A|^2 \sum_{i=1}^n |\Delta_i^n M|^2}\right] \leq \sqrt{\mathrm{E}\left[\sum_{i=1}^n |\Delta_i^n A|^2\right] \mathrm{E}\left[\sum_{i=1}^n |\Delta_i^n M|^2\right]} \leq \frac{T^{3/4}Kd}{\sqrt{n}} \tag{4.8}$$

を得る.

Step 3. 各 $i \in \{1, \ldots, n\}$ と $j, k \in \{1, \ldots, d\}$ に対して,

$$\xi_i^{jk} := (\Delta_i^n M^j)(\Delta_i^n M^k) - \Delta_i^n [M^j, M^k]$$

とおく. $[X^j, X^k] = [M^j, M^k]$ に注意すると,

$$\widehat{[M^j, M^k]}_T - [X^j, X^k]_T = \sum_{i=1}^n \xi_i^{jk} \tag{4.9}$$

と書き直せる. また, 部分積分公式より,

$$\xi_i^{jk} = \int_{t_{i-1}}^{t_i} (M_s^j - M_{t_{i-1}}^j) dM_s^k + \int_{t_{i-1}}^{t_i} (M_s^k - M_{t_{i-1}}^k) dM_s^j$$

と書き直せる. いま, Schwarz の不等式より,

$$\begin{aligned}
&\mathrm{E}\left[\int_{t_{i-1}}^{t_i} (M_s^j - M_{t_{i-1}}^j)^2 d\langle M^k, M^k\rangle_s\right] \\
&= \mathrm{E}\left[\int_{t_{i-1}}^{t_i} (M_s^j - M_{t_{i-1}}^j)^2 c_s^{kk} ds\right] \\
&\leq \sqrt{\mathrm{E}\left[\int_{t_{i-1}}^{t_i} (M_s^j - M_{t_{i-1}}^j)^4 ds\right] \mathrm{E}\left[\int_{t_{i-1}}^{t_i} |c_s^{kk}|^2 ds\right]}
\end{aligned}$$

が成り立つ. さらに, Burkholder–Davis–Gundy の不等式より,

$$\mathrm{E}\left[\int_{t_{i-1}}^{t_i} (M_s^j - M_{t_{i-1}}^j)^4 ds\right] \lesssim \mathrm{E}\left[\int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^s c_u^{jj} du\right)^2 ds\right] \leq \frac{T}{n} \mathrm{E}\left[\left(\int_{t_{i-1}}^{t_i} c_u^{jj} du\right)^2\right]$$

となるから,

$$
\begin{aligned}
\mathrm{E}\left[\int_{t_{i-1}}^{t_i}(M_s^j-M_{t_{i-1}}^j)^2 d\langle M^k,M^k\rangle_s\right] &\lesssim \sqrt{\frac{T}{n}\mathrm{E}\left[\left(\int_{t_{i-1}}^{t_i}c_u^{jj}du\right)^2\right]\mathrm{E}\left[\int_{t_{i-1}}^{t_i}|c_s^{kk}|^2ds\right]}\\
&\leq \frac{T}{n}\sqrt{\mathrm{E}\left[\int_{t_{i-1}}^{t_i}|c_u^{jj}|^2du\right]\mathrm{E}\left[\int_{t_{i-1}}^{t_i}|c_s^{kk}|^2ds\right]}
\end{aligned}
$$

を得る. ただし, 2 行目の不等式を得るのに Schwarz の不等式を用いた. j と k を入れ替えた式も同様にして示せるので, 命題 4.6 より $(\xi_i^{jk})_{i=1}^n$ はフィルトレーション $(\mathcal{F}_{t_i})_{i=0}^n$ に関するマルチンゲール差分列であり,

$$
\mathrm{E}\,|\xi_i^{jk}|^2 \lesssim \frac{T}{n}\sqrt{\mathrm{E}\left[\int_{t_{i-1}}^{t_i}|c_u^{jj}|^2du\right]\mathrm{E}\left[\int_{t_{i-1}}^{t_i}|c_s^{kk}|^2ds\right]}
$$

が成り立つ. 従って,

$$
\begin{aligned}
\mathrm{E}\left|\sum_{i=1}^n\xi_i^{jk}\right|^2 &= \sum_{i=1}^n\mathrm{E}\,|\xi_i^{jk}|^2 \lesssim \frac{T}{n}\sum_{i=1}^n\sqrt{\mathrm{E}\left[\int_{t_{i-1}}^{t_i}|c_u^{jj}|^2du\right]\mathrm{E}\left[\int_{t_{i-1}}^{t_i}|c_s^{kk}|^2ds\right]}\\
&\leq \frac{T}{n}\sqrt{\mathrm{E}\left[\int_0^T|c_u^{jj}|^2du\right]\mathrm{E}\left[\int_0^T|c_s^{kk}|^2ds\right]} \leq \frac{TK^2}{n}
\end{aligned}
$$

が成り立つ. よって,

$$
\begin{aligned}
\mathrm{E}\left[\left\|\widehat{[M,M]}_T-[X,X]_T\right\|_F\right] &\leq \sqrt{\mathrm{E}\left[\left\|\widehat{[M^j,M^k]}_T-[X^j,X^k]_T\right\|_F^2\right]}\\
&= \sqrt{\sum_{j,k=1}^d\mathrm{E}\left|\sum_{i=1}^n\xi_i^{jk}\right|^2} \lesssim dK\sqrt{\frac{T}{n}} \qquad (4.10)
\end{aligned}
$$

が成り立つ. (4.6)–(4.10) を組みわせて示すべき不等式を得る. □

実現共分散行列の (確率的な意味での) 収束レートを導出するだけであれば, 係数過程に対する可積分性の仮定は,「局所化の議論」と呼ばれる論法を用いることで大幅に緩めることができる.

系 4.2. T を n に依存しない正の定数とする. また, W,b,σ はいずれも n に依存しない (従って次元 d も n に依存しない) ものとして,

$$
\int_0^T|b_s|^2ds+\int_0^T\|c_s\|_F^2ds<\infty \quad \text{a.s.} \qquad (4.11)
$$

を満たすと仮定する. このとき,

$$
\left\|\widehat{[X,X]}_T-[X,X]_T\right\|_F = O_p(n^{-1/2}) \qquad (n\to\infty)
$$

が成り立つ.

証明. 各 $N \in \mathbb{N}$ について

$$\tau_N := \inf\left\{t \geq 0 : \int_0^t |b_s|^2 ds + \int_0^t \|c_s\|_F^2 ds > N\right\}$$

と定める. τ_N は停止時刻であることが容易に確認できる. 従って, 命題 4.7 より, 任意の $t \geq 0$ に対して

$$X_t^{\tau_N} = X_0 + \int_0^t b_s 1_{(0,\tau_N]}(s) ds + \int_0^t \sigma_s 1_{(0,\tau_N]}(s) dW_s$$

が成り立つ. さらに,

$$\int_0^T |b_s|^2 1_{(0,\tau_N]} ds + \int_0^t \|c_s\|_F^2 1_{(0,\tau_N]} ds \leq N$$

が成り立つ. 実際, $t < \tau_N$ の場合は τ_N の定義より明らかであり, $\tau_N \leq t$ の場合は, 特に $\tau_N < \infty$ なので, 任意の $k \in \mathbb{N}$ に対して

$$\int_0^{(\tau_N - 1/k)_+} |b_s|^2 ds + \int_0^{(\tau_N - 1/k)_+} \|c_s\|_F^2 ds \leq N$$

が成り立つことが τ_N の定義から従う. 上の式で $k \to \infty$ とすることで示すべき不等式を得る. 以上と定理 4.6 より, N, d, T にのみ依存するある定数 $C_N > 0$ が存在して

$$\mathrm{E}\left[\left\|\widehat{[X^{\tau_N}, X^{\tau_N}]}_T - [X^{\tau_N}, X^{\tau_N}]_T\right\|_F\right] \leq \frac{C_N}{\sqrt{n}}$$

が成り立つ. 命題 4.4 より $[X^{\tau_N}, X^{\tau_N}] = [X, X]^{\tau_N}$ であることに注意すると, 任意の $K > 0$ に対して,

$$\begin{aligned}
&\mathrm{P}\left(\sqrt{n}\left\|\widehat{[X,X]}_T - [X,X]_T\right\|_F > K\right) \\
&\leq \mathrm{P}\left(\sqrt{n}\left\|\widehat{[X^{\tau_N}, X^{\tau_N}]}_T - [X^{\tau_N}, X^{\tau_N}]_T\right\|_F > K\right) + \mathrm{P}(\tau_N \leq T) \leq \frac{C_N}{K} + \mathrm{P}(\tau_N \leq T)
\end{aligned}$$

が成り立つ. 従って,

$$\limsup_{K\to\infty} \sup_{n\in\mathbb{N}} \mathrm{P}\left(\sqrt{n}\left\|\widehat{[X,X]}_T - [X,X]_T\right\|_F > K\right) \leq \mathrm{P}(\tau_N \leq T)$$

を得る. 仮定より $\mathrm{P}(\tau_N \leq T) \to 0$ $(N \to \infty)$ であるから, 上の不等式において $N \to \infty$ とすれば, 示すべき結論を得る. □

注意 4.5. 仮定 (4.11) の下で, 実現共分散行列の収束レートだけでなく漸近混合正規性も導出できることが知られている. すなわち, $n \to \infty$ のとき $\sqrt{n}(\widehat{[X,X]}_T - [X,X]_T)$ がある混合正規分布に安定収束することが示せる. (Jacod & Protter, 2012, Theorem 5.4.2) 参照.

5 実現共分散行列: 高次元の場合

不等式 (4.4) の右辺が 0 に収束するためには, K, T が n に依存しない限りは $d = o(\sqrt{n})$ という条件が必要である. 次元 d がサンプル数 n と同程度, もしくははるかに大きくなるような高次元の設定においても実現共分散行列の推定誤差が (何らかの意味で) 小さくなることを保証するために, 3.1 節と同様に, 最大値ノルム

$$\left|\widehat{[X,X]}_T - [X,X]_T\right|_\infty = \max_{1\le j,k\le d}\left|\widehat{[X^j,X^k]}_T - [X^j,X^k]_T\right|$$

によって推定誤差を評価することを試みる. (4.9) 式より, $\widehat{[X^j,X^k]}_T - [X^j,X^k]_T$ の主要項はマルチンゲールで表せるため, マルチンゲールに対する Bernstein 不等式 (定理 3.1) を適用してそのような評価を得ることができそうである. そのためには, (4.9) 式のマルチンゲールが Bernstein 条件 (3.1) を満たすことを確認する必要があり, Burkholder–Davis–Gundy の不等式 (定理 4.5) に現れる定数 $C_2(p)$ について, $p \to \infty$ のときの増大度を調べる必要がある. (長井, 1999, 定理 2.8.1) で与えられているような伊藤の公式による証明では, $C_2(p) = O(p)$ $(p \to \infty)$ となるように定数 $C_2(p)$ を取れることが示せるが, この評価は Bernstein 条件の確認には不十分である. 本節ではまず, $C_2(p) = O(\sqrt{p})$ $(p \to \infty)$ となるように取れることを示したのち, その結果とマルチンゲールに対する Bernstein 不等式を組み合わせることで, 実現共分散行列の推定誤差の最大値ノルムによる評価を導出する.

以下この講義ノートでは, $\log d = 0$ となる状況を避けるために $d > 1$ と仮定する.

5.1 Burkholder–Davis–Gundy の不等式における定数の精密評価

この小節の目標は次の結果を示すことである.

定理 5.1. ある普遍定数 $c > 0$ が存在して, 任意の連続局所マルチンゲール $M = (M_t)_{t\ge0}$ と停止時刻 τ, および定数 $p \ge 2$ に対して

$$\left\|\sup_{0\le t\le\tau}|M_t|\right\|_p \le c\sqrt{p}\left\|\langle M,M\rangle_\tau^{1/2}\right\|_p$$

が成り立つ.

ここでは (Barlow & Yor, 1982, Proposition 4.2) の証明の議論に沿って証明を与える.

補題 5.1 (Burkholder (1973), Lemma 7.1)**.** X, Y を 2 つの非負値確率変数とする. ある定数 $\beta > 1$ および $\delta, \varepsilon > 0$ が存在して, 任意の $\lambda > 0$ に対して

$$\mathrm{P}(X > \beta\lambda, Y \le \delta\lambda) \le \varepsilon\,\mathrm{P}(X > \lambda) \tag{5.1}$$

が成り立つと仮定する. このとき, $p \ge 1$ が $\beta^p\varepsilon < 1$ を満たすならば,

$$\mathrm{E}[X^p] \le \frac{(\beta/\delta)^p}{1-\beta^p\varepsilon}\,\mathrm{E}[Y^p]$$

が成り立つ.

証明. 各 $n = 1, 2, \ldots$ について, X を $X \wedge n$ に置き換えても (5.1) は成立することに注意すると, 単調収束定理より, X を $X \wedge n$ に置き換えた不等式を証明すれば十分である. 従って, $\mathrm{E}[X^p] < \infty$ と仮定して一般性を失わない. 補題 2.1 より,

$$\begin{aligned}\mathrm{E}[X^p] &= p \int_0^\infty u^{p-1}\,\mathrm{P}(X > u)du = p\beta^p \int_0^\infty \lambda^{p-1}\,\mathrm{P}(X > \beta\lambda)du \\ &\le p\beta^p \int_0^\infty \lambda^{p-1}\,\mathrm{P}(X > \beta\lambda, Y \le \delta\lambda)d\lambda + p\beta^p \int_0^\infty \lambda^{p-1}\,\mathrm{P}(Y > \delta\lambda)d\lambda\end{aligned}$$

が成り立つから, (5.1) より

$$\begin{aligned}\mathrm{E}[X^p] &\le p\beta^p\varepsilon \int_0^\infty \lambda^{p-1}\,\mathrm{P}(X > \lambda)d\lambda + p(\beta/\delta)^p \int_0^\infty u^{p-1}\,\mathrm{P}(Y > u)du \\ &= \beta^p\varepsilon\,\mathrm{E}[X^p] + (\beta/\delta)^p\,\mathrm{E}[Y^p]\end{aligned}$$

を得る. 整理して示すべき不等式を得る. □

注意 5.1. (5.1) のような形の不等式は (X, Y) に対する **good λ inequality** と呼ばれる.[*2] 形式的な定義は (Revuz & Yor, 1999, Chapter IV, Definition 4.8) を参照のこと.

補題 5.2. $A = (A_t)_{t\ge 0}, B = (B_t)_{t\ge 0}$ を 2 つの連続増加過程で $A_0 = B_0 = 0$ を満たすものとする. また, ある定数 $p, D > 0$ が存在して, $S \le T$ なる任意の有界な停止時刻 S, T に対して

$$\mathrm{E}\left[(A_T - A_S)^p\right] \le D\|B_T\|_\infty^p\,\mathrm{P}(S < T) \tag{5.2}$$

が成り立つと仮定する. このとき, 任意の定数 $\beta > 1, \lambda > 0, \delta > 0$ と任意の有界な停止時刻 τ に対して,

$$\mathrm{P}(A_\tau > \beta\lambda, B_\tau \le \delta\lambda) \le \frac{D}{(\beta-1)^p}\delta^p\,\mathrm{P}(A_\tau > \lambda)$$

が成り立つ.

証明. $S := \inf\{t \ge 0 : A_t > \lambda\} \wedge \tau, T := \inf\{t \ge 0 : B_t > \delta\lambda\} \wedge \tau$ とおく. よく知られているように, S, T はともに停止時刻となる. また, τ は有界であるから, S, T もともに有界である. いま, $A_\tau > \lambda$ ならば, A の連続性よりある $t < \tau$ が存在して $A_t > \lambda$ となるから, $S \le t < \tau$ である. 従って, $\beta > 1$ に注意して,

$$\mathrm{P}(A_\tau > \beta\lambda, B_\tau \le \delta\lambda) = \mathrm{P}(A_\tau > \beta\lambda, S < \tau, B_\tau \le \delta\lambda)$$

を得る. 次に, もし $T < \tau$ ならば, ある $t < \tau$ が存在して $B_t > \delta\lambda$ となるから, $B_\tau \ge B_t > \delta\lambda$ となる. 従って,

$$\mathrm{P}(A_\tau > \beta\lambda, B_\tau \le \delta\lambda) \le \mathrm{P}(A_\tau > \beta\lambda, S < \tau, T \ge \tau) \le \mathrm{P}(A_T > \beta\lambda, S < \tau)$$

[*2] 風巻 (2003) では**分布関数不等式**と訳されている. これは, Burkholder (1973) においては, 当該不等式が "distribution function inequality" と呼ばれていることに由来する訳語だと思われる.

が成り立つ. ここで, $S < \tau$ のとき, A の連続性より $A_S = \lambda$ となるから, 特に $A_{T \wedge S} \le A_S = \lambda$ である. 従って,

$$\mathrm{P}(A_\tau > \beta\lambda, B_\tau \le \delta\lambda) \le \mathrm{P}(A_T - A_{T\wedge S} > (\beta - 1)\lambda)$$

が成り立つから, Markov の不等式と (5.2) より

$$\mathrm{P}(A_\tau > \beta\lambda, B_\tau \le \delta\lambda) \le \frac{\mathrm{E}\left[(A_T - A_{T\wedge S})^p\right]}{(\beta-1)^p\lambda^p} \le \frac{D\|B_T\|_\infty^p \,\mathrm{P}(S < T)}{(\beta-1)^p\lambda^p}$$

を得る. $S < T$ のとき $S < \tau$ だから, ある $t < \tau$ が存在して $\lambda < A_t \le A_\tau$ となる. 従って, $\mathrm{P}(S < T) \le \mathrm{P}(A_\tau > \lambda)$ が成り立つ. また, 任意の $t < T$ に対して $B_t \le \delta\lambda$ が成り立つから, B の連続性より $B_T \le \delta\lambda$ が成り立つ. 従って $\|B_T\|_\infty \le \delta\lambda$ であるから, 示すべき不等式が得られた. □

定理 5.1 は, 上の 2 つの補題と, Wiener 過程に関するよく知られた以下の 2 つの結果を組み合わせることで証明する. 第 1 の結果は Wiener 過程の強 Markov 性である.

定理 5.2 (Wiener 過程の強 Markov 性)**.** W を $\mathcal{B}$ 上の r 次元標準 Wiener 過程, τ を停止時刻とする. $\tau < \infty$ ならば, $(W_{\tau+t} - W_\tau)_{t\ge0}$ は $(\Omega, \mathcal{F}, (\mathcal{F}_{\tau+t})_{t\ge0}, \mathrm{P})$ 上の r 次元標準 Wiener 過程である.

証明. (Ikeda & Watanabe, 1989, Chapter II, Theorem 6.4) 参照. □

第 2 の結果を述べるために, 確率基底の拡大の概念を導入しておく:

定義 5.1 (確率基底の拡大)**.** 確率基底 $\tilde{\mathcal{B}} = (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\ge0}, \tilde{\mathrm{P}})$ が確率基底 $\mathcal{B}$ の**拡大**であるとは, ある確率基底 $\mathcal{B}' = (\Omega', \mathcal{F}', (\mathcal{F}'_t)_{t\ge0}, \mathrm{P}')$ が存在して以下の条件を満たすことをいう:

(i) $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathrm{P}})$ は $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}', \mathrm{P} \times \mathrm{P}')$ の完備化である.

(ii) 各 $t \ge 0$ について $\tilde{\mathcal{F}}_t = \bigcap_{s:s>t} \sigma\left((\mathcal{F}_t \otimes \mathcal{F}'_t) \cup \mathcal{N}\right)$ が成り立つ. ここに, $\mathcal{N}$ は $\tilde{\mathrm{P}}$-ゼロ集合の全体を表す.

$\mathcal{B}$ の拡大 $\tilde{B}$ が与えられたとき, $(\Omega, \mathcal{F}, \mathrm{P})$ 上で定義された確率変数 X は, $(\omega, \omega') \mapsto X(\omega)$ と同一視することで, $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathrm{P}})$ 上で定義された確率変数とも見なすことにする.

定理 5.3 (マルチンゲールの時間変更による表現)**.** $M = (M_t)_{t\ge0}$ を連続局所マルチンゲールとし, 各 $t \ge 0$ について停止時刻 τ_t を以下で定める:

$$\tau_t = \begin{cases} \inf\{u : \langle M, M\rangle_u > t\} & \text{if } t < \lim_{s\to\infty}\langle M, M\rangle_s, \\ \infty & \text{otherwsie.} \end{cases}$$

さらに, $\hat{\mathcal{F}}_t := \sigma(\bigcup_{s:s>0} \mathcal{F}_{\tau_t\wedge s})$ とおく. このとき, 確率基底 $(\Omega, \mathcal{F}, (\hat{\mathcal{F}}_t)_{t\ge0}, \mathrm{P})$ の拡大 $\tilde{\mathcal{B}} = (\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\ge0}, \tilde{\mathrm{P}})$ と $\tilde{\mathcal{B}}$ 上の標準 Wiener 過程 $W = (W_t)_{t\ge0}$ が存在して, 任意の $t \ge 0$ に対して $M_t = W_{\langle M,M\rangle_t}$ が成り立つ.

証明. (長井, 1999, 定理 2.7.6) とその後の注意を参照. もしくは (Ikeda & Watanabe, 1989, Chapter II, Theorem 7.2′) 参照. □

定理 5.1 の証明. 2 ステップに分けて証明する.

Step 1. M が $\mathcal{B}$ 上の標準 Wiener 過程で, かつ τ が有界な場合には定理の結論が成り立つことを示す. 各 $t \geq 0$ について $A_t := \sup_{0 \leq s \leq t} |M_s|$ とおく. 容易に確認できるように, $A = (A_t)_{t \geq 0}$ は連続増加過程である. 補題 5.2 を適用するために, $S \leq T$ かつ $\|T\|_\infty < \infty$ なる停止時刻 S, T を 1 組固定する. Wiener 過程の強 Markov 性より, $(M_{S+t} - M_S)_{t \geq 0}$ は $(\Omega, \mathcal{F}, (\mathcal{F}_{S+t})_{t \geq 0}, \mathrm{P})$ 上の標準 Wiener 過程である. 従って, Doob の不等式より

$$\begin{aligned}\mathrm{E}\left[\sup_{0 \leq t \leq \|T\|_\infty} |M_{S+t} - M_S|^p \mid \mathcal{F}_S\right] &\leq \left(\frac{p}{p-1}\right)^p \mathrm{E}\left[|M_{S+\|T\|_\infty} - M_S|^p \mid \mathcal{F}_S\right] \\ &= \left(\frac{p}{p-1}\right)^p \|T^{1/2}\|_\infty^p m_p\end{aligned}$$

が成り立つ. ここで, m_p は標準正規分布の p 次の絶対モーメントを表す.

$$|A_T - A_S| \leq \sup_{0 \leq t \leq \|T\|_\infty} |M_{S+t} - M_S| 1_{\{S<T\}}$$

であり, かつ $\{S < T\} \in \mathcal{F}_S$ であることから,

$$\mathrm{E}\left[|A_T - A_S|^p\right] \leq \left(\frac{p}{p-1}\right)^p m_p \|T^{1/2}\|_\infty^p \, \mathrm{P}(S < T)$$

を得る. 従って, $B_t = \sqrt{t}$, $D = (p/(p-1))^p m_p$ として (5.2) が成立するから, 補題 5.2 を $\beta = 2$, $\delta = (2^{p+1}D)^{-1/p}$ として適用すると, 任意の $\lambda > 0$ に対して

$$\mathrm{P}(A_\tau > 2\lambda, \tau^{1/2} \leq \delta\lambda) \leq D\delta^p \, \mathrm{P}(A_\tau > \lambda)$$

が成り立つ. $2^p D\delta^p = 1/2 < 1$ に注意すると, 補題 5.1 より

$$\mathrm{E}[A_\tau^p] \leq 2^{p+1}\delta^{-p} \, \mathrm{E}[\tau^{p/2}] = 4^{p+1} D \, \mathrm{E}[\tau^{p/2}]$$

を得る. Stirling 近似から

$$m_p = \frac{2^{p/2}\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}} < 2^{(p+1)/2}\left(\frac{p+1}{2}\right)^{p/2} e^{1/12}$$

が成り立つことに注意すると,

$$D^{1/p} \lesssim \frac{p}{p-1}\sqrt{\frac{p+1}{2}} \lesssim \sqrt{p}$$

である. 従って, $\|A_\tau\|_p \lesssim \left\|\tau^{1/2}\right\|_p$ を得る. $\langle M, M \rangle_\tau = \tau$ に注意すれば, 示すべき不等式が得られた.

Step 2. 一般の場合に定理の結論が成り立つことを示す. 定理 5.3 のように停止時刻 τ_t, 確率基底 $\tilde{\mathcal{B}}$ および標準 Wiener 過程 W を定める. Step 1 より, 任意の有界な停止時刻 T に対して

$$\left\|\sup_{0 \leq t \leq T} |W_t|\right\|_p \lesssim \sqrt{p}\left\|T^{1/2}\right\|_p \tag{5.3}$$

が成り立つ. この不等式は T が非有界な場合にも成り立つ. 実際, 任意の $K>0$ に対して, $T\wedge K$ は有界な停止時刻であるから,

$$\left\|\sup_{0\le t\le T\wedge K}|W_t|\right\|_p\lesssim\sqrt{p}\left\|(T\wedge K)^{1/2}\right\|_p$$

が成り立つ. 両辺で $K\to\infty$ とすると, 単調収束定理より (5.3) を得る.

W の定義より

$$\sup_{0\le t\le\tau}|M_t|=\sup_{0\le t\le T}|W_{\langle M,M\rangle_t}|\le\sup_{0\le t\le\langle M,M\rangle_\tau}|W_t|$$

が成り立つから, $\langle M,M\rangle_\tau$ が $(\tilde{\mathcal{F}}_t)$-停止時刻であることを示せれば, (5.3) を $T=\langle M,M\rangle_\tau$ に対して適用することで示すべき不等式が得られる. 任意の $t,s\ge0$ に対して

$$\{\langle M,M\rangle_{\tau\wedge s}>t\}=\{\tau\wedge s>\tau_t\}=\{\tau\wedge s>\tau_t\wedge s\}\in\mathcal{F}_{\tau_t\wedge s}\subset\hat{\mathcal{F}}_t$$

が成り立つから, $\langle M,M\rangle_{\tau\wedge s}$ は $(\hat{\mathcal{F}}_t)$-停止時刻である. 従って $\langle M,M\rangle_\tau=\lim_{s\to\infty}\langle M,M\rangle_{\tau\wedge s}$ も $(\hat{\mathcal{F}}_t)$-停止時刻であり, 故に $(\tilde{\mathcal{F}}_t)$-停止時刻でもある. □

注意 5.2. $p\in(0,\infty)$ に対して, 不等式

$$\|M_\tau\|_p\le c_p\left\|\langle M,M\rangle_\tau^{1/2}\right\|_p$$

を任意の連続局所マルチンゲール $M=(M_t)_{t\ge0}$ と有界停止時刻 τ に対して成立させるような最小の定数 c_p は Davis (1976) によって決定されている. また, $p\ge1$ に対しては, 上の不等式は $c_p=2\sqrt{p}$ として成立することが Carlen & Kree (1991) によって示されている. この結果を用いると, Doob の不等式より, 定理 5.1 は $c=4$ として成立することが従う.

5.2 実現共分散行列の推定誤差の最大値ノルムによる評価

4.2 節と同じ設定を考える. 命題 3.2 に対応する結果として, 以下の評価を示す.

定理 5.4. ある定数 $K>0$ と P-測度ゼロ集合 N が存在して, 任意の $s\in[0,T]$ に対して

$$|b_s|_\infty\vee\sqrt{|c_s|_\infty}\le K\qquad\text{on}\quad\Omega\setminus N$$

が成り立つと仮定する. このとき, ある普遍定数 $C>0$ が存在して

$$\left\|\left|\widehat{[X,X]}_T-[X,X]_T\right|_\infty\right\|_{\psi_1}\le CK^2\left(\frac{T^{3/2}+T\sqrt{\log(3d^2)}}{\sqrt{n}}+\frac{T^2+T\log(3d^2)}{n}\right)$$

が成り立つ.

証明. 2 ステップに分けて証明する.

Step 1. 定理 4.6 の証明と同じように記号を定義して, 分解 (4.5) を考える. Schwarz の不等式より

$$\begin{aligned}
&\left|\widehat{[X,X]}_T-[X,X]_T\right|_\infty \\
&\le \max_{1\le j\le d}\sum_{i=1}^n|\Delta_i^nA^j|^2+2\sqrt{\max_{1\le j\le d}\sum_{i=1}^n|\Delta_i^nA^j|^2\max_{1\le k\le d}\sum_{i=1}^n|\Delta_i^nM^k|^2}+\left|\widehat{[M,M]}_T-[X,X]_T\right|_\infty
\end{aligned}$$

が成り立つ. 仮定より $|\Delta_i^nA^j|\le KT/n$ であるから,

$$\max_{1\le j\le d}\sum_{i=1}^n|\Delta_i^nA^j|^2\le\frac{(KT)^2}{n}$$

である. 故に,

$$\left|\widehat{[X,X]}_T-[X,X]_T\right|_\infty\le\frac{(KT)^2}{n}+\frac{2KT}{\sqrt{n}}\sqrt{\max_{1\le k\le d}\sum_{i=1}^n|\Delta_i^nM^k|^2}+\left|\widehat{[M,M]}_T-[X,X]_T\right|_\infty$$

を得る. さらに,

$$\max_{1\le k\le d}\sum_{i=1}^n|\Delta_i^nM^k|^2\le\left|\widehat{[M,M]}_T-[X,X]_T\right|_\infty+\max_{1\le k\le d}[X^k,X^k]_T$$

であり, 仮定より $[X^k,X^k]_T\le K^2T$ であるから,

$$\begin{aligned}
\left|\widehat{[X,X]}_T-[X,X]_T\right|_\infty&\le\frac{(KT)^2}{n}+\frac{2K^2T^{3/2}}{\sqrt{n}}\\
&\quad+\frac{2KT}{\sqrt{n}}\sqrt{\left|\widehat{[M,M]}_T-[X,X]_T\right|_\infty}+\left|\widehat{[M,M]}_T-[X,X]_T\right|_\infty
\end{aligned}$$

を得る. 右辺第 3 項に相加相乗平均の不等式を適用して整理すると

$$\left|\widehat{[X,X]}_T-[X,X]_T\right|_\infty\le\frac{2(KT)^2}{n}+\frac{2K^2T^{3/2}}{\sqrt{n}}+2\left|\widehat{[M,M]}_T-[X,X]_T\right|_\infty$$

となる.

Step 2. $\left|\widehat{[M,M]}_T-[X,X]_T\right|_\infty$ を評価する. 各 $j,k\in\{1,\dots,d\}$ について, 分解 (4.9) を考える. 任意の整数 $p\ge2, i=1,\dots,n$ と $F\in\mathcal{F}_{t_{i-1}}$ に対して,

$$\left\|\xi_i^{jk}1_F\right\|_p\le\left\|(\Delta_i^nM^j)(\Delta_i^nM^k)1_F\right\|_p+\left\|\Delta_i^n[M^j,M^k]1_F\right\|_p$$

が成り立つ. 仮定より,

$$|\Delta_i^n[M^j,M^k]|\le\int_{t_{i-1}}^{t_i}|c_s^{jk}|ds\le\frac{K^2T}{n}$$

である. 一方で, Schwarz の不等式より,

$$\left\|(\Delta_i^nM^j)(\Delta_i^nM^k)1_F\right\|_p\le\left\|\Delta_i^nM^j1_F\right\|_{2p}\left\|\Delta_i^nM^k1_F\right\|_{2p}$$

であり, 定理 5.1 と仮定より

$$\left\|\Delta_i^n M^j 1_F\right\|_{2p} \lesssim \sqrt{p}\left\|\sqrt{\int_{t_{i-1}}^{t_i} c_s^{jj} ds 1_F}\right\|_{2p} \leq \sqrt{p\frac{K^2T}{n}}\left\|1_F\right\|_{2p}$$

が成り立つから,

$$\left\|(\Delta_i^n M^j)(\Delta_i^n M^k)1_F\right\|_p \lesssim p\frac{K^2T}{n}\left\|1_F\right\|_p$$

を得る. 以上より,

$$\left\|\xi_i^{jk}1_F\right\|_p \lesssim p\frac{K^2T}{n}\left\|1_F\right\|_p$$

が成り立つから, ある普遍定数 $c_1 > 0$ が存在して,

$$\mathrm{E}\left[\left|\xi_i^{jk}\right|^p \mid \mathcal{F}_{t_{i-1}}\right] \leq p^p\left(\frac{c_1K^2T}{n}\right)^p$$

が成り立つ. Stirling 近似を用いて整理すると,

$$\mathrm{E}\left[\left|\xi_i^{jk}\right|^p \mid \mathcal{F}_{t_{i-1}}\right] \leq p!\left(\frac{c_1eK^2T}{n}\right)^p = \frac{p!}{2}\cdot\frac{2(c_1eK^2T)^2}{n^2}\cdot\left(\frac{c_1eK^2T}{n}\right)^{p-2}$$

を得る. $(\xi_i^{jk})_{i=1}^n$ はフィルトレーション $(\mathcal{F}_{t_i})_{i=0}^n$ に関するマルチンゲール差分列となっていたから, 定理 3.1 より, ある普遍定数 $c_2 > 0$ が存在して, 任意の $u \geq 0$ に対して

$$\mathrm{P}\left(\left|[\widehat{M^j, M^k}]_T - [X^j, X^k]_T\right| \geq u\right) \leq 2\exp\left(-\frac{u^2}{2c_2(K^4T^2/n + uK^2T/n)}\right)$$

が成り立つ. 従って, 補題 3.3 より

$$\left\|\left|[\widehat{M, M}]_T - [X, X]_T\right|_\infty\right\|_{\psi_1} \lesssim \frac{K^2T\sqrt{\log(3d^2)}}{\sqrt{n}} + \frac{K^2T\log(3d^2)}{n}$$

が従う. Step 1 と 2 の結果を組み合わせて, 示すべき不等式を得る. □

系 4.2 と同様に, $\left|[\widehat{X, X}]_T - [X, X]_T\right|_\infty$ の収束レートを導出するだけであれば, 係数過程に対する有界性の仮定は局所化の議論によって緩和することができる.

定義 5.2 (局所有界過程)**.** d 次元確率過程 $H = (H_t)_{t\geq 0}$ が**局所有界**であるとは, ある局所化列 $(\tau_k)_{k=1}^\infty$ が存在して, 任意の $k \in \mathbb{N}$ に対して $\sup_{\omega\in\Omega}\sup_{0<t\leq\tau_k(\omega)}|H_t(\omega)| < \infty$ が成り立つことをいう.

注意 5.3. 上の定義では, Jacod & Protter (2012) に従って, $\sup_{0\leq t\leq\tau_k}|H_t|$ ではなく $\sup_{0<t\leq\tau_k}|H_t|$ の有界性を要求している. これは, H_0 の有界性を要求することを回避するためである.

系 5.1. T は n に依存しないと仮定する. さらに, ある P-測度ゼロ集合 N と n に依存しない局所有界過程 $H = (H_t)_{t\geq 0}$ が存在して, 任意の $s \in [0, T]$ に対して

$$|b_s|_\infty \vee |c_s|_\infty \leq |H_s| \qquad \text{on} \quad \Omega\setminus N$$

が成り立つと仮定する. このとき, $\log d = o(n)$ $(n \to \infty)$ ならば,

$$\left|\widehat{[X,X]}_T - [X,X]_T\right|_\infty = O_p\left(\sqrt{\frac{\log d}{n}}\right) \qquad (n \to \infty)$$

が成り立つ.

証明. 仮定より, 定義 5.2 の性質を満たすような局所化列 $(\tau_k)_{k=1}^\infty$ が存在する. 任意に $k \in \mathbb{N}$ を固定して, $C_k := \sup_{\omega\in\Omega}\sup_{0<t\leq\tau_k(\omega)}|H_t(\omega)|$ とおく. H は n に依存しないので, C_k も n に依存しないことに注意する. また, 仮定より, 任意の $s \in [0,T]$ に対して

$$|b_s|_\infty 1_{(0,\tau_k]}(s) \vee |c_s|_\infty 1_{(0,\tau_k]}(s) \leq C_k \qquad \text{on} \quad \Omega\setminus N$$

が成り立つ. 命題 4.7 より, 任意の $t \geq 0$ に対して

$$X_t^{\tau_k} = X_0 + \int_0^t b_s 1_{(0,\tau_k]}(s)ds + \int_0^t \sigma_s 1_{(0,\tau_k]}(s)dW_s$$

が成り立つことに注意すると, 定理 5.4 より

$$\left\|\left|\widehat{[X^{\tau_k},X^{\tau_k}]}_T - [X^{\tau_k},X^{\tau_k}]_T\right|_\infty\right\|_{\psi_1} \lesssim C_k^2\left(\frac{T^{3/2}+T\sqrt{\log(3d^2)}}{\sqrt{n}} + \frac{T^2+T\log(3d^2)}{n}\right)$$

が成り立つ. 従って, Markov の不等式と仮定 $\log d = o(n)$ より

$$\begin{aligned}
&\limsup_{K\to\infty}\sup_{n\in\mathbb{N}} \mathrm{P}\left(\left|\widehat{[X,X]}_T - [X,X]_T\right|_\infty > K\sqrt{\frac{\log d}{n}}\right)\\
&\leq \limsup_{K\to\infty}\sup_{n\in\mathbb{N}} \mathrm{P}\left(\left|\widehat{[X^{\tau_k},X^{\tau_k}]}_T - [X^{\tau_k},X^{\tau_k}]_T\right|_\infty > K\sqrt{\frac{\log d}{n}}\right) + \mathrm{P}(\tau_k \leq T)\\
&\leq \mathrm{P}(\tau_k \leq T)
\end{aligned}$$

を得る. $k \to \infty$ のとき $\mathrm{P}(\tau_k \leq T) \to 0$ となるから, 上の不等式において $k \to \infty$ とすれば, 示すべき結論を得る. □

注意 5.4. 系 5.1 で係数過程に課した仮定は, 高次元高頻度データの統計解析では標準的である. 例えば (Fan, Furger & Xiu, 2016, Assumption 1) 参照.

6 グラフィカル Lasso への応用

まず導入として, 1 節と同様に, 平均 0, 共分散行列 Σ の d 次元正規分布に従う独立同分布な確率変数列 $X_1,\dots,X_n$ が観測データとして与えられている状況を考える. Σ は正則であると仮定し, 逆行列 $\Theta_0 := \Sigma^{-1}$ を推定する問題を考える. Θ_0 は**精度行列**と呼ばれることがある. ナイーブなアイディアとして, 標本共分散行列

$$\widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$$

の逆行列で Θ_0 を推定する方法が考えられる. しかし, $\widehat{\Sigma}_n$ のランクは必ず n 以下となるから, 次元 d がサンプル数 n よりも大きいような高次元の状況では, そもそも $\widehat{\Sigma}_n$ の逆行列が存在しない. 高次元の設定下で Θ_0 を推定するためのアプローチは様々なものが提案されているが, ここでは**グラフィカル Lasso** として知られる方法を考える. グラフィカル Lasso では, Θ_0 のスパース性, すなわちほとんどの成分が 0 であることを仮定して, Θ_0 を ℓ_1-正則化最尤法で推定する. いまの場合, 対数尤度関数は本質的に以下の関数の定数倍で与えられる:

$$\ell_n(\Theta; \hat{\Sigma}_n) = \log\det\Theta - \mathrm{tr}(\hat{\Sigma}_n\Theta), \qquad \Theta \in \mathcal{S}_d^+.$$

ただし, $\mathcal{S}_d^+$ は d 次正定値対称行列全体の集合を表す. 実際, $\ell_n(\Theta; \hat{\Sigma}_n)$ は, 対数尤度関数から未知パラメータと無関係な項を取り除いたものを $2/n$ 倍することで得られる. グラフィカル Lasso 推定量は, $\lambda > 0$ を正則化の強さを調整するパラメータとして,

$$-\ell_n(\Theta; \hat{\Sigma}_n) + \lambda \sum_{j,k:j\neq k} |\Theta_{jk}| \tag{6.1}$$

を最小化する $\Theta \in \mathcal{S}_d^+$ として定義される. この最適化問題は確率 1 でただ一つの解を持つことが以下の命題から従う:

命題 6.1. A を d 次半正定値対称行列とする. A の対角成分がすべて正ならば,

$$-\ell_n(\Theta; A) + \lambda \sum_{j,k:j\neq k} |\Theta_{jk}|$$

を最小化する $\Theta \in \mathcal{S}_d^+$ がただ一つ存在する.

証明. (Duchi, Gould & Koller, 2008, Lemma 1) 参照. □

ここまでは対数尤度の計算のためにデータの正規性を課していたが, グラフィカル Lasso 推定量自体は, (6.1) を最小化する $\Theta \in \mathcal{S}_d^+$ として正規性の仮定なしに定義することが可能である. より一般に, $\hat{\Sigma}_n$ を (命題 6.1 の条件を満たすような) 別の Σ の推定量に置き換えても, 形式的には (6.1) を最小化する $\Theta \in \mathcal{S}_d^+$ としてグラフィカル Lasso 推定量を定義できる. 特に, $\hat{\Sigma}_n$ の代わりに実現共分散行列を考えることで, 累積共分散行列の逆行列の推定問題にもこのアプローチを適用できる. 本節では, $\hat{\Sigma}_n$ が最大値ノルムに関して Σ の一致推定量であるならば, 適当な仮定の下でグラフィカル Lasso 推定量も (作用素ノルムに関して) Θ_0 の一致推定量となることを示す. 従って, 前節までに示した結果を援用することで, データが劣 Gauss 型独立標本や連続セミマルチンゲールの離散観測で与えられている場合に, (累積) 共分散行列の逆行列に対する一致推定量が得られる.

以下の議論で用いる行列に関する記号を導入しておく:

- $d \times d'$ 行列 $A = (A_{jk})_{1\leq j\leq d, 1\leq k\leq d'}$ に対して, A の ℓ_1-ノルムを $|A|_1$, 作用素ノルムを $\|A\|$ でそれぞれ表す:

$$|A|_1 := \sum_{j=1}^{d}\sum_{k=1}^{d'} |A_{jk}|, \qquad \|A\| := \sup_{x\in\mathbb{R}^{d'}:|x|\leq 1} |Ax|$$

- 2 つの d 次正方行列 A, B に対して, A と B の Hadamard 積を $A \circ B$ で表す. すなわち, $A \circ B$ は A と B の成分どうしの積からなる d 次正方行列である.
- 正方行列 A に対して, A の対角成分からなる対角行列を $\operatorname{diag}(A)$ で表す. また, $A^- := A - \operatorname{diag}(A)$ と定める.
- d 次対称行列全体の集合を $\mathcal{S}_d$ で表す. $A \in \mathcal{S}_d$ に対して, A の最大固有値と最小固有値をそれぞれ $\Lambda_{\max}(A), \Lambda_{\min}(A)$ で表す (A の固有値はすべて実数となることに注意).
- d 次正方行列 $A = (A_{jk})_{1 \le j,k \le d}$ に対して, $S(A) := \{(j,k) : A_{jk} \neq 0, j \neq k\}, s(A) := \#S(A)$ とおく. ここで, 有限集合 S に対してその要素数を $\#S$ で表す.

6.1　オラクル不等式

Lasso 型推定量の予測誤差や推定誤差の評価は, 多くの場合に非確率論的な部分と確率論的な部分に切り分けることができる (例えば (Bühlmann & van de Geer, 2011, Chapter 6) 参照). グラフィカル Lasso も例外ではなく, 推定量に特有の議論は非確率論的な部分のみであり, その部分は観測データの設定とは無関係に議論できるので, まずはその部分を取り扱う.

命題 6.2. $A_0, A \in \mathcal{S}_d$ とし, ある定数 $\lambda_0 > 0$ が存在して $|A - A_0|_\infty \le \lambda_0$ が成り立つとする. また, ある定数 $L > 1$ が存在して

$$L^{-1} \le \Lambda_{\min}(A_0) \le \Lambda_{\max}(A_0) \le L$$

が成り立つとし, $B_0 := A_0^{-1}$ とおく. さらに, $s := s(B_0), c_L := 8L^2$ とし, ある定数 $\lambda > 0$ に対して,

$$2\lambda_0 \le \lambda \le \frac{1}{4Lc_L}, \qquad 4c_L s\lambda^2 + c_L \| \operatorname{diag}(A) - \operatorname{diag}(A_0)\|_F^2 \le \lambda_0/(2L)$$

が成り立つと仮定する. このとき, $B \in \mathcal{S}_d^+$ が

$$\operatorname{tr}(BA) - \log\det(B) + \lambda|B^-|_1 \le \operatorname{tr}(B_0 A) - \log\det(B_0) + \lambda|B_0^-|_1, \tag{6.2}$$

を満たすならば,

$$\|B - B_0\|_F^2 / c_L + \lambda|B^- - B_0^-|_1 \le 4c_L s\lambda^2 + c_L \| \operatorname{diag}(A) - \operatorname{diag}(A_0)\|_F^2 \tag{6.3}$$

が成り立つ.

注意 6.1. (6.3) 式のような形の不等式は **(スパース度) オラクル不等式**と呼ばれる. これは, (6.3) 式の右辺は, 第 2 項を無視すれば,「非ゼロパラメータ数 (s) $\times$ (A_0 の) 推定誤差の二乗 (λ^2)」の定数倍という形をしているため, あたかも B_0 のゼロ成分を既知とみなして, 非ゼロ成分のみを誤差 λ で推定した場合の推定誤差を,「推定量」B が Frobenius ノルムの意味で達成できる, と不等式 (6.3) が主張しているとみなせることに由来する (オラクルとは「神託」という意味であり, 神しか知らないはずの真のモデルが既知と思える, というニュアンスがある). なお, 後述するように, 推定量を適切に構成することで, (6.3) 式の右辺第 2 項は実際に無視できる.

命題 6.2 の証明のために 2 つ補題を用意する.

補題 6.1. 関数 $\mathcal{S}_d^+ \ni A \mapsto -\log\det(A) \in \mathbb{R}$ は凸である.

証明. 任意に $A, B \in \mathcal{S}_d^+$ および $t \in [0,1]$ をとり, $C := A^{-1/2}BA^{-1/2}$ とおく. このとき,

$$\log\det((1-t)A + tB) = \log\det(A) + \log\det((1-t)I_d + tC)$$

となる. いま, C の固有値を $\lambda_1, \ldots, \lambda_d$ とすると,

$$-\log\det((1-t)I_d + tC) = -\sum_{j=1}^{d} \log\det((1-t) + t\lambda_j)$$

と書ける. $-\log$ は凸だから,

$$-\log\det((1-t)I_d + tC) \leq -t\sum_{j=1}^{d} \log\det(\lambda_j) = -t\log\det(C) = t\log\det(A) - t\log\det(B)$$

となる. 従って,

$$-\log\det((1-t)A + tB) \leq -(1-t)\log\det(A) - t\log\det(B)$$

を得る. □

補題 6.2 (Janková & van de Geer (2018), Lemma 14.4.1)**.** 命題 6.2 の仮定に加えて, $\Delta := B - B_0$ が $\|\Delta\|_F \leq 1/(2L)$ を満たすと仮定する. このとき, $\Delta + B_0$ は正定値となる. さらに,

$$\mathcal{E}(\Delta) := \mathrm{tr}\,(\Delta A_0) - \{\log\det\,(\Delta + B_0) - \log\det\,(B_0)\}$$

とおくと,

$$\mathcal{E}(\Delta) \geq c_L^{-1}\|\Delta\|_F^2 \tag{6.4}$$

が成り立つ.

証明. $D := B_0^{-1/2}\Delta B_0^{-1/2}$ とおく. 仮定より $\|\Delta\|_F \leq 1/(2L)$ であるから,

$$\|D\| \leq \|B_0^{-1}\|\|\Delta\| \leq \Lambda_{\min}(B_0)^{-1}\|\Delta\|_F \leq \frac{1}{2}$$

が成り立つ. 従って $I_d + D$ は正定値であるから, $B_0 + \Delta = B_0^{1/2}(I_d + D)B_0^{1/2}$ も正定値である.

次に, 任意の $t \in [0,1]$ に対して $I_d + tD$ は正定値であるから, 関数 $f : [0,1] \to \mathbb{R}$ を $f(t) = \log\det(I_d + tD)$ $(t \in [0,1])$ で定義することができる. D の固有値を $\mu_1, \ldots, \mu_d$ とすると,

$$f(t) = \sum_{j=1}^{d} \log(1 + t\mu_j)$$

と書けるから,

$$f'(t) = \sum_{j=1}^{d} \frac{\mu_j}{1+t\mu_j}, \qquad f''(t) = -\sum_{j=1}^{d} \frac{\mu_j^2}{(1+t\mu_j)^2}$$

が成り立つ. 特に, $f'(0) = \mathrm{tr}(D) = \mathrm{tr}(\Delta A_0)$ である. また,

$$\begin{aligned} f(1) - f(0) &= \log\left(\det(B_0)^{-1/2}\det(B_0+\Delta)\det(B_0)^{-1/2}\right) - \log\det(I_d) \\ &= \log\det(\Delta + B_0) - \log\det(B_0) \end{aligned}$$

である. 従って, 部分積分法により

$$\mathcal{E}(\Delta) = f'(0) - \{f(1) - f(0)\} = -\int_0^1 (1-t) f''(t) dt$$

を得る. 故に,

$$\mathcal{E}(\Delta) \geq \|D\|_F^2 \min_{t\in[0,1]} \min_{1\leq j\leq d} \frac{1}{(1+t\mu_j)^2}$$

が成り立つ. ここで,

$$\|\Delta\|_F = \left\|B_0^{1/2} D B_0^{1/2}\right\|_F \leq \|B_0\| \, \|D\|_F \leq L\, \|D\|_F$$

だから, $\|D\|_F \geq L^{-1}\|\Delta\|_F$ であり, また, 任意の $t \in [0,1]$ に対して,

$$\min_{1\leq j\leq d} \frac{1}{(1+t\mu_j)^2} = \Lambda_{\mathrm{min}}((I_d + tD)^{-1})^2 = \Lambda_{\mathrm{max}}(I_d+tD)^{-2} \geq \frac{4}{9} \geq \frac{1}{8}$$

が成り立つ. □

命題 6.2 の証明. 2 ステップに分けて証明する.

Step 1. まず, $\Delta := B - B_0$ が $\|\Delta\|_F \leq 1/(2L)$ を満たす場合に命題の結論が成り立つことを示す. このとき, $\mathcal{E}(\Delta)$ を補題 6.2 のように定義することができて, 不等式 (6.4) が成り立つ. さらに, (6.2) より,

$$\begin{aligned} &\mathcal{E}(\Delta) + \lambda|B^-|_1 \\ &= -\mathrm{tr}\left(\Delta(A - A_0)\right) + \left\{\mathrm{tr}\left(BA\right) - \log\det(B) + \lambda|B^-|_1\right\} - \mathrm{tr}\left(B_0 A\right) + \log\det(B_0) \\ &\leq -\mathrm{tr}\left(\Delta(A - A_0)\right) + \left\{\mathrm{tr}\left(B_0 A\right) - \log\det(B_0) + \lambda|B_0^-|_1\right\} - \mathrm{tr}\left(B_0 A\right) + \log\det(B_0) \\ &= -\mathrm{tr}\left(\Delta(A - A_0)\right) + \lambda|B_0^-|_1 \end{aligned} \tag{6.5}$$

が成り立つ. 任意の $A_1, B_1 \in \mathcal{S}_d$ に対して $\mathrm{tr}(A_1B_1) = \mathrm{tr}(A_1^- B_1^-) + \mathrm{tr}(\mathrm{diag}(A_1)\,\mathrm{diag}(B_1))$ および $|\mathrm{tr}(A_1B_1)| \leq |A_1 \circ B_1|_1$ となることに注意すると,

$$\begin{aligned} |\mathrm{tr}\left(\Delta(A - A_0)\right)| &\leq |\Delta^-|_1 |A^- - A_0^-|_\infty + \|\mathrm{diag}(\Delta)\|_F \|\mathrm{diag}(A) - \mathrm{diag}(A_0)\|_F \\ &\leq \lambda_0 |\Delta^-|_1 + \|\mathrm{diag}(A) - \mathrm{diag}(A_0)\|_F \|\mathrm{diag}(\Delta)\|_F \end{aligned}$$

が成り立つ．ここで, 2 番目の不等式の導出に仮定 $|A-A_0|_\infty \le \lambda_0$ を用いた．この不等式と (6.4)–(6.5) を組み合わせて,

$$c_L^{-1}\|\Delta\|_F^2 + \lambda|B^-|_1 \le \lambda_0|\Delta^-|_1 + \|\operatorname{diag}(A)-\operatorname{diag}(A_0)\|_F\|\operatorname{diag}(\Delta)\|_F + \lambda|B_0^-|_1$$

を得る.

$S := S(B_0)$ とおく．さらに, $I \subset \{1,\dots,d\}^2$ と d 次正方行列 $U = (U^{ij})_{1\le i,j\le d}$ に対して, d 次正方行列 $U_I = (U_I^{ij})_{1\le i,j\le d}$ を $U_I^{ij} = U^{ij}1_{\{(i,j)\in I\}}$ で定める．このとき, 定義と仮定より, $|B^-|_1 = |B_S^-|_1 + |B_{S^c}^-|_1, |\Delta^-|_1 = |\Delta_S^-|_1 + |B_{S^c}^-|_1, |B_0^-|_1 \le |\Delta_S^-|_1 + |B_S^-|_1$ および $\lambda \ge 2\lambda_0$ が成り立つ. 従って,

$$c_L^{-1}\|\Delta\|_F^2 + \frac{\lambda}{2}|B_{S^c}^-|_1 \le \frac{3\lambda}{2}|\Delta_S^-|_1 + \|\operatorname{diag}(A)-\operatorname{diag}(A_0)\|_F\|\operatorname{diag}(\Delta)\|_F$$

が成り立つ. 以上より,

$$\begin{aligned}
&2c_L^{-1}\|\Delta\|_F^2 + \lambda|\Delta^-|_1 \\
&= 2c_L^{-1}\|\Delta\|_F^2 + \lambda(|B_{S^c}^-|_1 + |\Delta_S^-|_1) \\
&\le 4\lambda|\Delta_S^-|_1 + 2\|\operatorname{diag}(A)-\operatorname{diag}(A_0)\|_F\|\operatorname{diag}(\Delta)\|_F \\
&\le 4\lambda\sqrt{s}\|\Delta_S^-\|_F + 2\|\operatorname{diag}(A)-\operatorname{diag}(A_0)\|_F\|\operatorname{diag}(\Delta)\|_F \quad (\because \text{Schwarz の不等式}) \\
&\le 4s\lambda^2c_L + c_L^{-1}\|\Delta_S^-\|_F^2 + c_L\|\operatorname{diag}(A)-\operatorname{diag}(A_0)\|_F^2 + c_L^{-1}\|\operatorname{diag}(\Delta)\|_F^2 \\
&\hfill (\because \text{相加相乗平均の不等式})
\end{aligned}$$

を得る. $\|\Delta\|_F^2 = \|\operatorname{diag}(\Delta)\|_F^2 + \|\Delta^-\|_F^2$ であるから,

$$c_L^{-1}\|\Delta\|_F^2 + \lambda|\Delta^-|_1 \le 4s\lambda^2c_L + c_L\|\operatorname{diag}(A)-\operatorname{diag}(A_0)\|_F^2$$

を得る.

Step 2. 一般の場合に示す. Step 1 より, $\|B-B_0\|_F \le 1/(2L)$ を示せば十分である.

$M = 1/(2L)$, $\alpha = M/(M+\|B-B_0\|_F)$ とし, $\tilde{B} = \alpha B + (1-\alpha)B_0$ とおく．定義より $\|\tilde{B}-B_0\|_F \le M = 1/(2L)$ である. さらに, 補題 6.1 および (6.2) より,

$$\operatorname{tr}(\tilde{B}A) - \log\det(\tilde{B}) + \lambda|B^-|_1 \le \operatorname{tr}(B_0A) - \log\det(B_0) + \lambda|B_0^-|_1$$

が成り立つ. 従って, Step 1 より

$$\|\tilde{B}-B_0\|_F^2/c_L + \lambda|\tilde{B}^- - B_0^-|_1 \le 4c_Ls\lambda^2 + c_L\|\operatorname{diag}(A)-\operatorname{diag}(A_0)\|_F^2$$

が成り立つ. 特に, 仮定より,

$$\|\tilde{B}-B_0\|_F^2 \le c_L\lambda_0/(2L) \le 1/(16L^2)$$

が成り立つから, $\|\tilde{B}-B_0\|_F \le 1/(4L) = M/2$ である. $\tilde{B}$ の定義より, これは $\|B-B_0\|_F \le M = 1/(2L)$ を意味する. □

6.2 劣 Gauss 型独立標本の場合

$X_1, \ldots, X_n$ を独立同分布な確率変数列とし, 平均 0, 共分散行列 $\Sigma = (\Sigma_{jk})_{1 \leq j,k \leq d}$ をもち, かつ, n に依存しない定数 $\kappa > 0$ が存在して $\max_{1 \leq j \leq d} \|X_{1j}\|_{\psi_2} \leq \kappa$ を満たすとする. Σ は正則であると仮定し, $X_1, \ldots, X_n$ が観測データとして与えられた際に, 逆行列 $\Theta_0 := \Sigma^{-1}$ を推定する問題を考える. ここで, 次元 d は n に依存することを許す.

まず, 標本共分散行列

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^\top$$

に基づくグラフィカル Lasso 推定量のパフォーマンスを調べる. 各 $\lambda > 0$ に対して, (6.1) を最小化する $\Theta \in \mathcal{S}_d^+$ を $\hat{\Theta}_\lambda$ と書くことにする.

定理 6.1. n に依存しないある定数 $L > 1$ が存在して

$$L^{-1} \leq \Lambda_{\min}(\Theta_0) \leq \Lambda_{\max}(\Theta_0) \leq L \tag{6.6}$$

が成り立つと仮定する. また, $\log d = o(n)$ $(n \to \infty)$ を仮定する. このとき, 正数列 λ_n が $\lambda_n^{-1}\sqrt{(\log d)/n} \to 0$ $(n \to \infty)$ を満たし, かつ $s_n := s(\Theta_0)$ に対して $(s_n + d)\lambda_n \to 0$ $(n \to \infty)$ を満たすならば,

$$\left\| \hat{\Theta}_{\lambda_n} - \Theta_0 \right\|_F = O_p\left(\lambda_n \sqrt{s_n + d} \right)$$

が成り立つ.

証明. $E_n := \{|\hat{\Sigma}_n - \Sigma|_\infty \leq \lambda_n/2\}$ とおく. 命題 3.2 と Markov の不等式, および仮定より, $\mathrm{P}(E_n^c) \to 0$ $(n \to \infty)$ が成り立つ. また, E_n 上で $\|\operatorname{diag}(\hat{\Sigma}_n) - \operatorname{diag}(\Sigma)\|_F^2 \leq d\lambda_n^2/4$ が成り立つから, n が十分大きいとき, E_n 上で

$$8L^2(4s_n\lambda_n^2 + d\|\operatorname{diag}(\hat{\Sigma}_n) - \operatorname{diag}(\Sigma)\|_F^2/4) \leq \lambda_n/(4L)$$

が成り立つ. 従って, $\lambda_0 = \lambda_n/2, \lambda = \lambda_n$ として命題 6.2 を適用すると, E_n 上で

$$\left\| \hat{\Theta}_{\lambda_n} - \Theta_0 \right\|_F^2 /(8L^2) \leq 8L^2(4s_n\lambda_n^2 + d\lambda_n^2/4)$$

が成り立つ. 従って,

$$\begin{aligned}
&\limsup_{M\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\left\| \hat{\Theta}_{\lambda_n} - \Theta_0 \right\|_F > M\left(\lambda_n \sqrt{s_n + d} \right) \right) \\
&\leq \limsup_{M\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(64L^4\lambda_n^2(4s_n + d/4) > M\left(\lambda_n^2(s_n + d) \right) \right) = 0
\end{aligned}$$

となる. □

定理 6.1 より, $\hat{\Theta}_{\lambda_n}$ の収束レートを改善するには λ_n をできる限り小さくとる必要があるが, 一方で $\lambda_n^{-1}\sqrt{(\log d)/n} = o(1)$ という条件も満たす必要がある. 従って, λ_n は $\sqrt{(\log d)/n}$ よりわずかに大きいオーダーで取るのがベストだということになる. この場合の $\hat{\Theta}_{\lambda_n}$ の収束レートは

$O(\sqrt{(s_n+d)(\log d)/n})$ となるが, これが 0 に収束するためには少なくとも $d \ll n$ でなければならず, 特に次元がサンプル数よりも大きいような状況に対応できない. この問題は, 標本共分散行列ではなく標本相関行列 $\hat{R}_n := \operatorname{diag}(\hat{\Sigma}_n)^{-1/2}\hat{\Sigma}_n \operatorname{diag}(\hat{\Sigma}_n)^{-1/2}$ に基づくグラフィカル Lasso 推定量を考えることで解消できる (Rothman et al., 2008). 各 $\lambda > 0$ に対して,

$$-\ell_n(K;\hat{R}_n) + \lambda \sum_{j,k:j\neq k} |K_{jk}|$$

を最小化する $K \in \mathcal{S}_d^+$ を $\hat{K}_\lambda$ と書くことにする. $\hat{K}_\lambda$ は $K_0 := \operatorname{diag}(\Sigma)^{1/2}\Theta_0 \operatorname{diag}(\Sigma)^{1/2}$ を推定することが期待されるので, Θ_0 を推定するには,

$$\tilde{\Theta}_\lambda := \operatorname{diag}(\hat{\Sigma}_n)^{-1/2}\hat{K}_\lambda \operatorname{diag}(\hat{\Sigma}_n)^{-1/2}$$

を考えるのが自然である.

定理 6.2. n に依存しないある定数 $L > 1$ が存在して (6.6) が成り立つと仮定する. また, $\log d = o(n)$ $(n \to \infty)$ を仮定する. このとき, 正数列 λ_n が $\lambda_n^{-1}\sqrt{(\log d)/n} \to 0$ $(n \to \infty)$ を満たし, かつ $s_n := s(\Theta_0)$ に対して $(s_n+1)\lambda_n \to 0$ $(n \to \infty)$ を満たすならば,

$$\left\|\hat{K}_{\lambda_n} - K_0\right\|_F = O_p\left(\lambda_n\sqrt{s_n+1}\right) \tag{6.7}$$

および

$$\left\|\tilde{\Theta}_{\lambda_n} - \Theta_0\right\| = O_p\left(\lambda_n\sqrt{s_n+1}\right) \tag{6.8}$$

が成り立つ.

証明. 命題 3.2 と Markov の不等式より $|\hat{\Sigma}_n - \Sigma|_\infty = O_p(\sqrt{(\log d)/n})$ が成り立つ. また, (6.6) より $L^{-1} \leq \min_j \Sigma_{jj} \leq \max_j \Sigma_{jj} \leq L$ が成り立つことに注意すると, $R_0 := \operatorname{diag}(\Sigma)^{-1/2}\Sigma \operatorname{diag}(\Sigma)^{-1/2}$ として $|\hat{R}_n - R_0|_\infty = O_p(\sqrt{(\log d)/n})$ が成り立つ. 従って, $E_n := \{|\hat{R}_n - R_0|_\infty \leq \lambda_n/2\}$ とおくと, 仮定より, $\mathrm{P}(E_n^c) \to 0$ $(n \to \infty)$ が成り立つ. また, n が十分大きいとき,

$$8L^2(4s_n\lambda_n^2 + d\|\operatorname{diag}(\hat{R}_n) - \operatorname{diag}(R_0)\|_F^2/4) = 16L^2 s_n\lambda_n^2 \leq \lambda_n/(4L)$$

が成り立つ. 従って, $\lambda_0 = \lambda_n/2, \lambda = \lambda_n$ として命題 6.2 を適用すると, E_n 上で

$$\left\|\hat{K}_{\lambda_n} - K_0\right\|_F^2/(8L^2) \leq 16L^2 s_n\lambda_n^2$$

が成り立つ. 以上より (6.7) が従う. さらに,

$$\begin{aligned}\left\|\tilde{\Theta}_{\lambda_n} - \Theta_0\right\| \leq &\left\|\operatorname{diag}(\hat{\Sigma}_n)^{-1/2}\right\| \left\|\hat{K}_{\lambda_n} - K_0\right\| \left\|\operatorname{diag}(\hat{\Sigma}_n)^{-1/2}\right\| \\ &+ \left\|\operatorname{diag}(\hat{\Sigma}_n)^{-1/2} - \operatorname{diag}(\Sigma)^{-1/2}\right\| \|K_0\| \left\|\operatorname{diag}(\hat{\Sigma}_n)^{-1/2}\right\| \\ &+ \left\|\operatorname{diag}(\Sigma)^{-1/2}\right\| \|K_0\| \left\|\operatorname{diag}(\hat{\Sigma}_n)^{-1/2} - \operatorname{diag}(\Sigma)^{-1/2}\right\|\end{aligned}$$

より (6.8) が従う. □

注意 6.2. $\|\operatorname{diag}(\hat{\Sigma}_n)^{-1/2}-\operatorname{diag}(\Sigma)^{-1/2}\|_F=O(\sqrt{d/n})$ であるので, Frobenius ノルムで計測した $\tilde{\Theta}_{\lambda_n}$ の推定誤差の収束レートは, $\hat{\Theta}_{\lambda_n}$ と ($\log d$ の項を無視すると) 同じオーダーしか得られない.

注意 6.3. $\tilde{\Theta}_\lambda$ は,

$$-\ell_n(\Theta;\hat{\Sigma}_n)+\lambda\sum_{j,k:j\neq k}\sqrt{\hat{\Sigma}_{n,jj}\hat{\Sigma}_{n,kk}}|\Theta_{jk}|$$

を最小化する $\Theta\in\mathcal{S}_d^+$ となっている. すなわち, (6.1) の罰則項に重みをつけた最小化問題の解として得られる. この意味で, $\Theta\in\mathcal{S}_d^+$ は**重み付きグラフィカル Lasso 推定量**と呼ばれることがある.

6.3 連続セミマルチンゲールの場合

4.2 節と同じ設定を考える. 累積共分散行列 $[X,X]_T$ が確率 1 で正則であると仮定して, その逆行列 $\Theta_0:=[X,X]_T^{-1}$ を推定する問題を考える. ここで, 次元 d は n に依存することを許すが, 観測区間の終点 T は n に依存しない状況を考える.

上述したように, (6.1) 式において $\hat{\Sigma}_n$ を $\widehat{[X,X]}_T$ に置き換えたものを最小にする $\Theta\in\mathcal{S}_d^+$ を考えることで, いまの設定でも Θ_0 のグラフィカル Lasso 推定量を構成できる. 前小節で見たように, 標本共分散行列の代わりに標本相関行列を考えた方が理論上のパフォーマンスがよい推定量が得られるので, ここではそのような推定量のみ考察する. すなわち,

$$\hat{R}_X:=\operatorname{diag}(\widehat{[X,X]}_T)^{-1/2}\widehat{[X,X]}_T\operatorname{diag}(\widehat{[X,X]}_T)^{-1/2}$$

として, 各 $\lambda>0$ に対して,

$$-\ell_n(K;\hat{R}_X)+\lambda\sum_{j,k:j\neq k}|K_{jk}|$$

を最小化する $K\in\mathcal{S}_d^+$ を $\hat{K}_\lambda$ と書くことにする. Θ_0 の推定量は,

$$\tilde{\Theta}_\lambda:=\operatorname{diag}(\widehat{[X,X]}_T)^{-1/2}\hat{K}_\lambda\operatorname{diag}(\widehat{[X,X]}_T)^{-1/2}$$

で与えられる. 高頻度データ解析の文脈では, $\hat{R}_X$ は**実現相関行列**と呼ばれる.

定理 6.3. $n\to\infty$ のとき

$$\Lambda_{\max}([X,X]_T)+1/\Lambda_{\min}([X,X]_T)=O_p(1) \tag{6.9}$$

が成り立つと仮定する. また, $\log d=o(n)$ $(n\to\infty)$ を仮定する. このとき, 正数列 λ_n が $\lambda_n^{-1}\sqrt{(\log d)/n}\to 0$ $(n\to\infty)$ を満たし, かつ $s_n:=s(\Theta_0)$ に対して $(s_n+1)\lambda_n=o_p(1)$ $(n\to\infty)$ を満たすならば, 系 5.1 と同じ仮定の下で,

$$\left\|\hat{K}_{\lambda_n}-R_X^{-1}\right\|_F=O_p\left(\lambda_n\sqrt{s_n+1}\right) \tag{6.10}$$

および

$$\left\|\tilde{\Theta}_{\lambda_n}-\Theta_0\right\|=O_p\left(\lambda_n\sqrt{s_n+1}\right) \tag{6.11}$$

が成り立つ. ここに, $R_X:=\operatorname{diag}([X,X]_T)^{-1/2}[X,X]_T\operatorname{diag}([X,X]_T)^{-1/2}$ である.

証明. まず, (6.9) より

$$\max_{1\le j\le d}\left([X^j,X^j]_T+1/[X^j,X^j]_T\right)=O_p(1) \tag{6.12}$$

が成り立つ. このことと系 5.1 より,

$$|\hat{R}_X-R_X|_\infty=O_p\left(\sqrt{\frac{\log d}{n}}\right) \tag{6.13}$$

が成り立つ. 次に, 各 $L,n\in\mathbb{N}$ に対して, 集合 $\Omega_{n,L}\subset\Omega$ を

$$\Omega_{n,L}:=\{|\hat{R}_X-R_X|_\infty\le\lambda_n/2\}\cap\{L^{-1}\le\Lambda_{\min}(R_X)\le\Lambda_{\max}(R_X)\le L\}\cap\{4c_Ls_n\lambda_n\le 1/(4L)\}$$

で定める. ただし, $c_L:=8L^2$ と定める. このとき, (6.9), (6.12), (6.13) および $s_n\lambda_n=o_p(1)$ より

$$\lim_{L\to\infty}\limsup_{n\to\infty}\mathrm{P}(\Omega_{n,L}^c)=0$$

となる. $\Omega_{n,L}$ 上で $\lambda:=\lambda_n,\lambda_0:=\lambda_n/2$ として命題 6.2 を適用すると,

$$\|\hat{K}_{\lambda_n}-R_X^{-1}\|_F^2/c_L\le 4c_Ls_n\lambda_n^2\qquad\text{on }\Omega_{n,L}.$$

を得る. 従って,

$$\limsup_{n\to\infty}\mathrm{P}\left(\|\hat{K}_{\lambda_n}-R_X^{-1}\|_F>16L^2\sqrt{s_n}\lambda_n\right)\le\limsup_{n\to\infty}\mathrm{P}(\Omega_{n,L}^c)$$

となるから,

$$\limsup_{M\to\infty}\limsup_{n\to\infty}\mathrm{P}\left(\|\hat{K}_{\lambda_n}-R_X^{-1}\|_F>M\sqrt{s_n}\lambda_n\right)\le\limsup_{L\to\infty}\limsup_{n\to\infty}\mathrm{P}(\Omega_{n,L}^c)=0$$

を得る. すなわち, (6.10) が成り立つ. (6.11) は (6.8) と同様の議論で示せる. □

7 作用素ノルムによる評価

4.2 節と同じ設定を考える.

定理 7.1. ある定数 $\Lambda>0$ と P-測度ゼロ集合 N が存在して, 任意の $s\in[0,T]$ に対して

$$|b_s|_\infty^2\vee\|c_s\|\le\Lambda\qquad\text{on}\quad\Omega\setminus N$$

が成り立つと仮定する. このとき,

$$\left\|\left\|\widehat{[X,X]}_T-[X,X]_T\right\|\right\|_{\psi_1}\le C\Lambda\left((T^2+T)\frac{d}{n}+(T^{3/2}+T)\sqrt{\frac{d}{n}}\right)$$

が成り立つ.

定理 7.1 は, 作用素ノルムを有限個の 2 次形式の最大値で抑えて, 最大値ノルムの評価に帰着することで証明する. このために「ε-ネットの方法」と呼ばれる論法を用いる.

定義 7.1 (ネット・被覆数)**.** (S,ρ) を距離空間とする. $S_0 \subset S$ と $\varepsilon > 0$ に対して, $\mathcal{N} \subset S_0$ が (ρ に関する) S_0 の ε-**ネット**であるとは, 任意の $x \in S_0$ に対してある $y \in \mathcal{N}$ が存在して $\rho(x,y) \leq \varepsilon$ が成り立つことをいう. S_0 の ε-ネットの要素数の最小値を (ρ に関する) S_0 の ε-**被覆数**と呼び, 記号 $\mathcal{N}(S_0,\rho,\varepsilon)$ で表す. ただし, S_0 が有限な ε-ネットを含まない場合は $\mathcal{N}(S_0,\rho,\varepsilon) := \infty$ と定める.

定義 7.2 (パッキング数)**.** (S,ρ) を距離空間とする. $\varepsilon > 0$ に対して, $\mathcal{N} \subset S$ が (ρ に関して) ε-**separated** であるとは, 任意の相異なる 2 点 $x,y \in \mathcal{N}$ に対して $\rho(x,y) > \varepsilon$ が成り立つことをいう. $S_0 \subset S$ の ε-separated な部分集合の要素数の最大値を (ρ に関する) S_0 の ε-**パッキング数**と呼び, 記号 $\mathcal{P}(S_0,\rho,\varepsilon)$ で表す. ただし, S_0 が ε-separated な無限部分集合を含む場合は $\mathcal{P}(S_0,\rho,\varepsilon) := \infty$ と定める.

補題 7.1. (S,ρ) を距離空間とし, $S_0 \subset S, \varepsilon > 0$ とする. $N := \mathcal{P}(S_0,\rho,\varepsilon) < \infty$ ならば, 要素数 N の ε-separated な S_0 の部分集合は S_0 の ε-ネットとなる.

証明. $\mathcal{N} \subset S_0$ が要素数 N かつ ε-separated であったとする. 任意の $x \in S_0 \setminus \mathcal{N}$ に対して, N の最大性より $\mathcal{N} \cup \{x\}$ は ε-separated でないから, ある $y \in \mathcal{N}$ が存在して $\rho(x,y) \leq \varepsilon$ となる. 従って $\mathcal{N}$ は S_0 の ε-ネットである. □

補題 7.2 (Vershynin (2018), Lemma 4.2.8)**.** (S,ρ) を距離空間とする. 任意の $S_0 \subset S$ と $\varepsilon > 0$ に対して,

$$\mathcal{N}(S_0,\rho,\varepsilon) \leq \mathcal{P}(S_0,\rho,\varepsilon) \leq \mathcal{N}(S_0,\rho,\varepsilon/2).$$

証明. 左側の不等式は補題 7.1 から従う. 右側の不等式を背理法で示す. $N := \mathcal{N}(S_0,\rho,\varepsilon/2) < \mathcal{P}(S_0,\rho,\varepsilon)$ であったとすると, S_0 は要素数 N の $\varepsilon/2$-ネット $\mathcal{N}$ を含み, かつ要素数 $N+1$ の ε-separated な部分集合 $\mathcal{P}$ も含む. $\mathcal{P} \subset S_0 \subset \bigcup_{x\in\mathcal{N}}\{y \in S : \rho(y,x) \leq \varepsilon/2\}$ であるから, 鳩の巣原理よりある $x \in \mathcal{N}$ に対して $\{y \in S : \rho(y,x) \leq \varepsilon/2\}$ は $\mathcal{P}$ の相異なる 2 点 y,z を含む. しかし, このとき $\rho(y,z) \leq \rho(y,x) + \rho(x,z) \leq \varepsilon$ となって $\mathcal{P}$ が ε-separated であることに矛盾する. □

以下, $\mathbb{R}^d$ の部分集合に対するネットや ε-separated な部分集合を考える場合, 特に断らない限りは Euclid 距離に関するものを考えることにする. さらに, 記号の簡単のために, $S \subset \mathbb{R}^d$ と $\varepsilon > 0$ に対して, Euclid 距離に関する S の ε-被覆数および ε-パッキング数をそれぞれ記号 $\mathcal{N}(S,\varepsilon)$ および $\mathcal{P}(S,\varepsilon)$ で表す.

$\mathbb{R}^d$ 内の単位球面を $\mathbb{S}^{d-1}$ と書く:

$$\mathbb{S}^{d-1} := \{x \in \mathbb{R}^d : |x| = 1\}.$$

補題 7.3 (Vershynin (2018), Corollary 4.2.13)**.** 任意の $\varepsilon > 0$ に対して,

$$\mathcal{N}(\mathbb{S}^{d-1},\varepsilon) \leq \mathcal{P}(\mathbb{S}^{d-1},\varepsilon) \leq \left(1 + \frac{2}{\varepsilon}\right)^d.$$

証明. 左側の不等式は補題 7.2 から従うので, 右側の不等式を示す. 証明には「volume argument」と呼ばれる論法を用いる. まず, $\mathbb{S}^{d-1}$ はコンパクトだから全有界である. 従って $\mathcal{N}(\mathbb{S}^{d-1},\varepsilon/2) < \infty$

だから, 補題 7.2 より $N := \mathcal{P}(\mathbb{S}^{d-1}, \varepsilon) < \infty$ である. 定義より $\mathbb{S}^{d-1}$ の ε-separated な部分集合 $\mathcal{N}$ で要素数 N のものがとれる.

$$\#\mathcal{N} \le \left(1 + \frac{2}{\varepsilon}\right)^d \tag{7.1}$$

を示せば証明は完成する.

各 $x \in \mathbb{R}^d$ と $\delta > 0$ について, x を中心とする半径 δ の開球を $B(x;\delta)$ と書くことにする: $B(x;\delta) = \{y \in \mathbb{R}^d : |y - x| < \delta\}$. 任意の $x \in \mathbb{S}^{d-1}$ に対して $B(x;\varepsilon/2) \subset B(0;1+\varepsilon/2)$ が成り立つ. 実際, 任意の $y \in B(x;\varepsilon/2)$ に対して $|y| \le |y - x| + 1 < \varepsilon/2 + 1$ である. 従って $\bigcup_{x\in\mathcal{N}} B(x;\varepsilon/2) \subset B(0;1+\varepsilon/2)$ が成り立つ. 従って, 可測集合 $F \subset \mathbb{R}^d$ の Lebesgue 測度を $|F|$ と書くことにすると,

$$\left|\bigcup_{x\in\mathcal{N}} B(x;\varepsilon/2)\right| \le |B(0;1+\varepsilon/2)| = (1+\varepsilon/2)^d |B(0;1)|$$

が成り立つ. さらに, $\mathcal{N}$ の任意の相異なる 2 点 x, y に対して $B(x;\varepsilon/2) \cap B(y;\varepsilon/2) = \emptyset$ が成り立つ. 実際, そうでない場合, $z \in B(x;\varepsilon/2) \cap B(y;\varepsilon/2)$ なる点 z を取れるが, このとき $|x - y| \le |x - z| + |z - y| < \varepsilon$ となって $\mathcal{N}$ が ε-separated であることに矛盾する. 故に,

$$\left|\bigcup_{x\in\mathcal{N}} B(x;\varepsilon/2)\right| = \sum_{x\in\mathcal{N}} |B(x;\varepsilon/2)| = \#\mathcal{N} \cdot |B(0;\varepsilon/2)| = \#\mathcal{N} \cdot (\varepsilon/2)^d |B(0;1)|$$

が成り立つ. 以上より,

$$\#\mathcal{N} \cdot (\varepsilon/2)^d |B(0;1)| \le (1+\varepsilon/2)^d |B(0;1)|$$

であるから, 両辺を $(\varepsilon/2)^d |B(0;1)|$ で割って (7.1) を得る. □

補題 7.4 (Vershynin (2018), Exercise 4.4.3(b))**.** $\varepsilon \in (0, 1/2)$ とし, $\mathcal{N}$ を $\mathbb{S}^{d-1}$ の ε-ネットとする. このとき, 任意の d 次対称行列 A に対して,

$$\|A\| \le \frac{1}{1 - 2\varepsilon} \sup_{x\in\mathcal{N}} |x^\top A x|.$$

証明. A は対称だから, ある $x \in \mathbb{S}^{d-1}$ が存在して $\|A\| = |x^\top A x|$ となる. 実際, A の固有値のうち絶対値が最大となるようなものに対する固有ベクトルで長さ 1 のものがそのような x を与える. $\mathcal{N}$ は $\mathbb{S}^{d-1}$ の ε-ネットだから, ある $y \in \mathcal{N}$ が存在して $|x - y| \le \varepsilon$ となる. このとき,

$$\|A\| = |x^\top A x| \le |x - y||Ax| + |Ay||x - y| + |y^\top A y| \le 2\varepsilon\|A\| + \sup_{x\in\mathcal{N}} |x^\top A x|$$

となるので, 整理して示すべき不等式を得る. □

定理 7.1 の証明. 2 ステップに分けて証明する.

Step 1. 定理 4.6 の証明と同じように記号を定義して, 分解 (4.5) を考える. $R := \widehat{[M, M]}_T - [M, M]_T$ とおくと,

$$\left\|\widehat{[X, X]}_T - [X, X]_T\right\| \le \left\|\widehat{[A, A]}_T\right\| + 2\left\|\widehat{[A, M]}_T\right\| + \|R\|$$

と評価できる. 仮定より,

$$\left\|\widehat{[A,A]}_T\right\| \le \sum_{i=1}^n \left\|\Delta_i^n A(\Delta_i^n A)^\top\right\| = \sum_{i=1}^n |\Delta_i^n A|^2 \le \Lambda T^2 \frac{d}{n}$$

および

$$\left\|\widehat{[M,M]}_T\right\| \le \|R\| + \|[M,M]_T\| \le \|R\| + \Lambda T$$

が成り立つ. また, d 次元確率過程 U に対して $\Delta^n U := (\Delta_1^n U, \dots, \Delta_n^n U)$ と書くことにすると, $\widehat{[A,M]}_T = \Delta^n A(\Delta^n M)^\top$ と書き直せるから,

$$\begin{aligned}\left\|\widehat{[A,M]}_T\right\| &\le \|\Delta^n A\|\,\|\Delta^n M\| = \sqrt{\left\|\widehat{[A,A]}_T\right\|\left\|\widehat{[M,M]}_T\right\|} \\ &\le \sqrt{\Lambda T^2 \frac{d}{n}\|R\|} + \Lambda T^{3/2}\sqrt{\frac{d}{n}} \le \frac{\Lambda T^2}{2}\frac{d}{n} + \frac{\|R\|}{2} + \Lambda T^{3/2}\sqrt{\frac{d}{n}}\end{aligned}$$

を得る. 以上より,

$$\left\|\widehat{[X,X]}_T - [X,X]_T\right\| \le 2\Lambda T^2 \frac{d}{n} + 2\Lambda T^{3/2}\sqrt{\frac{d}{n}} + 2\,\|R\|. \tag{7.2}$$

Step 2. 補題 7.3 と 7.4 を $\varepsilon = 1/4$ として適用すると, ある空でない集合 $\mathcal{N} \subset \mathbb{S}^{d-1}$ が存在して $\#\mathcal{N} \le 9^d$ および

$$\|B\| \le 2\max_{x\in\mathcal{N}} |x^\top B x| \tag{7.3}$$

が任意の d 次対称行列 B に対して成り立つ.

次に, $x \in \mathcal{N}$ を 1 つ固定し, 確率過程 $M^x = (M_t^x)_{t\ge0}$ を $M_t^x = x^\top M_t$ $(t \ge 0)$ で定める. そして, 各 $i = 1, \dots, n$ について

$$\xi_i^x := x^\top \left\{(\Delta_i^n M)(\Delta_i^n M)^\top - \Delta_i^n [M,M]\right\} x = (\Delta_i^n M^x)^2 - \Delta_i^n [M^x, M^x]$$

とおく. 定義より

$$x^\top R x = \sum_{i=1}^n \xi_i^x$$

となる. 定理 3.1 を用いて右辺を評価する. まず, 部分積分公式より

$$\xi_i^x = 2\int_{t_{i-1}}^{t_i} (M_s^x - M_{t_{i-1}}^x) dM_s^x$$

と書ける.

$$\int_0^T x^\top c_s x ds \le \Lambda T$$

であるから, 命題 4.6 より $(M^x_{t\wedge T})_{t\geq 0}$ は連続マルチンゲールである. さらに,

$$\mathrm{E}\left[\int_{t_{i-1}}^{t_i}(M^x_s - M^x_{t_{i-1}})^2 d\langle M^x, M^x\rangle_s\right] \leq \Lambda\,\mathrm{E}\left[\int_{t_{i-1}}^{t_i}(M^x_s - M^x_{t_{i-1}})^2 ds\right]$$

$$= \Lambda\,\mathrm{E}\left[\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{s} x^\top c_u x du\right) ds\right] \leq \frac{\Lambda^2 T^2}{n^2}$$

が成り立つから, 再び命題 4.6 より $(\xi_i)_{i=1}^n$ は $(\mathcal{F}_{t_i})_{i=0}^n$ に関するマルチンゲール差分列である. 次に, 任意の整数 $p \geq 2, i = 1, \dots, n$ と $F \in \mathcal{F}_{t_{i-1}}$ に対して,

$$\|\xi^x_i 1_F\|_p \leq \left\|(\Delta^n_i M^x)^2 1_F\right\|_p + \|\Delta^n_i [M^x, M^x] 1_F\|_p$$

が成り立つ. 仮定より,

$$\|\Delta^n_i [M^x, M^x] 1_F\|_p = \left\|\int_{t_{i-1}}^{t_i} x^\top c_s x ds 1_F\right\|_p \leq \frac{\Lambda T}{n}\|1_F\|_p$$

である. 一方で, 定理 5.1 と仮定より,

$$\left\|(\Delta^n_i M^x)^2 1_F\right\|_p = \|\Delta^n_i M^x 1_F\|_{2p}^2 \lesssim p\left\|\sqrt{\int_{t_{i-1}}^{t_i} x^\top c_s x ds} 1_F\right\|_p^2 \leq \frac{p\Lambda T}{n}\|1_F\|_p$$

が成り立つ. 以上より,

$$\|\xi^x_i 1_F\|_p \leq \frac{p\Lambda T}{n}\|1_F\|_p$$

を得るから, ある普遍定数 $c_1 > 0$ が存在して,

$$\mathrm{E}[|\xi^x_i|^p \mid \mathcal{F}_{t_{i-1}}] \leq \frac{c_1 p\Lambda T}{n}$$

が成り立つ. Stirling 近似を用いて整理すると,

$$\mathrm{E}\left[|\xi^x_i|^p \mid \mathcal{F}_{t_{i-1}}\right] \leq p!\left(\frac{c_1 e\Lambda T}{n}\right)^p = \frac{p!}{2}\cdot\frac{2(c_1 e\Lambda T)^2}{n^2}\cdot\left(\frac{c_1 e\Lambda T}{n}\right)^{p-2}$$

を得る. 従って, 定理 3.1 を適用することができて, ある普遍定数 $c_2 > 0$ が存在して任意の $u \geq 0$ に対して

$$\mathrm{P}\left(|x^\top R x| \geq u\right) \leq 2\exp\left(-\frac{u^2}{2c_2(\Lambda^2 T^2/n + u\Lambda T/n)}\right)$$

が成り立つ. 従って, 補題 3.3 より

$$\left\|\max_{x\in\mathcal{N}}|x^\top R x|\right\|_{\psi_1} \lesssim \frac{\Lambda T\sqrt{d}}{\sqrt{n}} + \frac{\Lambda T d}{n}$$

が従う. この評価と (7.3) より,

$$\|\|R\|\|_{\psi_1} \lesssim \frac{\Lambda T\sqrt{d}}{\sqrt{n}} + \frac{\Lambda T d}{n}$$

が成り立つ. Step 1 と 2 の結果を組み合わせて, 示すべき不等式を得る. □

系 5.1 と同様の局所化の議論によって, 推定誤差の収束レートのみに興味があるのであれば, 係数過程に対する有界性の仮定は以下のように緩められる:

系 7.1. T は n に依存しないと仮定する. さらに, ある P-測度ゼロ集合 N と n に依存しない局所有界過程 $H = (H_t)_{t\geq 0}$ が存在して, 任意の $s \in [0, T]$ に対して

$$|b_s|_\infty \vee \|\sigma_s\| \leq |H_s| \qquad \text{on} \quad \Omega \setminus N$$

が成り立つと仮定する. このとき,

$$\left\| \widehat{[X, X]}_T - [X, X]_T \right\| = O_p\left(\sqrt{\frac{d}{n}} + \frac{d}{n} \right) \qquad (n \to \infty)$$

が成り立つ.

演習問題 1. 定理 7.1 から局所化の議論によって系 7.1 を導出せよ.

8 ファクターモデルへの応用

以下のモデルで与えられる d 次元確率過程 $Y = (Y_t)_{t\geq 0}$ を考える:

$$Y_t = \beta F_t + Z_t, \qquad t \geq 0. \tag{8.1}$$

ここに, $F = (F_t)_{t\geq 0}$ は r 次元連続セミマルチンゲール, $Z = (Z_t)_{t\geq 0}$ は d 次元連続セミマルチンゲールであり, $\beta \in \mathbb{R}^{d\times r}$ は非ランダムであるとする. モデル (8.1) は連続時間確率過程に対するファクターモデルと考えることができる: d 個の確率過程 $Y^1, \dots, Y^d$ に共通の r 個のファクター $F^1, \dots, F^r$ があり, β の第 k 列は k 番目のファクター F^k に対するファクターローディングに対応する. また, Z はファクターでは説明できない残差に対応する. 本節の目的は, 区間 $[0, T]$ を n 等分する時点 $t_i = Ti/n$ $(i = 0, 1, \dots, n)$ における Y の離散観測データ $(Y_{t_i})_{i=0}^n$ が与えられている状況下で, ファクターの数 r を推定することである.

6.3 節と同様に, d は n に依存するが, T は n に依存しない状況を考える. あとで見るように, 本節では, d が無限大に発散する状況を利用することで, r の一致推定量を構成する. 従って, ここでは高次元性は悪影響ではなくむしろ恩恵をもたらすことになり, いわゆる**次元の恩恵 (blessing of dimensionality)** の一例を与える.

8.1 線形代数からの準備

以下の議論では行列を摂動した際の固有値や特異値の挙動が重要な役割を果たすので, そのことに関連する結果のうち必要となるものを準備することから始める. この話題の系統的な扱いについては Horn & Johnson (2013) や Magnus & Neudecker (2019) を参照のこと.

まず, いくつか記号を導入する.

- d 次対称行列 A に対して, A の固有値を重複度を込めて降順に並べたものを $\lambda_1(A) \geq \cdots \geq \lambda_d(A)$ で表す.

- 各 $j=1,\dots,d$ について, $\mathbb{R}^d$ の j 次元部分ベクトル空間全体の集合を $\mathcal{L}_j$ で表す.
- $p,q\in\mathbb{N}$ に対して, $p\times q$ 零行列を $O_{p,q}$ で表す. $p=q$ の場合は $O_p=O_{p,p}$ と書く. また, 文脈から明らかな場合は添字は省略する.
- $A_j\in\mathbb{R}^{p_j\times q_j}$ $(j=1,\dots,k)$ に対して,

$$\mathrm{diag}(A_1,\dots,A_k):=\begin{pmatrix} A_1 & O & \cdots & O \\ O & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & A_k \end{pmatrix}$$

  と定める. 特に, 実数 $a_1,\dots,a_k$ に対して $\mathrm{diag}(a_1,\dots,a_k)$ は $a_1,\dots,a_k$ を対角成分に持つ k 次対角行列を表す.

定理 8.1 (Courant–Fischer のミニマックス定理)**.** A を d 次対称行列とする. 任意の $j=1,\dots,d$ に対して,

$$\lambda_j(A)=\min_{S\in\mathcal{L}_{d-j+1}}\max_{x\in S\setminus\{0\}}\frac{x^\top Ax}{x^\top x}=\max_{S\in\mathcal{L}_j}\min_{x\in S\setminus\{0\}}\frac{x^\top Ax}{x^\top x}$$

が成り立つ.

証明. $\lambda_k:=\lambda_k(A)$ $(k=1,\dots,d)$ とおく. $\mathbb{R}^d$ の正規直交基底 $x_1,\dots,x_d$ で $Ax_k=\lambda_k x_k$ $(k=1,\dots,d)$ を満たすものを 1 組とる. $x_j,\dots,x_d$ で張られる $\mathbb{R}^d$ の部分ベクトル空間を S_j とすると, $S_j\in\mathcal{L}_{d-j+1}$ である. いま, 任意に $x\in S_j$ をとると, ある $c_j,\dots,c_d\in\mathbb{R}$ を用いて $x=\sum_{k=j}^d c_kx_k$ と書けるから,

$$x^\top Ax=\sum_{k=j}^d\lambda_kc_k^2|x_k|^2\le\lambda_j\sum_{k=j}^d c_k^2|x_k|^2=\lambda_j|x|^2$$

が成り立つ. 従って,

$$\min_{S\in\mathcal{L}_{d-j+1}}\max_{x\in S\setminus\{0\}}\frac{x^\top Ax}{x^\top x}\le\max_{x\in S_j\setminus\{0\}}\frac{x^\top Ax}{x^\top x}\le\lambda_j$$

である. 一方で, 任意に $S\in\mathcal{L}_{d-j+1}$ をとると, $x_1,\dots,x_j$ で張られる $\mathbb{R}^d$ の部分ベクトル空間を S_j' としたとき, $\dim S+\dim S_j'=d+1>d$ となるから, $S\cap S_j'\neq\{0\}$ である. そこで, $x\in S\cap S_j'\setminus\{0\}$ を 1 つとると, $x\in S_j'$ よりある $c_1',\dots,c_j'\in\mathbb{R}$ を用いて $x=\sum_{k=1}^j c_k'x_k$ と書けるから,

$$x^\top Ax=\sum_{k=1}^j\lambda_k(c_k')^2|x_k|^2\ge\lambda_j\sum_{k=1}^d(c_k')^2|x_k|^2=\lambda_j|x|^2$$

が成り立つ. 従って,

$$\min_{S\in\mathcal{L}_{d-j+1}}\max_{x\in S\setminus\{0\}}\frac{x^\top Ax}{x^\top x}\ge\lambda_j$$

である. 以上より第 1 の等式が示された. 第 2 の等式は, $\lambda_{d-j+1}(-A)$ に第 1 の等式を適用することで得られる. □

系 8.1 (Weyl の不等式)**.** A, B を d 次対称行列とすると, 任意の $j = 1, \ldots, d$ に対して,

$$\lambda_j(A) + \lambda_d(B) \le \lambda_j(A+B) \le \lambda_j(A) + \lambda_1(B).$$

証明. 任意の $S \in \mathcal{L}_{d+j-1}$ に対して,

$$\max_{x \in S \setminus \{0\}} \frac{x^\top (A+B) x}{x^\top x} \le \max_{x \in S \setminus \{0\}} \frac{x^\top A x}{x^\top x} + \max_{x \in S \setminus \{0\}} \frac{x^\top B x}{x^\top x} \le \max_{x \in S \setminus \{0\}} \frac{x^\top A x}{x^\top x} + \lambda_1(B)$$

が成り立つから, 両辺の $S \in \mathcal{L}_{d+j-1}$ に関する最小値をとって右側の不等式を得る. 左側の不等式も同様の議論で示せる. □

系 8.2. $A \in \mathbb{R}^{r \times r}, B \in \mathbb{R}^{d \times r}$ とし, A は半正定値対称行列であると仮定する. このとき, 任意の $j = 1, \ldots, d$ に対して,

$$\lambda_r(A)\lambda_j(BB^\top) \le \lambda_j(BAB^\top) \le \lambda_1(A)\lambda_j(BB^\top).$$

証明. 仮定より $\lambda_r(A) \ge 0$ であることに注意する. Courant–Fischer のミニマックス定理より,

$$\begin{aligned}\lambda_j(BAB^\top) &= \max_{S \in \mathcal{L}_j} \min_{x \in S \setminus \{0\}} \frac{x^\top BAB^\top x}{x^\top x} \ge \max_{S \in \mathcal{L}_j} \min_{x \in S \setminus \{0\}} \frac{\lambda_r(A) x^\top BB^\top x}{x^\top x} \\ &= \lambda_r(A)\lambda_j(BB^\top)\end{aligned}$$

が成り立つ. 同様の議論で $\lambda_j(BAB^\top) \le \lambda_1(A)\lambda_j(BB^\top)$ も得られる. □

補題 8.1. $A \in \mathbb{R}^{d \times n}$ とする. 各 $j = 1, \ldots, d \wedge n$ に対して, $\lambda_j(A^\top A) = \lambda_j(AA^\top)$ が成り立つ.

証明. $A^\top A$ と $AA^\top$ はともに半正定値対称であるから, 両者の正の固有値が重複度も込めて一致することを示せばよい. $\lambda > 0$ を $A^\top A$ の重複度 m の固有値とする. 定義より, 1 次独立な $x_1, \ldots, x_m \in \mathbb{R}^d$ で $A^\top A x_j = \lambda x_j$ $(j = 1, \ldots, m)$ となるものを取ることができる. このとき, $Ax_1, \ldots, Ax_m$ は 1 次独立となる. 実際, $c_1, \ldots, c_m \in \mathbb{R}$ が $\sum_{j=1}^m c_j A x_j = 0$ を満たすとすると, 両辺に $A^\top$ を乗じて $\lambda \sum_{j=1}^m c_j x_j = 0$ を得るから, $\lambda \neq 0$ と $x_1, \ldots, x_m$ の 1 次独立性より $c_1 = \cdots = c_m = 0$ を得る. さらに, 各 $j = 1, \ldots, m$ に対して $AA^\top(Ax_j) = A(\lambda x_j) = \lambda(Ax_j)$ が成り立つから, $Ax_j (\neq 0)$ は λ に対する $AA^\top$ の固有ベクトルである. 従って λ は $AA^\top$ の固有値で, その重複度は m 以上である. 同様にして, $\lambda > 0$ が $AA^\top$ の重複度 m の固有値ならば, λ は $A^\top A$ の重複度 m 以上の固有値となることが示せる. 以上より $A^\top A$ と $AA^\top$ の正の固有値は重複度も込めて一致する. □

定義 8.1 (特異値)**.** $A \in \mathbb{R}^{d \times n}$ の**特異値** $\mathfrak{s}_1(A), \ldots, \mathfrak{s}_{d \wedge n}(A)$ を,

$$\mathfrak{s}_j(A) = \sqrt{\lambda_j(A^\top A)} = \sqrt{\lambda_j(AA^\top)} \qquad (j = 1, \ldots, d \wedge n)$$

で定義する.

定理 8.2 (特異値分解)**.** $A \in \mathbb{R}^{d\times n}$ とし, $D_0 := \mathrm{diag}(\mathfrak{s}_1(A), \dots, \mathfrak{s}_{d\wedge n}(A))$ とする. さらに, $d < n$ のときは $D := \begin{pmatrix} D_0 & O_{d,n-d}\end{pmatrix}$, $d = n$ のときは $D := D_0$, $d > n$ のときは $D := \begin{pmatrix} D_0 \\ O_{d-n,n}\end{pmatrix}$ として $D \in \mathbb{R}^{d\times n}$ を定める. このとき, ある直交行列 $U \in \mathbb{R}^{d\times d}, V \in \mathbb{R}^{n\times n}$ が存在して $A = UDV^\top$ と分解できる. この分解を A の**特異値分解**と呼ぶ.

証明. $AA^\top$ は直交行列によって対角化可能だから, ある d 次直交行列 U が存在して $U^\top(AA^\top)U = \mathrm{diag}(D_0^2, O)$ と書ける. $A^\top U$ の列ベクトルを $b_1, \dots, b_d \in \mathbb{R}^n$ とすると, $U^\top(AA^\top)U = (A^\top U)^\top(A^\top U) = (b_i \cdot b_j)_{1\le i,j\le d}$ と書き直せるので, $r := \mathrm{rank}\, A, v_j := \mathfrak{s}_j(A)^{-1}b_j$ $(j = 1, \dots, r)$ とおくと, $v_1, \dots, v_r$ は $\mathbb{R}^n$ の正規直交系をなし, かつ $b_j = 0\,(j = r+1, \dots, d)$ が成り立つ. $v_1, \dots, v_n$ が $\mathbb{R}^n$ の正規直交基底となるように $v_{r+1}, \dots, v_n \in \mathbb{R}^n$ をとり, $V := (v_1, \dots, v_n)$ と定める. 定義より V は n 次直交行列である. また,

$$U^\top AV = (b_i \cdot v_j)_{1\le i\le d, 1\le j\le n} = \begin{pmatrix} \mathrm{diag}(\mathfrak{s}_1(A), \dots, \mathfrak{s}_r(A)) & O_{r,n-r} \\ O_{d-r,r} & O_{d-r,n-r}\end{pmatrix} = D$$

となるから, $A = UDV^\top$ である. □

定理 8.3 (Horn & Johnson (2013), Corollary 7.3.5)**.** 任意の $A, B \in \mathbb{R}^{d\times n}$ と $j = 1, \dots, d\wedge n$ に対して,

$$|\mathfrak{s}_j(A + B) - \mathfrak{s}_j(A)| \le \|B\|.$$

証明. $d < n$ の場合は A, B の代わりに $A^\top, B^\top$ を考えることで, $d \ge n$ と仮定して一般性を失わない. $(d + n)$ 次正方行列 $\bar{A}, \bar{B}$ を

$$\bar{A} = \begin{pmatrix} O & A \\ A^\top & O\end{pmatrix}, \qquad \bar{B} = \begin{pmatrix} O & B \\ B^\top & O\end{pmatrix}$$

で定義する. このとき, $\bar{A}$ の固有値を重複度を込めて昇順に並べると,

$$-\mathfrak{s}_1(A) \le \cdots \le -\mathfrak{s}_n(A) \le \underbrace{0 = \cdots = 0}_{d-n} \le \mathfrak{s}_n(A) \le \cdots \le \mathfrak{s}_1(A) \tag{8.2}$$

となる. 実際, A の特異値分解を $A = UDV^\top$ とし, U の最初の n 列からなる $d \times n$ 行列を U_1, 残りの $d - n$ 列からなる行列を U_2 とする. $\hat{U} := U_1/\sqrt{2}, \hat{V} := V/\sqrt{2}$ とし,

$$\bar{U} := \begin{pmatrix} \hat{U} & -\hat{U} & U_2 \\ \hat{V} & \hat{V} & O\end{pmatrix}$$

と定める. このとき, $U_1^\top U_1 = I_n, U_1^\top U_2 = O$ および $U_2^\top U_2 = I_{d-n}$ となることに注意すると,

$$\bar{U}^\top\bar{U} = \begin{pmatrix} \hat{U}^\top & \hat{V}^\top \\ -\hat{U}^\top & \hat{V} \\ U_2^\top & O\end{pmatrix}\begin{pmatrix} \hat{U} & -\hat{U} & U_2 \\ \hat{V} & \hat{V} & O\end{pmatrix} = \begin{pmatrix} I_n & O_n & O_{n,d-n} \\ O_n & I_n & O_{n,d-n} \\ O_{d-n,n} & O_{d-n,n} & I_{d-n}\end{pmatrix} = I_{d+n}$$

となる. 従って, $\bar{U}$ は $(d+n)$ 次直交行列である. さらに, $U_1^\top AV = \mathrm{diag}(\mathfrak{s}_1(A), \dots, \mathfrak{s}_n(A)) =: D_0$, $U_2^\top AV = O$ に注意すると,

$$
\begin{aligned}
\bar{U}^\top \bar{A}\bar{U} &= \begin{pmatrix} \hat{U}^\top & \hat{V}^\top \\ -\hat{U}^\top & \hat{V} \\ U_2^\top & O \end{pmatrix} \begin{pmatrix} A\hat{V} & A\hat{V} & O \\ A^\top \hat{U} & -A\hat{U} & A^\top U_2 \end{pmatrix} \\
&= \begin{pmatrix} \hat{U}^\top A\hat{V} + \hat{V}A^\top \hat{U} & \hat{U}^\top A\hat{V} - \hat{V}A^\top \hat{U} & \hat{V}^\top AU_2 \\ -\hat{U}^\top A\hat{V} + \hat{V}A^\top \hat{U} & -\hat{U}^\top A\hat{V} - \hat{V}A^\top \hat{U} & \hat{V}^\top AU_2 \\ U_2^\top A\hat{V} & U_2^\top A\hat{V} & O_{d-n} \end{pmatrix} \\
&= \begin{pmatrix} D_0 & O & O \\ O & -D_0 & O \\ O & O & O_{d-n} \end{pmatrix}
\end{aligned}
$$

を得る. このことから $\bar{A}$ の固有値を昇順に並べると (8.2) となることが従う. 特に, $\lambda_j(\bar{A}) = \mathfrak{s}_j(A)$ である. 同様の議論によって, $\lambda_1(\bar{B}) = -\mathfrak{s}_1(B), \lambda_{d+n}(\bar{B}) = \mathfrak{s}_1(B)$ および $\lambda_j(\bar{A}+\bar{B}) = \mathfrak{s}_j(A+B)$ がわかる. 従って, Weyl の不等式より,

$$
\mathfrak{s}_j(A) - \mathfrak{s}_1(B) \le \mathfrak{s}_j(A+B) \le \mathfrak{s}_j(A) + \mathfrak{s}_1(B)
$$

を得る. $\mathfrak{s}_1(B) = \|B\|$ であるから, 示すべき結論が従う. □

8.2 推定量の構成

モデル (8.1) について以下の仮定をおく.

[A1] (i) F は n に依存しない (従って r も n に依存しない). さらに, $[F,F]_T$ は確率 1 で正則である.
(ii) ある r 次正定値対称行列 B が存在して $\|d^{-1}\beta^\top\beta - B\| \to 0$ $(n \to \infty)$ が成り立つ.
(iii) $\|[Z,Z]_T\| = O_p(1)$ $(n \to \infty)$.

推定量を構成するためのアイディアを説明するために, 一旦 $[F,Z]_T = 0$ という仮定をおく. これは, ファクターと残差の無相関性の仮定の確率過程版であり, 自然な仮定ではあるが, 後述する推定量の一致性を示すのには不要である. 仮定 $[F,Z]_T = 0$ の下で, Y の累積共分散行列は

$$
[Y,Y]_T = \beta[F,F]_T\beta^\top + [Z,Z]_T
$$

で与えられる. 従って, Weyl の不等式より, 各 $j = 1, \dots, d$ について

$$
\lambda_j(\beta^\top[F,F]_T\beta) - \|[Z,Z]_T\| \le \lambda_j([Y,Y]_T) \le \lambda_j(\beta^\top[F,F]_T\beta) + \|[Z,Z]_T\|
$$

が成り立つ. ここで, $\beta[F,F]_T\beta^\top$ のランクは高々 r であるから, $j > r$ のときは $\lambda_j(\beta[F,F]_T\beta^\top) = 0$ である. 従って, 仮定 [A1](iii) より $\lambda_j([Y,Y]_T) = O_p(1)$ である. 一方で, $j \le r$ の場合, 系 8.2 より $\lambda_j(\beta[F,F]_T\beta^\top) \ge \lambda_r([F,F]_T)\lambda_j(\beta\beta^\top)$ が成り立つ. 仮定 [A1](iii) より $\lambda_r([F,F]_T) > 0$ であり, また補題 8.1 と仮定 [A1](ii), および Weyl の不等式より $\lambda_j(\beta\beta^\top)/d = \lambda_j(B) + o_p(1)$ だ

から, 仮定 [A1](iii) とあわせると, $n \to \infty$ のとき $d \to \infty$ となるのであれば $\mathrm{P}(\lambda_j([Y,Y]_T)/d > \lambda_r([F,F]_T)\lambda_j(B)/2) \to 1$ が成り立つ. 故に, $j > r$ の場合は $\lambda_j([Y,Y]_T)$ は確率有界 (緊密) である一方で, $j \leq r$ の場合は (d のオーダーで) 発散することになる. 従って, $\lambda_j([Y,Y]_T)$ と $\lambda_{j+1}([Y,Y]_T)$ のギャップが非常に大きくなるような番号 j でもって r を推定するというアイディアが思いつく. 実際には $[Y,Y]_T$ は観測できないので実現共分散行列 $\widehat{[Y,Y]}_T$ で代用する必要がある.

上のアイディアに基づいて, Aït-Sahalia & Xiu (2017) は以下の推定量を提案した:

$$\hat{r}_n^{AX} := \arg\min_{j=0,1,\dots,r_{\max}} \left(d^{-1}\lambda_{j+1}(\widehat{[Y,Y]}_T) + (j+1)g(n,d) \right).$$

ここに, $r_{\max}$ は d 未満の正整数, $g(n,d)$ は正の値をとる n,d の関数である. $r_{\max}$ は想定されるファクター数の上界を表す. 次節で見るように, 理論上は $r_{\max} = d-1$ として問題ないが, 実用上はファクター数は少数であることを想定するため, 非常に大きなファクター数が推定されても解釈が難しくなるので, 有限標本でそのような無意味な推定結果が生じることを回避するためにあまり大きくない値を設定することが多い. これは計算効率の観点からも有用である. 一方で, $g(n,d)$ は「隣接する固有値間の非常に大きなギャップ」を検知するためのペナルティー関数として働く. すなわち,

$$d^{-1}\lambda_{j+1}(\widehat{[Y,Y]}_T) + (j+1)g(n,d) \tag{8.3}$$

において j を 1 つ増やした場合, この量はギャップ $d^{-1}(\lambda_{j+1}(\widehat{[Y,Y]}_T) - \lambda_{j+2}(\widehat{[Y,Y]}_T))$ だけ減少するが, $g(n,d)$ だけ増加する. 従って, ペナルティー関数による増加分を打ち消すほど固有値に大きなギャップが生じた場合に (8.3) は減少する. 特に, $j = r-1$ から j を 1 つ増やした際に固有値にはオーダー d のギャップが生じるため, $g(n,d) = o(1)$ となるようにペナルティー関数をとれば, $j < r$ の範囲でこのギャップがペナルティー関数の増加分で打ち消されることはない. 一方で, $j \geq r$ の場合に j を 1 つ増やした際の固有値の減少幅は $O(1)$ なので, d^{-1} よりも遅いオーダーで減少するように $g(n,d)$ をとれば, $j \geq r$ の範囲ではペナルティー関数の増加分が常に固有値のギャップによる減少分に打ち勝つことになる. 従って, (8.3) は $j = r$ のとき最小化されることが期待される. 実際には $[Y,Y]$ を $\widehat{[Y,Y]}_T$ に置き換えた際に生じる推定誤差を考慮する必要があるため, ペナルティー関数にはもう少し強い条件を課す必要がある (次節参照).

同様のアイディアに基づいて, Pelger (2019) は以下の推定量を提案した:

$$\hat{r}_n^{P}(\gamma) := \max\left\{ j \in \{1,\dots,r_{\max}\} : ER_j > 1+\gamma \right\}.$$

ただし,

$$ER_j := \frac{\lambda_j(\widehat{[Y,Y]}_T)/d + g(n,d)}{\lambda_{j+1}(\widehat{[Y,Y]}_T)/d + g(n,d)} \qquad (j = 1,\dots,d-1)$$

であり, $r_{\max}, g(n,d)$ は上と同様で, γ は正の定数である. ペナルティー関数 $g(n,d)$ を上と同じ条件を満たすようにとると, ER_r は $n \to \infty$ のとき正に発散することが期待される一方で, $j > r$ の場合の ER_j は分母分子ともに第 2 項が支配的となるため, 1 に収束することが期待される. 従って, ER_j が $1+\gamma$ を超えるような j の最大値は r となることが期待される.

8.3　推定量の一致性

$\widehat{[Y,Y]}_T$ の固有値の漸近挙動を評価するために, 残差過程に次の仮定をおく:

[A2] Z は以下の形で書ける:

$$Z_t = Z_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \qquad t \geq 0.$$

ここに, $b = (b_s)_{s\geq 0}$ は d 次元発展的可測過程, $\sigma = (\sigma_s)_{s\geq 0}$ は $\mathbb{R}^{d\times d'}$ 値発展的可測過程, $\mathcal{B}$ 上の d' 次元標準 Wiener 過程 $W = (W_t)_{t\geq 0}$ であり, ある P-測度ゼロ集合 N と n に依存しない局所有界過程 $H = (H_t)_{t\geq 0}$ が存在して, 任意の $s \in [0,T]$ に対して

$$|b_s|_\infty \vee \|\sigma_s\| \leq |H_s| \qquad \text{on} \quad \Omega \setminus N$$

が成り立つ.

注意 8.1. 仮定 [A1](iii) は仮定 [A2] の下では自動的に成立する. 実際,

$$\|[Z,Z]_T\| \leq \int_0^T \|\sigma_s \sigma_s^\top\| ds \leq \int_0^T |H_s|^2 ds$$

が成り立つが, H の局所有界性より $\int_0^T |H_s|^2 ds < \infty$ であるから, $\|[Z,Z]_T\| = O_p(1)$ が従う.

補題 8.2. $(\xi_n)_{n=1}^\infty$ を 0 に確率収束する確率変数列とする. このとき, 確率変数 η が $\mathrm{P}(\eta > 0) = 1$ を満たすならば, $\mathrm{P}(|\xi_n| > \eta) \to 0$ $(n \to \infty)$ が成り立つ.

証明. $\varepsilon > 0$ を任意に 1 つ固定する.

$$\mathrm{P}(|\xi_n| > \eta) \leq \mathrm{P}(|\xi_n| > \varepsilon) + \mathrm{P}(\eta < \varepsilon)$$

が成り立つから, $\xi_n \to^p 0$ より

$$\limsup_{n\to\infty} \mathrm{P}(|\xi_n| > \eta) \leq \mathrm{P}(\eta < \varepsilon)$$

を得る. しかし, $\varepsilon \to 0$ のとき $\mathrm{P}(\eta < \varepsilon) \to \mathrm{P}(\eta \leq 0) = 0$ となるから, 上の式で $\varepsilon \to 0$ とすることで示すべき主張を得る. □

補題 8.3. [A1]–[A2] を仮定する. このとき, 次のことが成り立つ.

(a) $n \to \infty$ のとき $d \to \infty$ となるならば, 任意の $j \in \{1,\dots,r\}$ と $\varepsilon \in (0,1)$ に対して, $\mathrm{P}(\lambda_j(\widehat{[Y,Y]}_T)/d > \varepsilon\lambda_r([F,F]_T)\lambda_j(B)) \to 1$ $(n \to \infty)$.

(b) $\lambda_{r+1}(\widehat{[Y,Y]}_T) = O_p(1 + d/n)$ $(n \to \infty)$.

証明. $n \to \infty$ の極限を考えるので, 以下 $n > r$ と仮定する. まず, d 次元確率過程 U に対して $\Delta^n U := (\Delta_1^n U, \dots, \Delta_n^n U)$ と定めると, $\widehat{[Y,Y]}_T = \Delta^n Y (\Delta^n Y)^\top$ と書けるから, $j = 1,\dots, d\wedge n$ に対して $\lambda_j(\widehat{[Y,Y]}_T) = \mathfrak{s}_j(\Delta_j^n Y)^2$ が成り立つ. $\Delta^n Y = \beta\Delta^n F + \Delta^n Z$ だから, 定理 8.3 より

$$|\mathfrak{s}_j(\Delta^n Y) - \mathfrak{s}_j(\beta\Delta^n F)| \leq \|\Delta^n Z\|$$

が成り立つ. 従って,

$$\mathfrak{s}_j(\beta\Delta^n F) - \|\Delta^n Z\| \leq \sqrt{\lambda_j(\widehat{[Y,Y]}_T)} \leq \mathfrak{s}_j(\beta\Delta^n F) + \|\Delta^n Z\| \tag{8.4}$$

が成り立つ.

$j \leq r$ の場合, $\mathfrak{s}_j(\beta\Delta^n F)^2 = \lambda_j(\beta\widehat{[F,F]}_T\beta^\top)$ に注意すると, 系 8.2 より

$$\mathfrak{s}_j(\beta\Delta^n F) \geq \sqrt{\lambda_r(\widehat{[F,F]}_T)\lambda_j(\beta\beta^\top)} \tag{8.5}$$

が成り立つ. F は n に依存しないことに注意すると, 命題 4.10 より $\|\widehat{[F,F]}_T - [F,F]_T\| \to^p 0$ $(n \to \infty)$ だから, Weyl の不等式より $\lambda_r(\widehat{[F,F]}_T) \to^p \lambda_r([F,F]_T)$ $(n \to \infty)$ が成り立つ. また, 仮定 [A1](ii) と Weyl の不等式より $\lambda_j(\beta\beta^\top)/d \to \lambda_j(B)$ $(n \to \infty)$ が成り立つ. 従って,

$$\lambda_r(\widehat{[F,F]}_T)\lambda_j(\beta\beta^\top)/d \to^p \lambda_r([F,F]_T)\lambda_j(B)$$

である. 一方で, 系 7.1 より $\|\Delta^n Z\|^2 = \|\widehat{[Z,Z]}_T\| = O_p(1 + d/n)$ である. 従って, $n \to \infty$ のとき $d \to \infty$ となるならば, $\|\Delta^n Z\|^2/d \to^p 0$ が成り立つ. 以上より, $n \to \infty$ のとき $d \to \infty$ となるならば,

$$\sqrt{\lambda_r(\widehat{[F,F]}_T)\lambda_j(\beta\beta^\top)/d} - \|\Delta^n Z\|/\sqrt{d} \to^p \sqrt{\lambda_r([F,F]_T)\lambda_j(B)}$$

が成り立つから, $\lambda_r([F,F]_T)\lambda_j(B) > 0$ に注意すると, 補題 8.2 より

$$\mathrm{P}\left(\sqrt{\lambda_r(\widehat{[F,F]}_T)\lambda_j(\beta\beta^\top)/d} - \|\Delta^n Z\|/\sqrt{d} > \sqrt{\varepsilon\lambda_r([F,F]_T)\lambda_j(B)}\right) \to 0 \tag{8.6}$$

が成り立つ. (8.4)–(8.6) を組み合わせて (a) を得る.

$j = r+1$ の場合, $\beta\Delta^n F$ のランクが r であることに注意すると, $\mathfrak{s}_{r+1}(\beta\Delta^n F) = 0$ である. 系 7.1 より $\|\Delta^n Z\|^2 = \|\widehat{[Z,Z]}_T\| = O_p(1 + d/n)$ であったから, (8.4) より (b) を得る. □

定理 8.4. [A1]–[A2] を仮定する. また, ペナルティー関数は $n \to \infty$ のとき $g(n,d) \to 0$ かつ $(n \wedge d)g(n,d) \to \infty$ を満たすとする. このとき, $r_{\max} \geq r$ ならば, $n \to \infty$ のとき

$$\mathrm{P}\left(\hat{r}_n^{AX} = r\right) \to 1, \qquad \mathrm{P}\left(\hat{r}_n^{P}(\gamma) = r\right) \to 1$$

が成り立つ.

証明. ペナルティー関数に対する仮定から特に $n \wedge d \to \infty$ とならなければならないから, $d \to \infty$ となることに注意する.

まず $\hat{r}_n^{AX}$ について考える.

$$\begin{aligned}
&\mathrm{P}\left(\hat{r}_n^{AX} < r\right) \\
&= \mathrm{P}\left(d^{-1}\lambda_{j+1}(\widehat{[Y,Y]}_T) + (j+1)g(n,d) < d^{-1}\lambda_{r+1}(\widehat{[Y,Y]}_T) + (r+1)g(n,d) \text{ for some } j < r\right) \\
&\leq \mathrm{P}\left(d^{-1}\lambda_r(\widehat{[Y,Y]}_T) < d^{-1}\lambda_{r+1}(\widehat{[Y,Y]}_T) + rg(n,d)\right)
\end{aligned}$$

であるが, 補題 8.3(b) とペナルティー関数に関する仮定より $d^{-1}\lambda_{r+1}(\widehat{[Y,Y]}_T)+rg(n,d)=o_p(1)$ であるから, 補題 8.2 と 8.3(a) より $\mathrm{P}\left(\hat{r}_n^{AX}<r\right)\to 0$ を得る. 一方で,

$$
\begin{aligned}
&\mathrm{P}\left(\hat{r}_n^{AX}>r\right)\\
&=\mathrm{P}\left(d^{-1}\lambda_{j+1}(\widehat{[Y,Y]}_T)+(j+1)g(n,d)<d^{-1}\lambda_{r+1}(\widehat{[Y,Y]}_T)+(r+1)g(n,d)\text{ for some } r<j\le r_{\max}\right)\\
&\le \mathrm{P}\left(g(n,d)<d^{-1}\lambda_{r+1}(\widehat{[Y,Y]}_T)\right)
\end{aligned}
$$

であり, 補題 8.3(b) より $d^{-1}\lambda_{r+1}(\widehat{[Y,Y]}_T)=O_p(d^{-1}+n^{-1})=O_p((d\wedge n)^{-1})$ であるから, ペナルティー関数に関する仮定より $\mathrm{P}\left(\hat{r}_n^{AX}>r\right)\to 0$ を得る. 以上より $\mathrm{P}\left(\hat{r}_n^{AX}=r\right)\to 1$ である.

次に $\hat{r}_n^P(\gamma)$ について考える.

$$
\mathrm{P}\left(\hat{r}_n^P(\gamma)<r\right)\le \mathrm{P}\left(ER_r\le 1+\gamma\right)=\mathrm{P}\left(\lambda_r(\widehat{[Y,Y]}_1)/d\le(1+\gamma)\lambda_{r+1}(\widehat{[Y,Y]}_1)/d+\gamma g(n,d)\right)
$$

であるから, 補題 8.2 と 8.3 より $\mathrm{P}\left(\hat{r}_n^P(\gamma)<r\right)\to 0$ である. 一方で,

$$
\begin{aligned}
\mathrm{P}\left(\hat{r}_n^P(\gamma)>r\right)&=\mathrm{P}\left(ER_j>1+\gamma\text{ for some } j>r\right)\\
&=\mathrm{P}\left(\lambda_j(\widehat{[Y,Y]}_1)/d>(1+\gamma)\lambda_{j+1}(\widehat{[Y,Y]}_1)/d+\gamma g(n,d)\text{ for some } j>r\right)\\
&\le \mathrm{P}\left(\lambda_{r+1}(\widehat{[Y,Y]}_1)/d>\gamma g(n,d)\right)
\end{aligned}
$$

であるから, 補題 8.3 より $\mathrm{P}\left(\hat{r}_n^P(\gamma)>r\right)\to 0$ である. 以上より $\mathrm{P}\left(\hat{r}_n^P(\gamma)=r\right)\to 1$ である. □

9 高次元中心極限定理入門

9.1 モチベーション: 高次元共分散行列に対する同時推測

これまでの議論の焦点は点推定であったが, より精緻なデータ解析を行うためには, 信頼区間の構成や仮説検定などの推測統計の話に踏み込む必要がある. 本節ではその際に重要な役割を果たす, Chernozhukov, Chetverikov および Kato らによる高次元中心極限定理とブートストラップ法の理論, 通称 **CCK 理論**を紹介し, 具体的な統計的応用についてもいくつか触れる.

モチベーションとなる統計的問題として, 高次元共分散行列に対する信頼区間を構成する問題を考える. 独立同分布な d 次元確率変数列 $X_1, \ldots, X_n$ が観測データとして与えられているとする. X_1 は共分散行列 Σ を持つものとし, Σ をデータから推定する問題を考える. 簡単のため X_1 は平均 0 であるとすると, 特に Σ に対する構造的な仮定がないのであれば, 中心化していない標本共分散行列

$$\widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$$

で Σ を点推定するのが自然である. 3.1 節ですでに見たように, X_1 の各成分が劣 Gauss 型であれば, $d \gg n$ なる高次元の状況においても $\widehat{\Sigma}_n$ は最大値ノルムに関して Σ の一致推定量となる. いま, $j, k \in \{1, \ldots, d\}$ を固定すると, 適当な仮定の下で, 中心極限定理により,

$$\frac{\widehat{\Sigma}_{n,jk} - \Sigma_{jk}}{\sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]}}$$

は $n \to \infty$ のとき標準正規分布に分布収束する. 従って, $\alpha \in (0,1)$ に対して, $z_{\alpha/2}$ を標準正規分布の上側 $100\alpha/2\%$ 点, すなわち $Z \sim \mathrm{N}(0,1)$ に対して $\mathrm{P}(Z > z_{\alpha/2}) = z_{\alpha/2}$ となるような点とすれば,

$$\left[\widehat{\Sigma}_{n,jk} - z_{\alpha/2}\sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]}, \widehat{\Sigma}_{n,jk} + z_{\alpha/2}\sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]}\right] \tag{9.1}$$

は Σ_{jk} に対する漸近 $100(1-\alpha)\%$ 信頼区間となる. 通常は $\mathrm{Var}[\widehat{\Sigma}_{n,jk}]$ が未知なのでデータから推定する必要があるが, これは例えば

$$V_{n,jk} := \frac{1}{n^2}\sum_{i=1}^n \left(X_{ij}X_{ik} - \widehat{\Sigma}_{n,jk}\right)^2$$

で推定してやればよい. 以上は Σ の特定の成分に着目した場合の話であったが, すべての成分に対して同時に統計推測をしたい場合はどうすれば良いだろうか? ナイーブには, 信頼区間 (9.1) を各成分について構成するという方法が考えられるが, その場合,「ある」成分が信頼区間に含まれない確率は

$$\mathrm{P}\left(\bigcup_{\substack{j,k=1\\ j\le k}}^d \left\{\Sigma_{jk} \notin \left[\widehat{\Sigma}_{n,jk} - z_{\alpha/2}\sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]}, \widehat{\Sigma}_{n,jk} + z_{\alpha/2}\sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]}\right]\right\}\right)$$

$$= \mathrm{P}\left(\max_{1 \le j \le k \le d} \left| \frac{\widehat{\Sigma}_{n,jk} - \Sigma_{jk}}{\sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]}} \right| > z_{\alpha/2} \right)$$

で与えられることになり, この確率が (漸近的に) α 以下になる保証は全くない. 上の等式からわかるように,「ある」成分が信頼区間に含まれない確率が漸近的に α となるようにするには, c_α を

$$\mathrm{P}\left(\max_{1 \le j \le k \le d} \left| \frac{\widehat{\Sigma}_{n,jk} - \Sigma_{jk}}{\sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]}} \right| > c_\alpha \right) \approx \alpha$$

が成り立つように定めて, Σ_{jk} に対する信頼区間を

$$\left[\widehat{\Sigma}_{n,jk} - c_\alpha \sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]},\ \widehat{\Sigma}_{n,jk} + c_\alpha \sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]} \right]$$

とすればよい. このためには, 最大値型統計量

$$\max_{1 \le j \le k \le d} \left| \frac{\widehat{\Sigma}_{n,jk} - \Sigma_{jk}}{\sqrt{\mathrm{Var}[\widehat{\Sigma}_{n,jk}]}} \right|$$

の分布を近似する必要があるが, $d \gg n$ なる高次元の設定でこれを達成するのは自明でない. 特に, 推定量の (漸近) 共分散行列に特に構造を課さない場合, 上の統計量の漸近分布は, (適切な正規化を施したとしても) 一般に解析的に表現することが困難である. 以下で詳細に見ていくように, Chernozhukov, Chetverikov & Kato (2013) に端を発する一連の研究では, ブートストラップ法をリーズナブルな正則条件の下で高次元の設定においても正当化することで, この困難を克服する.

以下本節では, $\log d > 1$ かつ $\log n > 1$ となるように $d \wedge n \ge 3$ を仮定する.

9.2 Chernozhukov–Chetverikov–Kato 理論

d 次元超矩形全体の集合を $\mathcal{R}_d$ と書くことにする:

$$\mathcal{R}_d := \left\{ \prod_{j=1}^d [a_j, b_j] : -\infty < a_j < b_j < \infty,\ j = 1, \dots, d \right\}.$$

d 次元確率変数 W に対して, 最大値統計量 $\max_{1 \le j \le d} |W_j|$ の分布を近似するには, $A \in \mathcal{R}_d$ に対して $\mathrm{P}(W \in A)$ を近似できれば十分である. 実際, 任意の $x \ge 0$ に対して,

$$\mathrm{P}\left(\max_{1 \le j \le d} |W_j| \le x \right) = \mathrm{P}\left(W \in [-x, x]^d \right)$$

と書き直せる.

定理 9.1 (Chernozhukov et al. (2022), Theorem 2.1)**.** $X_1, \dots, X_n$ を平均 0 の独立な d 次元確率変数列とし, $S_n = n^{-1/2} \sum_{i=1}^n X_i$ とおく. ある定数 $B_n \ge 1$ と $0 < b \le 1$ が存在して以下の 3 条件が成り立つと仮定する:

(i) すべての $i = 1, \ldots, n$ と $j = 1, \ldots, d$ について, $\|X_{ij}\|_{\psi_1} \le B_n$.
(ii) すべての $j = 1, \ldots, d$ について, $n^{-1} \sum_{i=1}^n \mathrm{E}[X_{ij}^4] \le B_n^2$.
(iii) すべての $j = 1, \ldots, d$ について, $n^{-1} \sum_{i=1}^n \mathrm{E}[X_{ij}^2] \ge b^2$.

このとき, ある普遍定数 $C > 0$ が存在して,

$$\sup_{A \in \mathcal{R}_d} |\mathrm{P}(S_n \in A) - \mathrm{P}(Z_n \in A)| \le \frac{C}{b} \left(\frac{B_n^2 \log^5(dn)}{n} \right)^{1/4} \tag{9.2}$$

が成り立つ. ここに, Z_n は S_n と同じ共分散行列を持つ平均 0 の d 次元正規確率変数である.

定理 9.1 において, S_n の共分散行列が既知であれば, Z_n の分布をモンテカルロシミュレーションで近似計算することができるので, 最大値型統計量 $\max_{1 \le j \le d} |S_{n,j}|$ の分布を $\max_{1 \le j \le d} |Z_{n,j}|$ の分布によって近似計算することができる. 通常は S_n の共分散行列は既知ではなく, その場合 Z_n の分布を直接シミュレーションすることはできないため, 代わりに**ブートストラップ法**を用いる. ここでは, 数学的に扱いやすい **Gauss 型ワイルドブートストラップ法**の正当性について議論する.

定理 9.2 (Chernozhukov et al. (2022), Theorem 2.2)**.** 定理 9.1 と同じ設定を考える. $e_1, \ldots, e_n$ を標準正規分布に従う独立確率変数列で $X = (X_1, \ldots, X_n)$ と独立なものとし,

$$S_n^* := \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i (X_i - \bar{X}_n), \quad \text{ただし}, \bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$$

と定める. さらに, $\alpha \in (0, 1)$ に対して, X が与えられた下での $\max_{1 \le j \le d} |S_{n,j}^*|$ の条件付き分布の $(1-\alpha)$-分位点を c_α^* とする:

$$c_\alpha^* := \inf \left\{ x \in \mathbb{R} : \mathrm{P} \left(\max_{1 \le j \le d} |S_{n,j}^*| \le x \mid X \right) \ge 1 - \alpha \right\}.$$

このとき, 定理 9.1 の仮定の下で, ある普遍定数 $C > 0$ が存在して,

$$\left| \mathrm{P} \left(\max_{1 \le j \le d} |S_{n,j}| > c_\alpha^* \right) - \alpha \right| \le \frac{C}{b} \left(\frac{B_n^2 \log^5(dn)}{n} \right)^{1/4} \tag{9.3}$$

が成り立つ.

定理 9.2 を用いると, 劣指数型の独立標本の母平均に対する同時信頼区間を構成できる: $Y_1, \ldots, Y_n$ を平均 $\mu \in \mathbb{R}^d$ を持つ独立同分布な d 次元確率変数列とし, ある定数 $K \ge 1$ と $0 < c \le 1$ が存在して $\max_{1 \le j \le d} \|Y_{1j}\|_{\psi_1} \le K$ および $\min_{1 \le j \le d} \mathrm{Var}[Y_{1j}] \ge c$ が成り立つとする. このとき, $X_i := Y_i - \mu$ $(i = 1, \ldots, n)$ とおくと, ある普遍定数 $C \ge 1$ が存在して, 定理 9.1 の仮定が $B_n = CK^2, b = \sqrt{c}$ として成立することが確認できる.

$$\bar{Y}_n := \frac{1}{n} \sum_{i=1}^n Y_i, \qquad \bar{Y}_n^* := \frac{1}{n} \sum_{i=1}^n e_i (Y_i - \bar{Y}_n)$$

と定めると,

$$\bar{Y}_n^* = \frac{1}{n}\sum_{i=1}^n e_i(X_i - \bar{X}_n)$$

と書き直せることに注意すると, $\alpha \in (0,1)$ に対して c_α^* を $Y = (Y_1, \dots, Y_n)$ が与えられた下での $\max_{1\le j\le d} |\bar{Y}_{n,j}^*|$ の条件付き分布の $(1-\alpha)$-分位点とすれば,

$$\left| \mathrm{P}\left(\max_{1\le j\le d} |\bar{Y}_{n,j} - \mu_j| > c_\alpha^* \right) - \alpha \right| = O\left(\left(\frac{\log^5(dn)}{n} \right)^{1/4} \right) \qquad (n \to \infty)$$

が成り立つ. 従って, $\log^5 d = o(n)$ であれば

$$\left[\bar{Y}_{n,j} - c_\alpha^*,\ \bar{Y}_{n,j} + c_\alpha^* \right] \qquad (j = 1, \dots, d) \tag{9.4}$$

は $\mu_1, \dots, \mu_d$ に対する漸近 $100(1-\alpha)\%$ 同時信頼区間を与える. ここで, Y が与えられた下での $\bar{Y}_n^*$ の条件付き分布は, $e_1, \dots, e_n$ をシミュレーションすることによって数値計算できるので, c_α^* はモンテカルロシミュレーションによって数値計算できることに注意しておく.

同時信頼区間 (9.4) は成分ごとのスケール不変性がないため応用上不便である. すなわち, j 番目の変数の観測データ $Y_{1j}, \dots, Y_{nj}$ を, 計測単位の変更などによって一斉に定数倍した場合, c_α^* の値が変化してしまい同時信頼区間の形状が変わってしまう. そのため, 成分ごとにスチューデント化した統計量に基づき同時信頼区間を構成するのが望ましい. この場合の正当化のために次の結果を用いる:

定理 9.3. 定理 9.2 と同じ設定を考える. 各 $j = 1, \dots, d$ について

$$\hat{\sigma}_{n,j} := \sqrt{\frac{1}{n}\sum_{i=1}^n (X_{ij} - \bar{X}_{n,j})^2}$$

と定める. $\alpha \in (0,1)$ に対して, X が与えられた下での $\max_{1\le j\le d} |S_{n,j}^*|/\hat{\sigma}_{n,j}$ の条件付き分布の $(1-\alpha)$-分位点を $c_\alpha^\star$ とする. このとき, 定理 9.1 の仮定の下で, ある普遍定数 $C > 0$ が存在して,

$$\left| \mathrm{P}\left(\max_{1\le j\le d} \frac{|S_{n,j}|}{\hat{\sigma}_{n,j}} > c_\alpha^\star \right) - \alpha \right| \le \frac{C}{b}\left(\frac{B_n^2 \log^5(dn)}{n} \right)^{1/4} \tag{9.5}$$

が成り立つ.

各 $j = 1, \dots, d$ について

$$\hat{\sigma}_{n,j} := \sqrt{\frac{1}{n}\sum_{i=1}^n (Y_{ij} - \bar{Y}_{n,j})^2}$$

と定めると, $X_i := Y_i - \mu$ $(i = 1, \dots, n)$ に対して

$$\hat{\sigma}_{n,j} = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_{ij} - \bar{X}_{n,j})^2}$$

と書き直せる. 従って, $\alpha \in (0,1)$ に対して $c^\star_\alpha$ を Y が与えられた下での $\max_{1\le j\le d}|\bar{Y}^*_{n,j}|/\hat{\sigma}_{n,j}$ の条件付き分布の $(1-\alpha)$-分位点とすれば, 定理 9.3 より

$$\left|\mathrm{P}\left(\max_{1\le j\le d}\frac{|\bar{Y}_{n,j}-\mu_j|}{\hat{\sigma}_{n,j}}>c^\star_\alpha\right)-\alpha\right|=O\left(\left(\frac{\log^5(dn)}{n}\right)^{1/4}\right)\qquad(n\to\infty)$$

が成り立つ. 従って, $\log^5 d=o(n)$ であれば

$$\left[\bar{Y}_{n,j}-c^\star_\alpha\hat{\sigma}_{n,j},\ \bar{Y}_{n,j}+c^\star_\alpha\hat{\sigma}_{n,j}\right]\qquad(j=1,\dots,d)\tag{9.6}$$

は $\mu_1,\dots,\mu_d$ に対する漸近 $100(1-\alpha)\%$ 同時信頼区間を与える.

注意 9.1. 定理 9.3 より,

$$\left|\mathrm{P}\left(c^\star_{1-\alpha/2}\le\max_{1\le j\le d}\frac{|S_{n,j}|}{\hat{\sigma}_{n,j}}\le c^\star_{\alpha/2}\right)-(1-\alpha)\right|\le\frac{C}{b}\left(\frac{B_n^2\log^5(dn)}{n}\right)^{1/4}$$

も成り立つので, 集合

$$\left\{\mu\in\mathbb{R}^d: c^\star_{1-\alpha/2}\le\max_{1\le j\le d}\frac{|\bar{Y}_{n,j}-\mu_j|}{\hat{\sigma}_{n,j}}\le c^\star_{\alpha/2}\right\}\tag{9.7}$$

に真の平均が含まれる確率は漸近的に $1-\alpha$ に近づく. このような集合を μ の漸近 $100(1-\alpha)\%$ 信頼域と呼ぶ. 同時信頼区間は矩形型の信頼域と見なすことができる. 正規確率ベクトルの最大値は多くの場合にその期待値に集中するという**超集中現象** ((Chatterjee, 2014, Section 9.6) や Tanguy (2015) 参照) を考慮すると, (9.6) よりも (9.7) の方がより精度の高い (= 体積が小さい) 信頼域を与えるのではないかということが Klaassen et al. (2023) で指摘されている (同論文の (8) 式の後の議論参照). 実際, Gauss 型グラフィカルモデルの推定の文脈で, 前者より後者のパフォーマンスが良いことを Klaassen et al. (2023) は数値実験で確認している.

■ 劣 Gauss 型独立標本の共分散行列に対する同時信頼区間

9.1 節で考えた設定の下で, Σ の成分に対する同時信頼区間を構成するには, 上の議論を $Y_i=\mathrm{vec}(X_iX_i^\top)$ の場合に適用すればよい. ただし, 行列 $A=(A_{jk})_{1\le j\le l,1\le k\le m}$ の**ベクトル化** $\mathrm{vec}(A)$ を以下で定義する:

$$\mathrm{vec}(A):=(A^{11},\dots,A^{l1},A^{12},\dots,A^{l2},\dots,A^{1m},\dots,A^{lm})^\top\in\mathbb{R}^{lm}.$$

具体的な手順は以下の通りである. 標準正規分布に従う独立確率変数列 $e_1,\dots,e_n$ を観測データと独立に生成し, $\widehat{\Sigma}_n$ のブートストラップ統計量

$$\widehat{\Sigma}^*_n=\frac{1}{n}\sum_{i=1}^n e_i(X_iX_i^\top-\widehat{\Sigma}_n)$$

を考える. $\alpha\in(0,1)$ に対して, 観測データが与えられた下での $\max_{1\le j\le k\le d}|\widehat{\Sigma}^*_{n,jk}|/\sqrt{V_{n,jk}}$ の条件付き分布の $(1-\alpha)$-分位点を $c^\star_\alpha$ とすると, $\log^5 d=o(n)$ ならば

$$\left[\widehat{\Sigma}_{n,jk}-c^\star_\alpha\sqrt{V_{n,jk}},\ \widehat{\Sigma}_{n,jk}+c^\star_\alpha\sqrt{V_{n,jk}}\right]\qquad(j,k=1,\dots,d)$$

は Σ の成分に対する漸近 $100(1-\alpha)\%$ 同時信頼区間を与える.

9.3 実現共分散行列への応用

4.2 節と同じ設定を考える.

定理 9.4. $b=0$ であり, かつ σ は非ランダムであると仮定する. さらに, 以下の 2 条件を仮定する:

(i) ある定数 $K\geq 1$ と $\gamma\in(0,1]$ が存在して,

$$\sup_{s\in[0,T]}|c_s|_\infty \vee \sup_{s,t,\in[0,T]:s\neq t}\frac{|c_s-c_t|_\infty}{|s-t|^\gamma}\leq K^2$$

が成り立つ.

(ii) ある定数 $v\in(0,1]$ が存在して, 任意の $j,k\in\{1,\dots,d\}$ に対して

$$\int_0^T\left(c_t^{jj}c_t^{kk}+\left(c_t^{jk}\right)^2\right)dt\geq v$$

が成り立つ.

このとき, $S_n:=\sqrt{n}\,\mathrm{vec}(\widehat{[X,X]}_T-[X,X]_T)$ とおくと, T のみに依存する定数 $C>0$ が存在して,

$$\sup_{A\in\mathcal{R}_{d^2}}|\operatorname{P}(S_n\in A)-\operatorname{P}(Z_n\in A)|\leq C\frac{K^2}{\sqrt{v}}\left\{\left(\frac{\log^5(dn)}{n}\right)^{1/4}+\sqrt{\frac{\log^2 d}{n^\gamma}}\right\} \tag{9.8}$$

が成り立つ. ここに, $Z_n\sim\mathrm{N}(0,\Sigma)$ であり, Σ は次式で定義される d^2 次半正定値対称行列である:

$$\Sigma_{(j-1)d+k,(l-1)d+m}:=\int_0^T\left(c_t^{jl}c_t^{km}+c_t^{jm}c_t^{kl}\right)dt,\qquad j,k,l,m=1,\dots,d.$$

注意 9.2. 定理 9.4 で係数過程 b,σ に置いた仮定のうち,「$b=0$」という仮定は, 標準的な議論によって「b が条件 (i) に対応する条件を満たす」という仮定に置き換えることができる. 一方で,「σ が非ランダムである」という仮定を置き換えるには本質的な困難が生じる. これは, σ がランダムな場合, S_n が独立確率ベクトルの和とはならないことと, S_n の漸近共分散行列に相当する量がランダムとなるために Z_n を混合正規分布に従う確率ベクトルに置き換える必要が生じるからである. この問題は, σ の Malliavin の意味での微分可能性を仮定した上で, S_n を二重 Skorohod 積分とみて Stein の方法を適用すると解消できることが Koike (2019) で示されている. 詳細はこの講義ノートで扱える範疇を超えるため, 原論文を参照のこと.

独立同分布データの場合と異なり, 定理 9.4 の設定においては, 標本共分散行列にあたる統計量

$$\sum_{i=1}^n\mathrm{vec}\left(\Delta_i^nX(\Delta_i^nX)^\top-\widehat{[X,X]}_T\right)\mathrm{vec}\left(\Delta_i^nX(\Delta_i^nX)^\top-\widehat{[X,X]}_T\right)^\top \tag{9.9}$$

は Σ の一致推定量とはならない. これは, $\Delta_i^nX(\Delta_i^nX)^\top$ の期待値が一般には $\widehat{[X,X]}_T$ と大きく異なることに起因する. Σ の推定量として, 以下の統計量が Barndorff-Nielsen & Shephard (2004) で

提案されている:

$$\widehat{\Sigma} := n \sum_{i=1}^{n} \chi_i \chi_i^\top - \frac{n}{2} \sum_{i=1}^{n-1} \left(\chi_i \chi_{i+1}^\top + \chi_{i+1} \chi_i^\top \right).$$

ただし, 各 $i = 1, \ldots, n$ について

$$\chi_i := \mathrm{vec}\left(\Delta_i^n X (\Delta_i^n X)^\top \right)$$

と定める.

命題 9.1. 定理 9.4 の仮定の下で, T のみに依存する定数 $C > 0$ が存在して, 任意の $j, k, l, m \in \{1, \ldots, d\}$ と $p \geq 1$ に対して,

$$\left\| \widehat{\Sigma}_{(j-1)d+k,(l-1)d+m} - \Sigma_{(j-1)d+k,(l-1)d+m} \right\|_p \leq CK^4 \left(\sqrt{\frac{p}{n}} + \frac{p^3 \log^2 n}{n} + \frac{1}{n^\gamma} \right)$$

が成り立つ.

累積共分散行列 $[X, X]_T$ に対する同時信頼区間の構成に応用するために, スチューデント化した統計量の最大値

$$U_n := \max_{1 \leq j \leq k \leq d} \frac{\sqrt{n} \left| \widehat{[X^j, X^k]}_T - [X^j, X^k]_T \right|}{\widehat{\varsigma}_{jk}}$$

の分布をブートストラップ法で近似する. ただし,

$$\widehat{\varsigma}_{jk} := \sqrt{\widehat{\Sigma}_{(j-1)d+k,(j-1)d+k}}$$

と定める. 重み付け変数 $e_1, \ldots, e_n$ を観測データと独立に生成し, 実現共分散行列の (ワイルド) ブートストラップ統計量

$$\widehat{[X, X]}_T^* = \left(\widehat{[X^j, X^k]}_T^* \right)_{1 \leq j,k \leq d} := \sum_{i=1}^{n} e_i \Delta_i^n X (\Delta_i^n X)^\top$$

を考える. このとき, U_n のブートストラップ統計量は

$$U_n^* := \max_{1 \leq j \leq k \leq d} \frac{\sqrt{n} \left| \widehat{[X^j, X^k]}_T^* \right|}{\widehat{\varsigma}_{jk}}$$

で与えられる.

(9.9) が Σ の一致推定量とはならないのと同様の理由で, 通常のワイルドブートストラップ法のように $e_1, \ldots, e_n$ を独立確率変数としてしまうと, U_n^* の分布は U_n の分布を正しく近似できないことが知られている. この問題は, $e_1, \ldots, e_n$ に適切な従属構造を導入することで解消できる.

定理 9.5. $e_1, \ldots, e_n$ は X と独立な平均 0 の正規確率変数列で

$$\mathrm{E}[e_i e_{i+h}] = \begin{cases} 1 & \text{if } h = 0, \\ -\frac{1}{2} & \text{if } h = 1, \\ 0 & \text{otherwise} \end{cases}$$

を満たすものとする. $\alpha \in (0,1)$ に対して, X が与えられた下での U_n^* の条件付き分布の $(1-\alpha)$-分位点を $c_\alpha^\star$ とする. また,

$$\frac{\log^5(dn)\log^4 n}{n} \leq 1 \tag{9.10}$$

が成り立つと仮定する. このとき, 定理 9.4 の仮定の下で, T のみに依存する定数 $C > 0$ が存在して,

$$|\mathrm{P}(U_n > c_\alpha^\star) - \alpha| \leq C\frac{K^2}{\sqrt{v}}\left\{\left(\frac{\log^5(dn)}{n}\right)^{1/4} + \sqrt{\frac{\log^2 d}{n^\gamma}}\right\} \tag{9.11}$$

が成り立つ.

定理 9.5 の条件を満たすような重み付け変数 $e_1, \dots, e_n$ は以下の MA(1) モデルによって容易に生成できる:

$$e_i = \eta_i - \eta_{i-1}, \qquad i = 1, \dots, n.$$

ここに, $(\eta_i)_{i=0}^n$ は平均 0, 分散 $\frac{1}{2}$ の正規分布に従う独立同分布列である. また, このことから, 定理 9.5 のブートストラップ法は, Hounyo (2017) の **wild blocks of blocks bootstrap** と本質的に同等であることもわかる. 定理 9.5 より, 定理 9.4 の仮定の下で, T, K, v が n に依存せず, かつ $(\log d)^{5\vee(2/\gamma)} = o(n/\log^4 n)$ ならば

$$\left[\widehat{[X^j, X^k]}_T - c_\alpha^\star \widehat{\varsigma}_{n,jk},\ \widehat{[X^j, X^k]}_T + c_\alpha^\star \widehat{\varsigma}_{n,jk}\right] \qquad (j, k = 1, \dots, d)$$

は $[X, X]_T$ の成分に対する漸近 $100(1-\alpha)\%$ 同時信頼区間を与える.

9.4 定理 9.1 の証明

まずいくつか記号を導入する.

- 2 つのベクトル $x, y \in \mathbb{R}^d$ に対して, $x_j \leq y_j$ がすべての $j = 1, \dots, d$ について成り立つことを記号 $x \leq y$ で表す.
- 2 つの実数 a, b に対して, ある正の普遍定数 C が存在して $a \leq Cb$ が成り立つことを, 記号 $a \lesssim b$ で表す.
- 標準正規分布の確率密度関数を ϕ, 累積分布関数を Φ でそれぞれ表す.
- r 回微分可能な関数 $f : \mathbb{R}^d \to \mathbb{R}$ と $j_1, \dots, j_r \in \{1, \dots, d\}$ に対して, 偏導関数

$$\frac{\partial^r f}{\partial x_{j_1} \cdots \partial x_{j_r}}$$

  のことを $\partial_{j_1,\dots,j_r} f$ と略記する. $j_1 = \cdots = j_r =: j$ の場合はさらに $\partial_{j_1,\dots,j_r} f = \partial_j^r f$ と略記する.
- C^k 級関数 $f : \mathbb{R}^d \to \mathbb{R}$ で, 任意の $r \in \{1, \dots, k\}$ と $j_1, \dots, j_r \in \{1, \dots, d\}$ に対して $\partial_{j_1,\dots,j_r} f$ が有界であるようなもの全体の集合を $C_{b^*}^k(\mathbb{R}^d)$ で表す (f 自身は有界でなくてもよいことに注意). $C_{b^*}^k(\mathbb{R}^d)$ に属する関数のうち有界なもの全体の集合を $C_b^k(\mathbb{R}^d)$ で表す.

- 微分可能な関数 $f:\mathbb{R}^d\to\mathbb{R}$ に対して, f の勾配を ∇f で表す: $\nabla f(x)=(\partial_1 f(x),\dots,\partial_d f(x))^\top$. f が 2 回微分可能な場合, f の Hesse 行列を $\mathrm{Hess}\, f$ で表す: $\mathrm{Hess}\, f(x)=(\partial_{ij}f(x))_{1\le i,j\le d}$.
- 関数 $h:\mathbb{R}^m\to\mathbb{R}^n$ に対して, $\|h\|_\infty:=\sup_{x\in\mathbb{R}^m}|h(x)|$ とおく.
- 実数の族 $a=(a_{i_1,\dots,i_k})_{1\le i_1,\dots,i_k\le d}$ および $b=(b_{i_1,\dots,i_k})_{1\le i_1,\dots,i_k\le d}$ が与えられたとき,

$$a\pm b:=(a_{i_1,\dots,i_k}\pm b_{i_1,\dots,i_k})_{1\le i_1,\dots,i_k\le d},\qquad ca:=(ca_{i_1,\dots,i_k})_{1\le i_1,\dots,i_k\le d}\quad (c\in\mathbb{R}),$$

$$\langle a,b\rangle:=\sum_{i_1,\dots,i_k=1}^{d}a_{i_1,\dots,i_k}b_{i_1,\dots,i_k},\qquad |a|:=\sqrt{\langle a,a\rangle}=\sqrt{\sum_{i_1,\dots,i_k=1}^{d}a_{i_1,\dots,i_k}^2}$$

  と定める. $k=2$ の場合は $|a|=\|a\|_F$ となることに注意. また, 便宜上, 実数 $a,b\in\mathbb{R}$ に対して $\langle a,b\rangle:=ab$ と定めることにする.
- 2 つの実数の族 $A=(A_{i_1,\dots,i_k})_{1\le i_1,\dots,i_k\le d}$ および $B=(B_{j_1,\dots,j_l})_{1\le j_1,\dots,j_l\le d}$ が与えられたとき,

$$A\otimes B:=(A_{i_1,\dots,i_k}B_{j_1,\dots,j_l})_{1\le i_1,\dots,i_k,j_1,\dots,j_l\le d}$$

  と定義する. 便宜上 $c\in\mathbb{R}$ に対して $c\otimes A:=A\otimes c:=cA$ と定める. 特に, k 個の d 次元ベクトル $x_1,\dots,x_k\in\mathbb{R}^d$ に対しては,

$$x_1\otimes\cdots\otimes x_k=(x_{1,j_1}\cdots x_{k,j_d})_{1\le j_1,\dots,j_k\le d}$$

  となる. ただし, 各 i について $x_{i,j}$ は x_i の第 j 成分を表す. $x_1=\cdots=x_d=:x$ の場合は $x_1\otimes\cdots\otimes x_k=x^{\otimes k}$ と書くことにする. なお, 便宜上 $x^{\otimes 0}:=1$ と定める.
- 関数 $f:\mathbb{R}^d\to\mathbb{R}$ が点 $x\in\mathbb{R}^d$ で k 回微分可能なとき,

$$\nabla^k f(x):=(\partial j_1,\dots,j_k f(x))_{1\le i_1,\dots,i_k\le d}$$

  と定義する. $\nabla^1 f(x)=\nabla f(x)$ および $\nabla^2 f(x)=\mathrm{Hess}\, f(x)$ となることに注意. なお, 便宜上 $\nabla^0 f:=f$ と定める.

次に, 定理 9.1 を示すには, 定理 9.1 の仮定の下で

$$\sup_{x\in\mathbb{R}^d}|\mathrm{P}(S_n\le x)-\mathrm{P}(Z_n\le x)|\lesssim\frac{1}{b}\left(\frac{B_n^2\log^5(dn)}{n}\right)^{1/4}\tag{9.12}$$

が成り立つことを示せば十分であることに注意する. 実際, 一般に d 次元確率変数 Y に対して $Y^\diamond:=(Y^\top,-Y^\top)^\top$ と書くことにすると, $X_1^\diamond,\dots,X_n^\diamond$ も定理 9.1 の仮定を満たすから, (9.12) より

$$\sup_{x\in\mathbb{R}^{2d}}|\mathrm{P}(S_n^\diamond\le x)-\mathrm{P}(Z_n^\diamond\le x)|\lesssim\frac{1}{b}\left(\frac{B_n^2\log^5(dn)}{n}\right)^{1/4}$$

が成り立つ.

$$\sup_{x\in\mathbb{R}^{2d}}|\mathrm{P}(S_n^\diamond\le x)-\mathrm{P}(Z_n^\diamond\le x)|=\sup_{A\in\mathcal{R}_d}|\mathrm{P}(S_n\in A)-\mathrm{P}(Z_n\in A)|$$

が成り立つことが容易に確認できるから, (9.2) が従う. そこで, 以下では (9.12) が成立することを証明する.

9.4.1　反集中不等式

一般に, 2 つの d 次元確率変数 W, Z と Borel 集合 $A \subset \mathbb{R}^d$ に対して $|\,\mathrm{P}(W \in A) - \mathrm{P}(Z \in A)|$ を評価する問題では, W か Z, あるいはその両方が A の境界に「集中しない」ことを示す評価, いわゆる**反集中不等式**が重要な役割を果たす.[*3] いまの場合, d 次元正規確率変数 Z が特定の超矩形の境界に「集中しない」ことを示す反集中不等式が必要となるが, そのような結果として次が知られている:

定理 9.6 (Nazarov の不等式)**.** Z を平均 0 の d 次元正規確率変数で $\underline{\sigma} := \min_{1\leq j\leq d} \sqrt{\mathrm{E}[Z_j^2]} > 0$ を満たすものとする. このとき, 任意の $y \in \mathbb{R}^d$ と $\varepsilon > 0$ に対して

$$\mathrm{P}(Z \leq y + \varepsilon) - \mathrm{P}(Z \leq y) \leq \frac{\varepsilon}{\underline{\sigma}}\left(\sqrt{2\log d} + 2\right)$$

が成り立つ.

ここでは (Deng & Zhang, 2020, Theorem 10) の証明に倣った議論で定理 9.6 を示す. そのために 2 つ補題を準備する. 1 つ目の補題は Birnbaum (1942) による Mills 比に対する古典的な不等式である.

補題 9.1 (Birnbaum の不等式)**.** 任意の $x \geq 0$ に対して

$$\frac{\phi(x)}{1 - \Phi(x)} \leq \frac{x + \sqrt{x^2 + 4}}{2} \leq 1 + x.$$

証明. 右側の不等式は, 任意の $a, b \geq 0$ に対して $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ が成り立つことから従う. 一方で, Z を標準正規確率変数とすると, Schwarz の不等式より

$$\mathrm{E}[Z1_{\{Z>x\}}]^2 \leq \mathrm{E}[Z^2 1_{\{Z>x\}}]\,\mathrm{P}(Z > x) = \mathrm{E}[Z^2 1_{\{Z>x\}}]\{1 - \Phi(x)\}$$

が成り立つ. ここで,

$$\mathrm{E}[Z1_{\{Z>x\}}] = \int_x^\infty z\phi(z)dz = \phi(x)$$

および

$$\mathrm{E}[Z^2 1_{\{Z>x\}}] = \int_x^\infty z^2\phi(z)dz = x\phi(x) + \int_x^\infty \phi(z)dz = x\phi(x) + 1 - \Phi(x)$$

であるから,

$$\phi(x)^2 \leq x\phi(x)\{1 - \Phi(x)\} + \{1 - \Phi(x)\}^2$$

を得る. 従って, $h(x) = \phi(x)/\{1 - \Phi(x)\}$ とおくと,

$$h(x)^2 \leq xh(x) + 1$$

[*3] Stein の方法の文献では, このような評価を「集中不等式」と呼ぶこともある. 例えば (Chen, Shao & Goldstein, 2011, Section 3.4.1) 参照.

が成り立つ. 実数 y が不等式 $y^2 \le xy + 1$ を満たすための必要十分条件は $(x - \sqrt{x^2+4})/2 \le y \le (x + \sqrt{x^2+4})/2$ であることから, 左側の不等式を得る. $\square$

2 つ目の補題は正規確率ベクトルの最大値の密度の明示的な表現であり, こちらも古典的である.

補題 9.2 (Chernozhukov, Chetverikov & Kato (2015), Lemma 5)**.** $\zeta = (\zeta_1, \dots, \zeta_d)^\top$ を d 次元正規確率変数で $\mathrm{Var}[\zeta_j] = 1$ $(j = 1, \dots, d)$ を満たすものとし, $\mu_j := \mathrm{E}[\zeta_j]$ $(j = 1, \dots, d)$ とおく. $j \neq k$ なる任意の $j, k \in \{1, \dots, d\}$ に対して $\rho_{jk} := \mathrm{Corr}[\zeta_j, \zeta_k] < 1$ が成り立つならば, $\max_{1\le j\le d} \zeta_j$ は次式で与えられる関数 $f : \mathbb{R} \to [0, \infty)$ を密度にもつ:

$$f(x) = \sum_{j=1}^d \phi(x - \mu_j)\,\mathrm{P}(\zeta_k - \rho_{jk}\zeta_j \le (1 - \rho_{jk})x \text{ for all } k \neq j), \quad x \in \mathbb{R}. \tag{9.13}$$

証明. まず, $j \neq k$ なる任意の $j, k \in \{1, \dots, d\}$ に対して $\mathrm{P}(\zeta_j = \zeta_k) = 0$ が成り立つことを示す. 実際, $\zeta_j' := \zeta_j - \rho_{jk}\zeta_k$ とおくと, ζ_j' と ζ_k の共分散は 0 で, かつその結合分布は正規だから, 独立である. さらに,

$$\mathrm{Var}[\zeta_j'] = \mathrm{Var}[\zeta_j] - 2\rho_{jk}\,\mathrm{Cov}[\zeta_j, \zeta_k] + \rho_{jk}^2\,\mathrm{Var}[\zeta_k] = 1 - \rho_{jk}^2$$

であるから, $|\rho_{jk}| < 1$ ならば ζ_j' は密度を持つ. 故に, この場合は任意の $x \in \mathbb{R}$ に対して $\mathrm{P}(\zeta_j' = x) = 0$ が成り立つから, $\mu_k := \mathrm{E}[\zeta_k]$ とおくと,

$$\mathrm{P}(\zeta_j = \zeta_k) = \mathrm{P}(\zeta_j' = (1 - \rho_{jk})\zeta_k) = \int_{-\infty}^{\infty} \mathrm{P}(\zeta_j' = (1 - \rho_{jk})z)\phi(z - \mu_k)dz = 0$$

となる. 一方で, $|\rho_{jk}| = 1$ の場合は, 仮定より $\rho_{jk} = -1$ でなければならないから, $\zeta_k - \mu_k = -(\zeta_j - \mu_j)$ である. 従って $\mathrm{P}(\zeta_j = \zeta_k) = \mathrm{P}(2\zeta_j = \mu_k + \mu_j) = 0$ となる.

次に, 任意に $x \in \mathbb{R}$ を固定する.

$$\mathrm{P}\left(\max_{1\le j\le d} \zeta_j \le x\right) = \int_{-\infty}^x f(z)dz \tag{9.14}$$

が成り立つことを示せば証明は完成する. 上で示したことから,

$$\mathrm{P}\left(\max_{1\le j\le d} \zeta_j \le x\right) = \sum_{j=1}^d \mathrm{P}\left(\zeta_j = \max_{1\le k\le d} \zeta_k, \zeta_j \le x\right) \tag{9.15}$$

が成り立つ. 各 $j = 1, \dots, d$ について,

$$\begin{aligned}\mathrm{P}\left(\zeta_j = \max_{1\le k\le d} \zeta_k, \zeta_j \le x\right) &= \mathrm{P}\left(\zeta_k \le \zeta_j \text{ for all } k \neq j, \zeta_j \le x\right) \\ &= \mathrm{P}\left(\zeta_k' \le (1 - \rho_{jk})\zeta_j \text{ for all } k \neq j, \zeta_j \le x\right)\end{aligned}$$

と書けるが, ζ_j と $\{\zeta_k' : k \neq j\}$ は独立であるから,

$$\mathrm{P}\left(\zeta_j = \max_{1\le k\le d} \zeta_k, \zeta_j \le x\right) = \int_{-\infty}^x \mathrm{P}\left(\zeta_k' \le (1 - \rho_{jk})z \text{ for all } k \neq j\right)\phi(z - \mu_j)dz$$

を得る. この式を (9.15) に代入して (9.14) を得る. $\square$

注意 9.3. (9.13) 式は,

$$f(x)=\sum_{j=1}^{d}\phi(x-\mu_j)\,\mathrm{P}(\zeta_k\leq x \text{ for all } k\neq j \mid \zeta_j=x),\quad x\in\mathbb{R}.$$

と書き直せる. 実際, 各 $j=1,\dots,d$ について, ζ_j と $\{\zeta_k-\rho_{jk}\zeta_j:k\neq j\}$ が独立であることに注意すると,

$$\begin{aligned}\mathrm{P}(\zeta_k\leq x \text{ for all } k\neq j \mid \zeta_j=x)&=\mathrm{P}(\zeta_k-\rho_{jk}\zeta_j\leq(1-\rho_{jk})x \text{ for all } k\neq j \mid \zeta_j=x)\\&=\mathrm{P}(\zeta_k-\rho_{jk}\zeta_j\leq(1-\rho_{jk})x \text{ for all } k\neq j)\end{aligned}$$

が成り立つ.

定理 9.6 の証明. 各 $j=1,\dots,d$ について, $\sigma_j:=\sqrt{\mathrm{E}[Z_j^2]}$, $\zeta_j:=(Z_j-y_j)/\sigma_j$ とおくと

$$\mathrm{P}(Z\leq y+\varepsilon)-\mathrm{P}(Z\leq y)\leq\mathrm{P}\left(0<\max_{1\leq j\leq d}\zeta_j\leq\frac{\varepsilon}{\underline{\sigma}}\right)$$

が成り立つ. 従って, 任意の $\varepsilon>0$ に対して

$$\mathrm{P}\left(0<\max_{1\leq j\leq d}\zeta_j\leq\varepsilon\right)\leq\varepsilon\left(\sqrt{2\log d}+2\right)$$

が成り立つことを示せばよい.

相異なる $j_1,j_2\in\{1,\dots,d\}$ で $\mathrm{Corr}(\zeta_{j_1},\zeta_{j_2})=1$ かつ $y_{j_1}/\sigma_{j_1}\leq y_{j_2}/\sigma_{j_2}$ なるものがあった場合, $\max_{1\leq j\leq d}\zeta_j=\max_{1\leq j\leq d:j\neq j_2}\zeta_j$ が成り立つことに注意すると, $j\neq k$ なる任意の $j,k\in\{1,\dots,d\}$ に対して $\rho_{jk}:=\mathrm{Corr}[\zeta_j,\zeta_k]<1$ が成り立つと仮定して一般性を失わない. $(\zeta_1,\dots,\zeta_d)^\top$ は d 次元正規確率変数であり, $\mathrm{Var}[\zeta_j]=1$ $(j=1,\dots,d)$ を満たすから, 補題 9.2 より, $\max_{1\leq j\leq d}\zeta_j$ は (9.13) 式で与えられる密度 f を持つ. 従って, 任意の $z\in\mathbb{R}$ に対して

$$f(z)\leq\sqrt{2\log d}+2$$

を示せば証明は完成する.

任意に $j\in\{1,\dots,d\}$ を固定する. ζ_j と $\{\zeta_k-\rho_{jk}\zeta_j:k\neq j\}$ は独立だから, $\zeta_j':=\zeta_j-z$, $\mu_j:=\mathrm{E}[\zeta_j]-z=\mathrm{E}[\zeta_j']$ とおくと,

$$\begin{aligned}&\mathrm{P}(\zeta_k-\rho_{jk}\zeta_j\leq(1-\rho_{jk})z \text{ for all } k\neq j)\{1-\Phi(-\mu_j)\}\\&=\mathrm{P}(\zeta_k'-\rho_{jk}\zeta_j'\leq 0 \text{ for all } k\neq j)\,\mathrm{P}(\zeta_j'>0)\\&=\mathrm{P}(\zeta_k'-\rho_{jk}\zeta_j'\leq 0 \text{ for any } k\neq j,\zeta_j'>0)\\&\leq\mathrm{P}(\zeta_k'<\zeta_j' \text{ for any } k\neq j,\zeta_j'>0)=:\mathrm{P}(A_j)\end{aligned}$$

が成り立つ. 従って,

$$f(z)=\sum_{j=1}^{d}\phi(-\mu_j)\,\mathrm{P}(\zeta_k-\rho_{jk}\zeta_j\leq(1-\rho_{jk})z \text{ for all } k\neq j)$$

$$\le \sum_{j=1}^{d} \phi(-\mu_j) \min\left\{1, \frac{\mathrm{P}(A_j)}{1-\Phi(-\mu_j)}\right\}$$
$$\le \sum_{j=1}^{d} \left(\sup_{x>\sqrt{2\log d}} \phi(x) + \sup_{x\le\sqrt{2\log d}} \frac{\phi(x)}{1-\Phi(x)} \mathrm{P}(A_j) \right)$$

を得る. ここで,

$$\sup_{x>\sqrt{2\log d}} \phi(x) = \phi(\sqrt{2\log d}) \le \frac{1}{d},$$
$$\sup_{x\le 0} \frac{\phi(x)}{1-\Phi(x)} = \frac{\phi(0)}{1-\Phi(0)} = \frac{2}{\sqrt{2\pi}} < 1$$

であり, また, Birnbaum の不等式より

$$\sup_{0<x\le\sqrt{2\log d}} \frac{\phi(x)}{1-\Phi(x)} \le 1 + \sqrt{2\log d}$$

が成り立つ. 従って,

$$f(z) \le 1 + (1+\sqrt{2\log d}) \sum_{j=1}^{d} \mathrm{P}(A_j)$$

を得る. ここで, $A_1, \dots, A_d$ は明らかに disjoint であるから, $\sum_{j=1}^{d} \mathrm{P}(A_j) = \mathrm{P}(\bigcup_{j=1}^{d} A_j) \le 1$ である. 故に, $f(z) \le 2 + \sqrt{2\log d}$ である. □

9.4.2 最大値関数の平滑化

十分滑らかな関数 $h : \mathbb{R}^d \to \mathbb{R}$ に対して $|\mathrm{E}\, h(S_n) - \mathrm{E}\, h(Z_n)|$ を評価する方法は様々なものが知られている. 一般に, d 次元確率変数 W とベクトル $x \in \mathbb{R}^d$ に対して,

$$\mathrm{P}(W \le x) = \mathrm{E}\left[1_{(-\infty,0]}\left(\max_{1\le j\le d}(W_j - x_j)\right)\right]$$

と書き直せる. そこで, 右辺の期待値の中身を W の滑らかな関数で,「次元 d に関して効率的に」近似することを考える. 指示関数 $1_{(-\infty,0]}$ は d とは無関係だから標準的な方法で平滑化すればよい. 従って, 最大値関数 $w \mapsto \max_{1\le j\le d} w_j$ をうまく平滑化する必要がある. そのような平滑化関数として, Chernozhukov, Chetverikov & Kato (2013) は次の関数を用いるのが有効であることを示した: $\beta > 0$ に対して, 関数 $F_\beta : \mathbb{R}^d \to \mathbb{R}$ を

$$F_\beta(w) = \beta^{-1} \log\left(\sum_{j=1}^{d} e^{\beta w_j}\right) \qquad (w \in \mathbb{R}^d)$$

で定める. 以下の補題から, β を大きくすると F_β は最大値関数をよく近似する:

補題 9.3. 任意の $\beta > 0$ と $w \in \mathbb{R}^d$ に対して,

$$\max_{1\le j\le d} w_j \le F_\beta(w) \le \max_{1\le j\le d} w_j + \beta^{-1} \log d.$$

演習問題 2. 補題 9.3 を証明せよ.

F_β の微分はいくつかの望ましい性質を持つ.

補題 9.4. 任意の $\beta > 0, j \in \{1, \ldots, d\}$, および $w, w' \in \mathbb{R}^d$ に対して次が成り立つ.

(a) $\partial_j F_\beta(w) \geq 0$.

(b) $\partial_j F_\beta(w + w') \leq e^{2\beta|w'|_\infty} \partial_j F_\beta(w)$.

(c) $\sum_{k=1}^d \partial_k F_\beta(w) = 1$.

証明. 直接計算によって,

$$\partial_j F_\beta(w) = \frac{e^{\beta w_j}}{\sum_{k=1}^d e^{\beta w_k}} \tag{9.16}$$

を得る. この表現から直ちに示すべき主張が従う. □

補題 9.5. 任意の $\beta > 0$ に対して, $\mathbb{R}^d$ 上の非負値関数の族 $(U^0_{jk})_{1 \leq j,k \leq d}$, $(U^0_{jkl})_{1 \leq j,k,l \leq d}$, $(U^0_{jklm})_{1 \leq j,k,l,m \leq d}$ が存在して以下の性質を満たす:

(i) 任意の $w \in \mathbb{R}^d$ と $j, k, l, m \in \{1, \ldots, d\}$ に対して,

$$|\partial_{jk} F_\beta(w)| \leq U^0_{jk}(w), \qquad |\partial_{jkl} F_\beta(w)| \leq U^0_{jkl}(w), \qquad |\partial_{jklm} F_\beta(w)| \leq U^0_{jklm}(w).$$

(ii) ある普遍定数 $C > 0$ が存在して, 任意の $w \in \mathbb{R}^d$ に対して

$$\sum_{j,k=1}^d U^0_{jk}(w) \leq C\beta, \qquad \sum_{j,k,l=1}^d U^0_{jkl}(w) \leq C\beta^2, \qquad \sum_{j,k,l,m=1}^d U^0_{jklm}(w) \leq C\beta^3.$$

(iii) ある普遍定数 $C' > 0$ が存在して, 任意の $w, w' \in \mathbb{R}^d$ と $j, k, l, m \in \{1, \ldots, d\}$ に対して, $\beta|w'|_\infty \leq 1$ ならば,

$$U^0_{jk}(w + w') \leq C' U^0_{jk}(w), \quad U^0_{jkl}(w + w') \leq C' U^0_{jkl}(w), \quad U^0_{jklm}(w + w') \leq C' U^0_{jklm}(w).$$

証明. (9.16) 式を微分して,

$$\begin{aligned}
\partial_{jk} F_\beta(w) &= \beta \left\{ \partial_j F_\beta(w) 1_{\{j=k\}} - \partial_j F_\beta(w) \partial_k F_\beta(w) \right\}, \\
\partial_{jkl} F_\beta(w) &= \beta \left\{ \partial_{jl} F_\beta(w) 1_{\{j=k\}} - \partial_{jl} F_\beta(w) \partial_k F_\beta(w) - \partial_j F_\beta(w) \partial_{kl} F_\beta(w) \right\}
\end{aligned}$$

および

$$\begin{aligned}
\partial_{jklm} F_\beta(w) = \beta\{&\partial_{jlm} F_\beta(w) 1_{\{j=k\}} - \partial_{jlm} F_\beta(w) \partial_k F_\beta(w) \\
&- \partial_{jl} F_\beta(w) \partial_{km} F_\beta(w) - \partial_{jm} F_\beta(w) \partial_{kl} F_\beta(w) - \partial_j F_\beta(w) \partial_{klm} F_\beta(w)\}
\end{aligned}$$

を得る. 従って,

$$U^0_{jk}(w) = \beta \left\{ \partial_j F_\beta(w) 1_{\{j=k\}} + \partial_j F_\beta(w) \partial_k F_\beta(w) \right\},$$

$$U^0_{jkl}(w) = \beta \left\{ U^0_{jl}(w) 1_{\{j=k\}} + U^0_{jl}(w) \partial_k F_\beta(w) + \partial_j F_\beta(w) U^0_{kl}(w) \right\}$$

および

$$\begin{aligned} U^0_{jklm}(w) = \beta\{ & U^0_{jlm}(w) 1_{\{j=k\}} + U^0_{jlm}(w) \partial_k F_\beta(w) \\ & + U^0_{jl}(w) U^0_{km}(w) + U^0_{jm}(w) U^0_{kl}(w) + \partial_j F_\beta(w) U^0_{klm}(w) \} \end{aligned}$$

と定めれば, 補題 9.4 より性質 (i)–(iii) が満たされる. □

関数 F_β と Nazarov の不等式を用いると, 以下の形の「平滑化不等式」が得られる.

命題 9.2. $g_0 : \mathbb{R} \to [0,1]$ を可測関数とし, $t \le 0$ のとき $g_0(t) = 1, t \ge 1$ のとき $g_0(t) = 0$ を満たすものとする. また, Z を平均 0 の d 次元正規確率変数で $\underline{\sigma} := \min_{1 \le j \le d} \sqrt{\mathrm{E}[Z_j^2]} > 0$ を満たすものとする. さらに, $\varepsilon > 0$ とし, $\beta := \varepsilon^{-1} \log d$ とおく. このとき, 任意の d 次元確率変数 W に対して,

$$\sup_{x \in \mathbb{R}^d} |\mathrm{P}(W \le x) - \mathrm{P}(Z \le x)| \le \Delta_\varepsilon(W, Z) + \frac{2\varepsilon}{\underline{\sigma}} \left(\sqrt{2 \log d} + 2 \right) \tag{9.17}$$

が成り立つ. ここに,

$$\Delta_\varepsilon(W, Z) := \sup_{y \in \mathbb{R}^d} \left| \mathrm{E}[g_0(\varepsilon^{-1} F_\beta(W - y))] - \mathrm{E}[g_0(\varepsilon^{-1} F_\beta(Z - y))] \right| \tag{9.18}$$

と定める.

証明. $x \in \mathbb{R}^d$ を任意に 1 つ固定する. このとき,

$$\begin{aligned} \mathrm{P}(W \le x) &= \mathrm{P}\left(\max_{1 \le j \le d} (W_j - x_j - \varepsilon) + \varepsilon \le 0 \right) \\ &\le \mathrm{P}(F_\beta(W - x - \varepsilon) \le 0) \qquad (\because \text{補題 } 9.3) \\ &\le \mathrm{E}[g_0(\varepsilon^{-1} F_\beta(W - x - \varepsilon))] \qquad (\because g_0 \ge 1_{(-\infty, 0]}) \\ &\le \mathrm{E}[g_0(\varepsilon^{-1} F_\beta(Z - x - \varepsilon))] + \Delta_\varepsilon(W, Z) \\ &\le \mathrm{P}(F_\beta(Z - x - \varepsilon) < \varepsilon) + \Delta_\varepsilon(W, Z) \qquad (\because g_0 \le 1_{(-\infty, 1]}) \\ &\le \mathrm{P}(Z \le x + 2\varepsilon) + \Delta_\varepsilon(W, Z) \qquad (\because \text{補題 } 9.3) \end{aligned}$$

が成り立つ. 従って, Nazarov の不等式より

$$\mathrm{P}(W \le x) \le \mathrm{P}(Z \le x) + \Delta_\varepsilon(W, Z) + \frac{2\varepsilon}{\underline{\sigma}} \left(\sqrt{2 \log d} + 2 \right)$$

を得る. 同様の議論によって

$$\mathrm{P}(W \le x) \ge \mathrm{P}(Z \le x) - \Delta_\varepsilon(W, Z) - \frac{2\varepsilon}{\underline{\sigma}} \left(\sqrt{2 \log d} + 2 \right)$$

も得られる. □

命題 9.2 において現れる量 $\Delta_\varepsilon(W, Z)$ を評価するには, 関数 $w \mapsto g_0(\varepsilon^{-1} F_\beta(w - y))$ の微分の評価が必要になるが, これは以下の補題で与えられる:

補題 9.6. $g_0 : \mathbb{R} \to \mathbb{R}$ を C^4 級関数とし,

$$C_g := \max_{r=0,1,2,3,4} \sup_{t\in\mathbb{R}} |g_0^{(r)}(t)| < \infty$$

が成り立つと仮定する. また, $\varepsilon > 0, y \in \mathbb{R}^d$ とし, $\beta := \varepsilon^{-1}\log d$ と定め, 関数 $h : \mathbb{R}^d \to \mathbb{R}$ を $h(w) = g_0(\varepsilon^{-1}F_\beta(w-y))$ $(w \in \mathbb{R}^d)$ で定める. このとき, $\mathbb{R}^d$ 上の非負値関数の族 $(U_{jk})_{1\leq j,k\leq d}$, $(U_{jkl})_{1\leq j,k,l\leq d}, (U_{jklm})_{1\leq j,k,l,m\leq d}$ が存在して以下の性質を満たす:

(i) 任意の $w \in \mathbb{R}^d$ と $j,k,l,m \in \{1,\dots,d\}$ に対して,

$$|\partial_{jk}h(w)| \leq U_{jk}(w), \qquad |\partial_{jkl}h(w)| \leq U_{jkl}(w), \qquad |\partial_{jklm}h(w)| \leq U_{jklm}(w).$$

(ii) C_g にのみ依存する定数 $C > 0$ が存在して, 任意の $w \in \mathbb{R}^d$ に対して

$$\sum_{j,k=1}^d U_{jk}(w) \leq C\varepsilon^{-2}\log d, \qquad \sum_{j,k,l=1}^d U_{jkl}(w) \leq C\varepsilon^{-3}\log^2 d,$$
$$\sum_{j,k,l,m=1}^d U_{jklm}(w) \leq C\varepsilon^{-4}\log^3 d.$$

(iii) ある普遍定数 $C' > 0$ が存在して, 任意の $w, w' \in \mathbb{R}^d$ と $j,k,l,m \in \{1,\dots,d\}$ に対して, $\beta|w'|_\infty \leq 1$ ならば,

$$U_{jk}(w+w') \leq C'U_{jk}(w), \quad U_{jkl}(w+w') \leq C'U_{jkl}(w), \quad U_{jklm}(w+w') \leq C'U_{jklm}(w).$$

証明. $y = 0$ の場合に条件 (i)–(iii) を満たす $\mathbb{R}^d$ 上の非負値関数の族を $(U_{0,jk})_{1\leq j,k\leq d}$, $(U_{0,jkl})_{1\leq j,k,l\leq d}$, $(U_{0,jklm})_{1\leq j,k,l,m\leq d}$ と書くことにすれば, 一般の y の場合は $U_{jk}(w) = U_{0,jk}(w-y), U_{jkl}(w) = U_{0,jkl}(w-y), U_{jklm}(w) = U_{0,jklm}(w-y)$ と定めれば条件 (i)–(iii) を満たす $\mathbb{R}^d$ 上の非負値関数の族が得られる. 従って $y=0$ の場合を考えれば十分である. このとき,

$$\begin{aligned}
\partial_{jk}h(w) &= \varepsilon^{-2}g_0''(\varepsilon^{-1}F_\beta(w))\partial_jF_\beta(w)\partial_kF_\beta(w) + \varepsilon^{-1}g_0'(\varepsilon^{-1}F_\beta(w))\partial_{jk}F_\beta(w),\\
\partial_{jkl}h(w) &= \varepsilon^{-3}g_0^{(3)}(\varepsilon^{-1}F_\beta(w))\partial_jF_\beta(w)\partial_kF_\beta(w)\partial_lF_\beta(w)\\
&\quad + \varepsilon^{-2}g_0''(\varepsilon^{-1}F_\beta(w))\left\{\partial_{jl}F_\beta(w)\partial_kF_\beta(w) + \partial_jF_\beta(w)\partial_{kl}F_\beta(w) + \partial_{jk}F_\beta(w)\partial_lF_\beta(w)\right\}\\
&\quad + \varepsilon^{-1}g_0'(\varepsilon^{-1}F_\beta(w))\partial_{jkl}F_\beta(w),
\end{aligned}$$

および

$$\begin{aligned}
\partial_{jkl}h(w) &= \varepsilon^{-4}g_0^{(4)}(\varepsilon^{-1}F_\beta(w))\partial_jF_\beta(w)\partial_kF_\beta(w)\partial_lF_\beta(w)\partial_mF_\beta(w)\\
&\quad + \varepsilon^{-3}g_0^{(3)}(\varepsilon^{-1}F_\beta(w))\{\partial_{jm}F_\beta(w)\partial_kF_\beta(w)\partial_lF_\beta(w) + \partial_jF_\beta(w)\partial_{km}F_\beta(w)\partial_lF_\beta(w)\\
&\qquad + \partial_jF_\beta(w)\partial_kF_\beta(w)\partial_{lm}F_\beta(w) + \partial_mF_\beta(w)\partial_{jl}F_\beta(w)\partial_kF_\beta(w)\\
&\qquad + \partial_mF_\beta(w)\partial_jF_\beta(w)\partial_{kl}F_\beta(w) + \partial_mF_\beta(w)\partial_{jk}F_\beta(w)\partial_lF_\beta(w)\}\\
&\quad + \varepsilon^{-2}g_0''(\varepsilon^{-1}F_\beta(w))\{\partial_{jlm}F_\beta(w)\partial_kF_\beta(w) + \partial_{jl}F_\beta(w)\partial_{km}F_\beta(w)\\
&\qquad + \partial_{jm}F_\beta(w)\partial_{kl}F_\beta(w) + \partial_jF_\beta(w)\partial_{klm}F_\beta(w)
\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad + \partial_{jkm}F_\beta(w)\partial_l F_\beta(w) + \partial_{jk}F_\beta(w)\partial_{lm}F_\beta(w) \\
&\qquad\qquad + \partial_m F_\beta(w)\partial_{jkl}F_\beta(w)\} \\
&+ \varepsilon^{-1} g_0'(\varepsilon^{-1}F_\beta(w))\partial_{jklm}F_\beta(w)
\end{aligned}
$$

が成り立つ. 従って, 補題 9.4 で与えられる $\mathbb{R}^d$ 上の非負値関数の族を $(U^0_{jk})_{1\leq j,k\leq d}$, $(U^0_{jkl})_{1\leq j,k,l\leq d}$, $(U^0_{jklm})_{1\leq j,k,l,m\leq d}$ として,

$$
\begin{aligned}
U_{jk}(w) &= \varepsilon^{-2}C_g\partial_j F_\beta(w)\partial_k F_\beta(w) + \varepsilon^{-1}C_g U^0_{jk}(w), \\
U_{jkl}(w) &= \varepsilon^{-3}C_g\partial_j F_\beta(w)\partial_k F_\beta(w)\partial_l F_\beta(w) \\
&\quad + \varepsilon^{-2}C_g\left\{U^0_{jl}(w)\partial_k F_\beta(w) + \partial_j F_\beta(w)U^0_{kl}F_\beta(w) + U^0_{jk}F_\beta(w)\partial_l F_\beta(w)\right\} \\
&\quad + \varepsilon^{-1}C_g U^0_{jkl}(w),
\end{aligned}
$$

および

$$
\begin{aligned}
U_{jkl}(w) &= \varepsilon^{-4}C_g\partial_j F_\beta(w)\partial_k F_\beta(w)\partial_l F_\beta(w)\partial_m F_\beta(w) \\
&\quad + \varepsilon^{-3}C_g\{U^0_{jm}(w)\partial_k F_\beta(w)\partial_l F_\beta(w) + \partial_j F_\beta(w)U^0_{km}(w)\partial_l F_\beta(w) \\
&\qquad\qquad + \partial_j F_\beta(w)\partial_k F_\beta(w)U^0_{lm}(w) + \partial_m F_\beta(w)U^0_{jl}(w)\partial_k F_\beta(w) \\
&\qquad + \partial_m F_\beta(w)\partial_j F_\beta(w)U^0_{kl}(w) + \partial_m F_\beta(w)U^0_{jk}(w)\partial_l F_\beta(w)\} \\
&\quad + \varepsilon^{-2}C_g\{U^0_{jlm}(w)\partial_k F_\beta(w) + U^0_{jl}(w)U^0_{km}(w) + U^0_{jm}(w)U^0_{kl}(w) \\
&\qquad\qquad + \partial_j F_\beta(w)U^0_{klm}(w) + U^0_{jkm}(w)\partial_l F_\beta(w) + U^0_{jk}(w)U^0_{lm}(w) \\
&\qquad\qquad + \partial_m F_\beta(w)U^0_{jkl}(w)\} \\
&\quad + \varepsilon^{-1}C_g U^0_{jklm}(w)
\end{aligned}
$$

と定めれば, 補題 9.4–9.5 と β の定義より性質 (i)–(iii) が満たされる. □

命題 9.2 の条件を満たす滑らかな関数 g_0 の構成は標準的だが, 例えば以下の構成がよく利用される.

補題 9.7. 関数 $f_0 : \mathbb{R} \to \mathbb{R}$ を

$$
f_0(t) = \begin{cases} e^{-1/t} & \text{if } t > 0, \\ 0 & \text{if } t \leq 0 \end{cases}
$$

で定めて, 関数 $g_0 : \mathbb{R} \to [0,1]$ を

$$
g_0(t) = \frac{f_0(1-t)}{f_0(t) + f_0(1-t)} \qquad (t \in \mathbb{R})
$$

で定める. このとき, g_0 は C^∞ 級であり, $t \leq 0$ のとき $g_0(t) = 1$, $t \geq 1$ のとき $g_0(t) = 0$ を満たす.

演習問題 3. 補題 9.7 を示せ.

9.4.3 Stein の等式

命題 9.2 における $\Delta_\varepsilon(W,Z)$ の評価には **Stein の方法**を用いる. 本小節と次小節・次々小節では, Stein の方法を適用するための準備を行う.

補題 9.8 (Stein の等式: 1 次元の場合)**.** X を標準正規分布に従う確率変数とする. 関数 $f:\mathbb{R}\to\mathbb{R}$ が絶対連続であり, かつ $f'(X)\in L^1(\mathrm{P})$ を満たすならば, $\mathrm{E}[|Xf(X)|]\le|f(0)|\,\mathrm{E}[|X|]+\mathrm{E}[|f'(X)|]<\infty$ であり

$$\mathrm{E}[Xf(X)]=\mathrm{E}[f'(X)] \tag{9.19}$$

が成り立つ.

証明. $\mathrm{E}[|Xf(X)|]\le|f(0)|\,\mathrm{E}[|X|]+\mathrm{E}[|X\{f(X)-f(0)\}|]$ かつ $\mathrm{E}[X]=0$ であるから, $f(0)=0$ の場合に示せば十分である. このとき,

$$\begin{aligned}
\mathrm{E}[|Xf(X)|]&=\int_{-\infty}^{\infty}|xf(x)|\phi(x)dx=\int_{-\infty}^{\infty}|x|\left|\int_0^x f'(y)dy\right|\phi(x)dx\\
&=\int_0^{\infty}x\left|\int_0^x f'(y)dy\right|\phi(x)dx+\int_{-\infty}^0(-x)\left|\int_0^x f'(y)dy\right|\phi(x)dx\\
&\le\int_0^{\infty}x\left(\int_0^x|f'(y)|dy\right)\phi(x)dx+\int_{-\infty}^0(-x)\left(\int_x^0|f'(y)|dy\right)\phi(x)dx\\
&=-\int_0^{\infty}\left(\int_0^x|f'(y)|dy\right)\phi'(x)dx+\int_{-\infty}^0\left(\int_x^0|f'(y)|dy\right)\phi'(x)dx\\
&=-\int_0^{\infty}\left(\int_y^{\infty}\phi'(x)dx\right)|f'(y)|dy+\int_{-\infty}^0\left(\int_{-\infty}^y\phi'(x)dx\right)|f'(y)|dy\ (\because \text{Fubini の定理})\\
&=\int_0^{\infty}|f'(y)|\phi(y)dy+\int_{-\infty}^0|f'(y)|\phi(y)dy=\mathrm{E}[|f'(X)|]<\infty
\end{aligned}$$

であり, また,

$$\begin{aligned}
\mathrm{E}[Xf(X)]&=\int_{-\infty}^{\infty}xf(x)\phi(x)dx=-\int_{-\infty}^{\infty}\left(\int_0^x f'(y)dy\right)\phi'(x)dx\\
&=-\int_0^{\infty}\left(\int_0^x f'(y)dy\right)\phi'(x)dx+\int_{-\infty}^0\left(\int_x^0 f'(y)dy\right)\phi'(x)dx\\
&=-\int_0^{\infty}\left(\int_y^{\infty}\phi'(x)dx\right)f'(y)dy+\int_{-\infty}^0\left(\int_{-\infty}^y\phi'(x)dx\right)f'(y)dy\ (\because \text{Fubini の定理})\\
&=\int_0^{\infty}\phi(y)f'(y)dy+\int_{-\infty}^0\phi(y)f'(y)dy=\mathrm{E}[f'(X)]
\end{aligned}$$

となる. □

補題 9.9 (Stein の等式: 多次元の場合)**.** Z を平均 0 の d 次元正規確率変数とする. また, $j\in\{1,\dots,d\}$ とする. 関数 $f:\mathbb{R}^d\to\mathbb{R}$ が C^1 級であり, かつ $\mathrm{E}[|\nabla f(Z)|]<\infty$ および $\mathrm{E}[|Zf(Z)|]<\infty$ を満たすならば,

$$\mathrm{E}[Z_jf(Z)]=\sum_{k=1}^d\mathrm{E}[Z_jZ_k]\,\mathrm{E}[\partial_kf(Z)] \tag{9.20}$$

が成り立つ.

証明. $j=1$ と仮定して一般性を失わない. $\sigma := \sqrt{\mathrm{E}[Z_1^2]}$ とおく. $\sigma = 0$ の場合は $Z_1 \equiv 0$ となって主張は明らかに成り立つから, $\sigma > 0$ の場合を考える.

$Z_1' := Z_1/\sigma$ とおく. Z_1' は標準正規分布に従う. さらに, 各 $k = 2, \dots, d$ について,

$$Z_k' := Z_k - \mathrm{E}[Z_1' Z_k] Z_1'$$

とおく. 確率ベクトル $(Z_1', \dots, Z_d')$ は d 次元正規分布に従い, かつ $\mathrm{E}[Z_1' Z_k'] = 0$ $(k = 2, \dots, d)$ を満たすから, Z_1' と $\bar{Z} := (Z_2', \dots, Z_d')^\top$ は独立である. いま, 任意に $\bar{z} := (z_2, \dots, z_d)^\top \in \mathbb{R}^{d-1}$ を固定して, 関数 $f_{\bar{z}} : \mathbb{R} \to \mathbb{R}$ を

$$f_{\bar{z}}(z) = f(\sigma z, z_2 + \mathrm{E}[Z_1' Z_2] z, \dots, z_d + \mathrm{E}[Z_1' Z_d] z) \qquad (z \in \mathbb{R})$$

で定める. $f_{\bar{z}}$ は C^1 級であり,

$$\begin{aligned} f_{\bar{z}}'(z) &= \sigma \partial_1 f(\sigma z, z_2 + \mathrm{E}[Z_1' Z_2] z, \dots, z_d + \mathrm{E}[Z_1' Z_d] z) \\ &\quad + \sum_{k=2}^{d} \mathrm{E}[Z_1' Z_k] \partial_k f(\sigma z, z_2 + \mathrm{E}[Z_1' Z_2] z, \dots, z_d + \mathrm{E}[Z_1' Z_d] z) \\ &= \sigma^{-1} \sum_{k=1}^{d} \mathrm{E}[Z_1 Z_k] \partial_k f(\sigma z, z_2 + \mathrm{E}[Z_1' Z_2] z, \dots, z_d + \mathrm{E}[Z_1' Z_d] z) \end{aligned}$$

が成り立つ. 特に,

$$\mathrm{E}[|f_{\bar{Z}}'(Z_1')|] \le \sigma^{-1} \sum_{k=1}^{d} |\mathrm{E}[Z_1 Z_k]| \, \mathrm{E}[|\partial_k f(Z)|] < \infty$$

が成り立つ. ここで, Z_1' と $\bar{Z}$ は独立であるから, $\bar{Z}$ の分布を $\mathrm{P}^{\bar{Z}}$ と書くことにすれば, Fubini の定理より

$$\mathrm{E}[|f_{\bar{Z}}'(Z_1')|] = \int_{\mathbb{R}^{d-1}} \mathrm{E}[|f_{\bar{z}}'(Z_1')|] \, \mathrm{P}^{\bar{Z}}(d\bar{z})$$

が成り立つ. 従って, $\mathrm{P}^{\bar{Z}}$ に関してほとんどすべての $\bar{z} \in \mathbb{R}^{d-1}$ について $\mathrm{E}[|f_{\bar{z}}'(Z_1')|] < \infty$ が成り立つ. そのような $\bar{z}$ について, 補題 9.8 より

$$\mathrm{E}[Z_1' f_{\bar{z}}(Z_1')] = \mathrm{E}[f_{\bar{z}}'(Z_1')]$$

が成り立つ. 従って

$$\int_{\mathbb{R}} \mathrm{E}[Z_1 f_{\bar{z}}(Z_1')] \, \mathrm{P}^{\bar{Z}}(d\bar{z}) = \sigma \int_{\mathbb{R}} \mathrm{E}[f_{\bar{z}}'(Z_1')] \, \mathrm{P}^{\bar{Z}}(d\bar{z}) \tag{9.21}$$

を得る. ここで, Fubini の定理より, 右辺は

$$\sigma \, \mathrm{E}[f_{\bar{Z}}'(Z_1')] = \sum_{k=1}^{d} \mathrm{E}[Z_1 Z_k] \, \mathrm{E}[\partial_k f(Z)]$$

に一致する. 一方で, 再び Fubini の定理より

$$\int_{\mathbb{R}} \mathrm{E}[|Z_1 f_{\bar{z}}(Z_1')|] \, \mathrm{P}^{\bar{Z}}(d\bar{z}) = \mathrm{E}[|Z_1 f_{\bar{Z}}(Z_1')|] = \mathrm{E}[|Z_1 f(Z)|] < \infty$$

が成り立つから, (9.21) の左辺は, 再度 Fubini の定理を適用することで, $\mathrm{E}[Z_1 f_{\bar{Z}}(Z_1')] = \mathrm{E}[Z_1 f(Z)]$ に一致することがわかる. 以上より等式 (9.20) が得られた. □

9.4.4 Slepian 補間

以下特に断らない限り, Σ を d 次半正定値対称行列とし, $Z \sim \mathrm{N}(0, \Sigma)$ とする.

定義 9.1 (Slepian 補間)**.** 可測関数 $h : \mathbb{R}^d \to \mathbb{R}^k$ と $t \in [0,1]$ が任意の $x \in \mathbb{R}^d$ に対して $\mathrm{E}[|h(\sqrt{t}x + \sqrt{1-t}Z)|] < \infty$ を満たすとき, 関数 $S_t^\Sigma h : \mathbb{R}^d \to \mathbb{R}^k$ を

$$S_t^\Sigma h(x) = \mathrm{E}[h(\sqrt{t}x + \sqrt{1-t}Z)] \qquad (x \in \mathbb{R}^d)$$

で定める. 混乱の恐れが少ない場合, Σ を省略して $S_t^\Sigma h = S_t h$ と書く.

注意 9.4. Slepian 補間により定まる作用素の族 $(S_t)_{t\in[0,1)}$ は, 本質的には Σ を共分散行列とする Ornstein-Uhlenbeck 半群と同じものである. 実際, 任意の $t \in [0,\infty)$ に対して,

$$S_{e^{-2t}} h(x) = \mathrm{E}[h(e^{-t}x + \sqrt{1-e^{-2t}}Z)]$$

であるから, $(S_{e^{-2t}})_{t\geq 0}$ は Σ を共分散行列とする Ornstein-Uhlenbeck 半群である. Slepian 補間を使う方が記号が軽くなるため, この講義ノートではこちらを用いることにする.

命題 9.3. $h : \mathbb{R}^d \to \mathbb{R}^k$ を可測関数とし, 任意の $t \in [0,1]$ と $x \in \mathbb{R}^d$ に対して $\mathrm{E}[|h(\sqrt{t}x + \sqrt{1-t}Z)|] < \infty$ を満たすようなものとする. このとき, 任意の $s,t \in [0,1]$ に対して, $S_{st}h = S_s(S_t h)$ が成り立つ.

演習問題 4. 命題 9.3 を証明せよ.

定義 9.2 (Ornstein–Uhlenbeck 作用素)**.** 関数 $f : \mathbb{R}^d \to \mathbb{R}$ が 2 回微分可能なとき, 関数 $L_\Sigma f : \mathbb{R}^d \to \mathbb{R}$ を

$$L_\Sigma f(x) = \langle \Sigma, \mathrm{Hess}\, f(x) \rangle - x \cdot \nabla f(x) \qquad (x \in \mathbb{R}^d)$$

で定める. 混乱の恐れが少ない場合, Σ を省略して $L_\Sigma f = Lf$ と書く.

命題 9.4. $h \in C^2_{b*}(\mathbb{R}^d)$ ならば, 任意の $t \in (0,1)$ と $x \in \mathbb{R}^d$ に対して,

$$\frac{\partial}{\partial t} S_t h(x) = -\frac{1}{2t} L S_t h(x)$$

が成り立つ.

証明. 仮定より期待値と微分の順序交換が可能であり,

$$\frac{\partial}{\partial t} S_t h(x) = \frac{1}{2}\mathrm{E}\left[\left(\frac{x}{\sqrt{t}} - \frac{Z}{\sqrt{1-t}}\right) \cdot \nabla h(\sqrt{t}x + \sqrt{1-t}Z)\right], \tag{9.22}$$

$$\nabla S_t h(x) = \sqrt{t}\,\mathrm{E}\left[\nabla h(\sqrt{t}x + \sqrt{1-t}Z)\right], \tag{9.23}$$

$$\operatorname{Hess} S_t h(x) = t\,\mathrm{E}\left[\operatorname{Hess} h(\sqrt{t}x + \sqrt{1-t}Z)\right] \tag{9.24}$$

が成り立つ. 第 1 式と第 2 式より,

$$\frac{\partial}{\partial t} S_t h(x) = \frac{1}{2t}\left(x \cdot \nabla S_t h(x) - \frac{t}{\sqrt{1-t}}\,\mathrm{E}\left[Z \cdot \nabla h(\sqrt{t}x + \sqrt{1-t}Z)\right]\right)$$

を得る. さらに, Stein の等式と第 3 式より

$$\begin{aligned}\frac{t}{\sqrt{1-t}}\,\mathrm{E}\left[Z \cdot \nabla h(\sqrt{t}x + \sqrt{1-t}Z)\right] &= t\,\mathrm{E}[\langle \Sigma, \operatorname{Hess} h(\sqrt{t}x + \sqrt{1-t}Z)\rangle] \\ &= \langle \Sigma, \operatorname{Hess} S_t h(x)\rangle\end{aligned}$$

が成り立つ. 以上より結論が従う. □

9.4.5 Stein 方程式

定義 9.3 (Stein 方程式)**.** $h : \mathbb{R}^d \to \mathbb{R}$ を可測関数で $\mathrm{E}[|h(Z)|] < \infty$ を満たすものとする. f に関する偏微分方程式

$$L_\Sigma f(x) = h(x) - \mathrm{E}[h(Z)] \tag{9.25}$$

を h に対する共分散行列 Σ の **Stein 方程式 (Stein's equation)** と呼ぶ. 関数 $f : \mathbb{R}^d \to \mathbb{R}$ が方程式 (9.25) の解であるとは, f が 2 回微分可能であり, かつすべての $x \in \mathbb{R}^d$ について (9.25) が成り立つことをいう.

定義 9.3 の設定において, もし方程式 (9.25) の解が存在するならば, $\mathrm{E}[|h(W)|] < \infty$ なる任意の d 次元確率変数 W に対して

$$\mathrm{E}[h(W)] - \mathrm{E}[h(Z)] = \mathrm{E}[L_\Sigma f(W)]$$

が成り立つことになる. 従って, $\mathrm{E}[h(W)] - \mathrm{E}[h(Z)] \approx 0$ を示すには $\mathrm{E}[L_\Sigma f(W)] \approx 0$ を示せばよいことになる.

Stein 方程式 (9.25) の解の存在を保証するための十分条件として, 以下の結果を挙げておく.

命題 9.5. r を 2 以上の整数とする. $h \in C^r_{b*}(\mathbb{R}^d)$ ならば, 関数 $f : \mathbb{R}^d \to \mathbb{R}$ を

$$f(x) = -\int_0^1 \frac{1}{2t}\{S_t h(x) - \mathrm{E}[h(Z)]\}dt \qquad (x \in \mathbb{R}^d) \tag{9.26}$$

で定義することができる. さらに, $f \in C^r_{b*}(\mathbb{R}^d)$ であり, Stein 方程式 (9.25) の解となる.

証明. まず, 仮定より任意の $x, y \in \mathbb{R}^d$ に対して $|h(x) - h(y)| \le \|\nabla h\|_\infty |x - y|$ が成り立つから, 任意の $x \in \mathbb{R}^d$ と $t \in [0, 1]$ に対して,

$$\begin{aligned}|S_t h(x) - \mathrm{E}[h(Z)]| &\le \mathrm{E}[|h(\sqrt{t}x + \sqrt{1-t}Z) - h(Z)|] \\ &\le \|\nabla h\|_\infty\,\mathrm{E}[|\sqrt{t}x + (\sqrt{1-t} - 1)Z|] \\ &\le \|\nabla h\|_\infty(\sqrt{t}|x| + (1 - \sqrt{1-t})\,\mathrm{E}[|Z|])\end{aligned}$$

が成り立つ. $1-\sqrt{1-t}=\sqrt{1-t+t}-\sqrt{1-t}\le\sqrt{t}$ であるから,

$$\begin{aligned}\int_0^1 \frac{1}{t}|S_t h(x)-\mathrm{E}[h(Z)]|dt &\le \|\nabla h\|_\infty(|x|+\mathrm{E}[|Z|])\int_0^1\frac{1}{\sqrt{t}}dt\\ &=2\|\nabla h\|_\infty(|x|+\mathrm{E}[|Z|])<\infty\end{aligned}$$

を得る. 従って $f(x)$ は well-defined である.

次に, $f\in C^r_{b^*}(\mathbb{R}^d)$ であり, Stein 方程式 (9.25) の解となることを示す. 各 $k\in\{1,\dots,r\}$ と $t\in[0,1]$ について, 任意の $j_1,\dots,j_k\in\{1,\dots,d\}$ に対して

$$|\partial_{j_1,\dots,j_k}S_t h(x)|=|t^{k/2}\,\mathrm{E}[\partial_{j_1,\dots,j_k}h(\sqrt{t}x+\sqrt{1-t}Z)]|\le t^{k/2}\|\partial_{j_1,\dots,j_k}h\|_\infty$$

が成り立つから, f は r 回微分可能であり,

$$\partial_{j_1,\dots,j_k}f(x)=-\int_0^1\frac{1}{2t}\partial_{j_1,\dots,j_k}S_t h(x)dt$$

が成り立つ. 特に, 有界収束定理より $\partial_{j_1,\dots,j_k}f$ が連続であることがわかり, かつ

$$\begin{aligned}|\partial_{j_1,\dots,j_k}f(x)| &\le \|\partial_{j_1,\dots,j_k}h\|_\infty\int_0^1\frac{t^{k/2-1}}{2}dt\\ &=\frac{\|\partial_{j_1,\dots,j_k}h\|_\infty}{2}\left[\frac{t^{k/2}}{k/2}\right]_0^1=\frac{\|\partial_{j_1,\dots,j_k}h\|_\infty}{k}<\infty\end{aligned}$$

が成り立つから, $f\in C^r_{b^*}(\mathbb{R}^d)$ である. さらに, 命題 9.4 および微分積分学の基本定理より,

$$Lf(x)=-\int_0^1\frac{1}{2t}LS_t h(x)dt=\int_0^1\frac{\partial}{\partial t}S_t h(x)dt=S_1h(x)-S_0h(x)=h(x)-\mathrm{E}[h(Z)]$$

が成り立つ. 故に, f は Stein 方程式 (9.25) の解である. □

9.4.6 高次元正規分布の比較

これまでの準備を元に, まずは CCK 理論のプロトタイプとして以下の結果を証明する. この結果は後でブートストラップ法の正当性を証明する際にも有用である.

定理 9.7 (Chernozhukov et al. (2022), Proposition 2.1)**.** Σ,Σ' を d 次半正定値対称行列とし, $Z\sim\mathrm{N}(0,\Sigma)$, $Z'\sim\mathrm{N}(0,\Sigma')$ とする. このとき, $\underline{\sigma}:=\min_{j=1,\dots,d}\sqrt{\mathrm{E}[Z_j^2]}>0$ ならば, ある普遍定数 $C>0$ が存在して

$$\sup_{x\in\mathbb{R}^d}|\mathrm{P}(Z'\le x)-\mathrm{P}(Z\le x)|\le C\sqrt{\frac{|\Sigma'-\Sigma|_\infty(\log d)^2}{\underline{\sigma}^2}}$$

が成り立つ.

証明. Z と Z' は独立であると仮定して一般性を失わない. 関数 g_0 を補題 9.7 のように定める. $\varepsilon > 0$ を任意に 1 つ取り, $\beta := \varepsilon^{-1}\log d$ とおく. 命題 9.2 より

$$\sup_{x\in\mathbb{R}^d} |\mathrm{P}(Z' \le x) - \mathrm{P}(Z \le x)| \le \Delta_\varepsilon(Z', Z) + \frac{2\varepsilon}{\underline{\sigma}}\left(\sqrt{2\log d} + 2\right)$$

が成り立つ. ただし,

$$\Delta_\varepsilon(Z', Z) := \sup_{y\in\mathbb{R}^d} \left|\mathrm{E}[g_0(\varepsilon^{-1}F_\beta(Z'-y))] - \mathrm{E}[g_0(\varepsilon^{-1}F_\beta(Z-y))]\right|$$

である.

$\Delta_\varepsilon(Z', Z)$ を評価するために, 任意に $y \in \mathbb{R}^d$ を 1 つ固定し, 関数 $h : \mathbb{R}^d \to [0,1]$ を $h(w) = g_0(\varepsilon^{-1}F_\beta(w-y))$ $(w \in \mathbb{R}^d)$ で定める. 補題 9.6 より $h \in C_b^2(\mathbb{R}^d)$ であるから, 命題 9.5 より関数 $f : \mathbb{R}^d \to \mathbb{R}$ を (9.26) 式で定めることができて, f は Stein 方程式 (9.25) の解となる. (9.23)–(9.24) より, 任意の $x \in \mathbb{R}^d$ に対して

$$L_\Sigma f(x) = \frac{1}{2}\int_0^1 \left\{\frac{x\cdot \mathrm{E}\left[\nabla h(\sqrt{t}x + \sqrt{1-t}Z)\right]}{\sqrt{t}} - \left\langle \Sigma, \mathrm{E}\left[\operatorname{Hess} h(\sqrt{t}x + \sqrt{1-t}Z)\right]\right\rangle\right\} dt$$

が成り立つ. Z と Z' が独立であることに注意すると, この式から

$$\mathrm{E}[L_\Sigma f(Z')] = \frac{1}{2}\int_0^1 \left\{\frac{\mathrm{E}\left[Z'\cdot \nabla h(\sqrt{t}Z' + \sqrt{1-t}Z)\right]}{\sqrt{t}} - \left\langle \Sigma, \mathrm{E}\left[\operatorname{Hess} h(\sqrt{t}Z' + \sqrt{1-t}Z)\right]\right\rangle\right\} dt$$

を得る. ここで, Stein の等式より

$$\mathrm{E}\left[Z'\cdot \nabla h(\sqrt{t}Z' + \sqrt{1-t}Z)\right] = \sqrt{t}\,\mathrm{E}\left[\langle \Sigma', \operatorname{Hess} h(\sqrt{t}Z' + \sqrt{1-t}Z)\rangle\right]$$

が成り立つから, 代入して

$$\mathrm{E}[L_\Sigma f(Z')] = \frac{1}{2}\int_0^1 \left\langle \Sigma' - \Sigma, \mathrm{E}\left[\operatorname{Hess} h(\sqrt{t}Z' + \sqrt{1-t}Z)\right]\right\rangle dt$$

を得る. ここで, $x \notin [0,1]$ のとき $g_0'(x) = 0$ であるから, $w \in \mathbb{R}^d$ が $\nabla h(w) \neq 0$ を満たすならば $0 \le F_\beta(w-y) \le \varepsilon$ とならなければならない. 従って, 各 $t \in [0,1]$ について,

$$\begin{aligned}
&\operatorname{Hess} h(\sqrt{t}Z' + \sqrt{1-t}Z) \neq O\\
&\Rightarrow 0 \le F_\beta(\sqrt{t}Z' + \sqrt{1-t}Z - y) \le \varepsilon\\
&\Rightarrow -\varepsilon \le \max_{1\le j\le d}(\sqrt{t}Z_j' + \sqrt{1-t}Z_j - y_j) \le \varepsilon \quad (\because \text{補題 } 9.3)
\end{aligned} \tag{9.27}$$

が成り立つ. 故に, $A(t) := \{-\varepsilon \le \max_{1\le j\le d}(\sqrt{t}Z_j' + \sqrt{1-t}Z_j - y_j) \le \varepsilon\}$ とおけば,

$$\mathrm{E}[L_\Sigma f(Z')] = \frac{1}{2}\int_0^1 \left\langle \Sigma' - \Sigma, \mathrm{E}\left[\operatorname{Hess} h(\sqrt{t}Z' + \sqrt{1-t}Z)1_{A(t)}\right]\right\rangle dt$$

となるので,

$$|\operatorname{E}[L_\Sigma f(Z')]| \le \frac{1}{2}\int_0^1 |\Sigma' - \Sigma|_\infty \operatorname{E}\left[\sum_{j,k=1}^d |\partial_{jk} h(\sqrt{t}Z' + \sqrt{1-t}Z)| 1_{A(t)}\right] dt$$

が成り立つ. ここで, 補題 9.6 の性質 (i)–(ii) より, 任意の $w \in \mathbb{R}^d$ に対して

$$\sum_{j,k=1}^d |\partial_{jk} h(w)| \lesssim \varepsilon^{-2} \log d$$

が成り立つ. 従って,

$$|\operatorname{E}[L_\Sigma f(Z')]| \lesssim \varepsilon^{-2} (\log d) |\Sigma' - \Sigma|_\infty \int_0^1 \operatorname{P}(A(t)) dt \tag{9.28}$$

を得る. ここで, Z と Z' が独立であることに注意すると,

$$\begin{aligned}
\operatorname{P}(A(t)) &\le \sup_{z \in \mathbb{R}^d} \operatorname{P}\left(-\varepsilon \le \max_{1\le j\le d}(\sqrt{1-t}Z_j - z_j) \le \varepsilon\right) \\
&= \sup_{z \in \mathbb{R}^d} \left\{\operatorname{P}\left(\sqrt{1-t}Z \le z + \varepsilon\right) - \operatorname{P}\left(\sqrt{1-t}Z \le z - \varepsilon\right)\right\}
\end{aligned}$$

が成り立つから, Nazarov の不等式より

$$\operatorname{P}(A(t)) \lesssim \frac{\varepsilon\sqrt{\log d}}{\underline{\sigma}\sqrt{1-t}}$$

を得る. よって,

$$\int_0^1 \operatorname{P}(A(t)) dt \lesssim \frac{\varepsilon}{\underline{\sigma}} \int_0^1 \frac{1}{\sqrt{1-t}} dt \lesssim \frac{\varepsilon\sqrt{\log d}}{\underline{\sigma}} \tag{9.29}$$

が成り立つ. (9.28), (9.29) と f が Stein 方程式 (9.25) の解であることから,

$$|\operatorname{E}[h(Z')] - \operatorname{E}[h(Z)]| \lesssim \varepsilon^{-1} \frac{|\Sigma' - \Sigma|_\infty (\log d)^{3/2}}{\underline{\sigma}}$$

を得る. 右辺は y に依存しないから,

$$\Delta_\varepsilon(Z', Z) \lesssim \varepsilon^{-1} \frac{|\Sigma' - \Sigma|_\infty (\log d)^{3/2}}{\underline{\sigma}}$$

である. 故に,

$$\sup_{x \in \mathbb{R}^d} |\operatorname{P}(Z' \le x) - \operatorname{P}(Z \le x)| \lesssim \varepsilon^{-1} \frac{|\Sigma' - \Sigma|_\infty (\log d)^{3/2}}{\underline{\sigma}} + \frac{\varepsilon\sqrt{\log d}}{\underline{\sigma}}$$

が成り立つ. 従って,

$$\varepsilon = \sqrt{|\Sigma' - \Sigma|_\infty (\log d)}$$

ととれば示すべき不等式が得られる. □

9.4.7 Exchangeable pair

定理 9.1 の証明には Stein の方法における標準的技法の 1 つである **exchangeable pair approach** を用いる. まずは exchangeable pair の定義から始める. 以下この小節では, Ξ は可分位相空間を表すとする.

定義 9.4 (Exchangeable pair)**.** Ξ 値確率変数の組 (W, W') が **exchangeable pair** であるとは, (W, W') の確率分布と (W', W) の確率分布が (積空間 $\Xi \times \Xi$ に値をとる確率変数として) 一致することをいう.

注意 9.5. (W, W') が exchangeable pair ならば, W と W' は同分布となる.

例 9.1. W と W' を同じ分布に従う独立な確率変数とすると, (W, W') は明らかに exchangeable pair である.

例 9.2. (W, W') を Ξ 値確率変数の exchangeable pair, $\tilde{\Xi}$ を可分距離空間とする. このとき, 任意の可測関数 $g : \Xi \to \tilde{\Xi}$ に対して, $(g(W), g(W'))$ も明らかに exchangeable pair となる.

命題 9.6. $\xi_1, \dots, \xi_n$ を独立な Ξ 値確率変数列とし, Ξ^n 値確率変数 $\xi = (\xi_1, \dots, \xi_n)$ を考える. $\xi^* = (\xi_1^*, \dots, \xi_n^*)$ を ξ の独立なコピー, すなわち ξ^* は ξ と同じ分布をもち, かつ ξ と ξ^* は独立であるようなものとする. I を $\{1, \dots, n\}$ 上の一様分布に従う確率変数で, ξ, ξ^* と独立であるようなものとする. このとき, Ξ^n 値確率変数 $\xi' = (\xi_1', \dots, \xi_n')$ を

$$\xi_i' = \begin{cases} \xi_i^* & \text{if } I = i, \\ \xi_i & \text{otherwise}, \end{cases} \qquad i = 1, \dots, n,$$

で定めると, (ξ, ξ') は exchangeable pair となる. 従って, 任意の可分距離空間 $\tilde{\Xi}$ と可測関数 $g : \Xi^n \to \tilde{\Xi}$ に対して, $(g(\xi), g(\xi'))$ も exchangeable pair である.

演習問題 5. 命題 9.6 を証明せよ.

Exchangeable pair approach に基づく高次元中心極限定理として次の評価が導かれる.

定理 9.8. (W, W') を d 次元確率変数の exchangeable pair で $\mathrm{E}[|W' - W|^4] < \infty$ を満たすものとする. また, ある d 次正則行列 Λ と σ-加法族 $\mathcal{G}$, および $\mathcal{G}$-可測な d 次元確率変数 R が存在して, W は $\mathcal{G}$-可測であり, かつ

$$\mathrm{E}[W' - W \mid \mathcal{G}] = -\Lambda(W + R) \tag{9.30}$$

を満たすとする. このとき, ある普遍定数 $C > 0$ が存在して, 任意の $\varepsilon > 0$ に対して,

$$\sup_{x \in \mathbb{R}^d} |\mathrm{P}(W \le x) - \mathrm{P}(Z \le x)|$$
$$\le \frac{C}{\underline{\sigma}} \Bigg(\sqrt{\log d}\, \mathrm{E}\left[|\mathrm{E}[\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty + |R|_\infty \right] + \varepsilon^{-1} (\log d)^{3/2}\, \mathrm{E}\left[|V(\beta^{-1})|_\infty \right]$$

$$+\,\varepsilon^{-3}(\log d)^{7/2}\,\mathrm{E}\left[\max_{1\le j,k,l,m\le d}\mathrm{E}[|(\Lambda^{-1}D)_jD_kD_lD_m|1_{\{|D|_\infty\le\beta^{-1}\}}\mid\mathcal{G}]\right]+\varepsilon\sqrt{\log d}\Bigg)$$

が成り立つ. ただし, $D := W' - W$, $\beta = \varepsilon^{-1}\log d$ とし, $\tau \in (0,\infty]$ に対して

$$V(\tau) := \Sigma - \frac{1}{2}\,\mathrm{E}[\Lambda^{-1}DD^\top 1_{\{|D|_\infty\le\tau\}}\mid\mathcal{G}]$$

と定める.

定理 9.8 を証明するために 2 つ補題を準備する. 1 つ目は積分表現を見やすくするためのもので, 本質的ではない.

補題 9.10 (Taylor の定理の確率論的表現)**.** k を正整数とする. 関数 $f:\mathbb{R}^d\to\mathbb{R}$ が C^k 級ならば, 任意の $x,h\in\mathbb{R}^d$ に対して

$$f(x+h)-f(x)=\sum_{l=1}^{k-1}\frac{\langle h^{\otimes l},\nabla^l f(x)\rangle}{l!}+\frac{1}{(k-1)!}\,\mathrm{E}[(1-U)^{k-1}\langle h^{\otimes k},\nabla^k f(x+Uh)\rangle]$$

が成り立つ. ここに, U は $[0,1]$ 上の一様分布に従う確率変数である. ただし, $k=1$ の場合は右辺第 1 項を 0 とみなす.

演習問題 6. 補題 9.10 を証明せよ.

次の補題は exchangeable pair approach の肝となる重要な結果である.

補題 9.11 (Symmetry trick)**.** 定理 9.8 の仮定の下, 任意の $f\in C^4_{b^*}(\mathbb{R}^d)$ と $\tau\in(0,\infty]$ に対して,

$$\begin{aligned}\mathrm{E}[Lf(W)] &= \mathrm{E}[(R-\Lambda^{-1}D1_{\{|D|_\infty>\tau\}})\cdot\nabla f(W)]+\mathrm{E}[\langle V(\tau),\mathrm{Hess}\,f(W)\rangle]\\ &\quad+\frac{1}{4}\,\mathrm{E}[(1-U)\tilde{U}\langle\Lambda^{-1}D\otimes D^{\otimes 3},\nabla^4 f(W+(U+U'\tilde{U})D)\rangle 1_{\{|D|_\infty\le\tau\}}]\end{aligned}$$

が成り立つ. ここに, $\tilde{U}:=1-2U$ であり, U,U' は $[0,1]$ 上の一様分布に従う独立な確率変数で (W,W') と独立であるようなものとする.

証明. (W,W') が exchangeable pair であることから,

$$\mathrm{E}[\Lambda^{-1}D\cdot\{\nabla f(W)+\nabla f(W')\}1_{\{|D|_\infty\le\tau\}}]=-\,\mathrm{E}[\Lambda^{-1}D\cdot\{\nabla f(W')+\nabla f(W)\}1_{\{|D|_\infty\le\tau\}}]$$

が成り立つ. 従って,

$$\mathrm{E}[\Lambda^{-1}D\cdot\{\nabla f(W)+\nabla f(W')\}1_{\{|D|_\infty\le\tau\}}]=0 \tag{9.31}$$

を得る. 一方で, 補題 9.10 より

$$\begin{aligned}&\mathrm{E}[\Lambda^{-1}D\cdot\{\nabla f(W')-\nabla f(W)\}1_{\{|D|_\infty\le\tau\}}]=\sum_{j=1}^d\mathrm{E}[(\Lambda^{-1}D)_j\{\partial_j f(W')-\partial_j f(W)\}1_{\{|D|_\infty\le\tau\}}]\\ &=\sum_{j,k=1}^d\mathrm{E}[(\Lambda^{-1}D)_jD_k\partial_{jk}f(W)1_{\{|D|_\infty\le\tau\}}]+\sum_{j,k,l=1}^d\mathrm{E}[(1-U)(\Lambda^{-1}D)_jD_kD_l\partial_{jkl}f(W+UD)1_{\{|D|_\infty\le\tau\}}]\end{aligned}$$

$$= \mathrm{E}[\langle \mathrm{E}[\Lambda^{-1} D D^\top 1_{\{|D|_\infty \le \tau\}} \mid \mathcal{G}], \mathrm{Hess}\, f(W)\rangle] + E$$

が成り立つ. ただし,

$$\begin{aligned} E &:= \mathrm{E}[(1-U)\langle \Lambda^{-1} D \otimes D^{\otimes 2}, \nabla^3 f(W+UD)\rangle 1_{\{|D|_\infty \le \tau\}}] \\ &= \sum_{j,k,l=1}^d \mathrm{E}[(1-U)(\Lambda^{-1}D)_j D_k D_l \partial_{jkl} f(W+UD) 1_{\{|D|_\infty \le \tau\}}] \end{aligned}$$

と定める. この式と条件 (9.30) から,

$$\begin{aligned} &\mathrm{E}[\Lambda^{-1} D \cdot \{\nabla f(W) + \nabla f(W')\} 1_{\{|D|_\infty \le \tau\}}] \\ &= 2\,\mathrm{E}[\Lambda^{-1} D \cdot \nabla f(W) 1_{\{|D| \le \tau\}}] + \mathrm{E}[\Lambda^{-1} D \cdot \{\nabla f(W') - \nabla f(W)\} 1_{\{|D|_\infty \le \tau\}}] \\ &= 2\,\mathrm{E}[\Lambda^{-1} D \cdot \nabla f(W) 1_{\{|D|_\infty > \tau\}}] - 2\,\mathrm{E}[(W+R)\cdot \nabla f(W)] \\ &\quad + \mathrm{E}[\langle \mathrm{E}[\Lambda^{-1} D D^\top 1_{\{|D|_\infty \le \tau\}} \mid \mathcal{G}], \mathrm{Hess}\, f(W)\rangle] + E \end{aligned} \tag{9.32}$$

を得る. (9.31)–(9.32) より

$$\begin{aligned} \mathrm{E}[W \cdot \nabla f(W)] &= \mathrm{E}[(\Lambda^{-1} D 1_{\{|D|_\infty > \tau\}} - R) \cdot \nabla f(W)] \\ &\quad + \frac{1}{2}\left(\mathrm{E}[\langle \mathrm{E}[\Lambda^{-1} D D^\top 1_{\{|D|_\infty \le \tau\}} \mid \mathcal{G}], \mathrm{Hess}\, f(W)\rangle] + E\right) \end{aligned}$$

が成り立つから,

$$\begin{aligned} \mathrm{E}[Lf(W)] &= \mathrm{E}[\langle \Sigma, \mathrm{Hess}\, f(W)\rangle] - \mathrm{E}[W \cdot \nabla f(W)] \\ &= \mathrm{E}[(R - \Lambda^{-1} D 1_{\{|D|_\infty > \tau\}}) \cdot \nabla f(W)] + \mathrm{E}[\langle V(\tau), \mathrm{Hess}\, f(W)\rangle] - \frac{1}{2} E \end{aligned} \tag{9.33}$$

を得る. ここで, (W, W') が exchangeable pair であることを再び用いると,

$$\begin{aligned} E &= \mathrm{E}[(1-U)\langle \Lambda^{-1}(-D) \otimes (-D)^{\otimes 2}, \nabla^3 f(W' - UD)\rangle 1_{\{|D|_\infty \le \tau\}}] \\ &= -\,\mathrm{E}[(1-U)\langle \Lambda^{-1} D \otimes D^{\otimes 2}, \nabla^3 f(W + (1-U)D)\rangle 1_{\{|D|_\infty \le \tau\}}] \end{aligned} \tag{9.34}$$

を得る. 従って, 補題 9.10 より

$$\begin{aligned} E &= \frac{1}{2}\mathrm{E}[(1-U)\langle \Lambda^{-1} D \otimes D^{\otimes 2}, \nabla^3 f(W+UD) - \nabla^3 f(W+(1-U)D)\rangle 1_{\{|D|_\infty \le \tau\}}] \\ &= -\frac{1}{2}\mathrm{E}[(1-U)\langle \Lambda^{-1} D \otimes D^{\otimes 2}, \nabla^3 f(W+UD+(1-2U)D) - \nabla^3 f(W+UD)\rangle 1_{\{|D|_\infty \le \tau\}}] \\ &= -\frac{1}{2}\mathrm{E}[(1-U)\tilde{U}\langle \Lambda^{-1} D \otimes D^{\otimes 3}, \nabla^4 f(W+(U+U'\tilde{U})D)\rangle 1_{\{|D|_\infty \le \tau\}}] \end{aligned}$$

が成り立つ. この式を (9.33) に代入して示すべき等式を得る. □

注意 9.6. (a) 補題 9.11 の等式が D の 3 次の項を含まず 4 次の項を含むことは, 高次元正規近似に対するよい評価を得る上で非常に重要な役割を果たす. このような等式は exchangeable pair に内蔵されている対称性を用いた非常に単純な議論によって (9.34) 式を導くことで得られるが, この事実はごく最近になって Fang & Koike (2021) で指摘されるまで (なぜか) 知られていなかった. Fang & Koike (2022) は同様の議論によって Wasserstein 距離に対する正規近似の評価を改善している.

(b) 補題 9.6 の性質 (iii) を有効活用する上で, 刈り込みの議論 (truncation argument) が重要な役割を果たす. 標準的な exchangeable pair approach の議論では, 等式 (9.31) において $\tau = \infty$ としたものを考えるが, 補題 9.11 ではこの時点で刈り込みを導入することで, 刈り込みの議論を効率的に証明に組み込めるようにしている. このアイディアは本質的には Bonis (2020) によるものである. なお, 定理 9.1 を証明するだけであれば, 刈り込みの議論を別の段階で行っても問題ないため, $\tau = \infty$ の場合の結果で十分である.

定理 9.8 の証明. $\mathcal{G}$ と Z は独立であると仮定して一般性を失わない. 4 ステップに分けて証明する.
Step 1. 関数 g_0 を補題 9.7 のように定める. 命題 9.2 より

$$\sup_{x \in \mathbb{R}^d} |\mathrm{P}(W \le x) - \mathrm{P}(Z \le x)| \le \Delta_\varepsilon(W, Z) + \frac{2\varepsilon}{\underline{\sigma}} \left(\sqrt{2 \log d} + 2 \right) \tag{9.35}$$

が成り立つ. $\Delta_\varepsilon(W, Z)$ を評価するために, 任意に $y \in \mathbb{R}^d$ を 1 つ固定し, 関数 $h : \mathbb{R}^d \to [0, 1]$ を $h(w) = g_0(\varepsilon^{-1} F_\beta(w - y))$ $(w \in \mathbb{R}^d)$ で定める. 補題 9.6 より $h \in C_b^2(\mathbb{R}^d)$ であるから, 命題 9.5 より関数 $f : \mathbb{R}^d \to \mathbb{R}$ を (9.26) 式で定めることができて, f は Stein 方程式 (9.25) の解となる.

(9.23)–(9.24) に注意すると, 微分と積分の順序交換より

$$\mathrm{E}[Lf(W)] = -\int_0^1 \frac{1}{2t} \mathrm{E}[LS_t h(W)] dt \tag{9.36}$$

が成り立つ. 各 $t \in [0, 1]$ について, 補題 9.6 より $h \in C_b^4(\mathbb{R}^d)$ であるから, 補題 9.11 を $\tau = \beta^{-1}$ として適用すると,

$$\begin{aligned} \mathrm{E}[LS_t h(W)] &= \mathrm{E}[(\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} - R) \cdot \nabla S_t h(W)] + \mathrm{E}[\langle V(\beta^{-1}), \mathrm{Hess}\, S_t h(W) \rangle] \\ &\quad + \frac{1}{4} \mathrm{E}[(1 - U)\tilde{U} \langle \Lambda^{-1} D \otimes D^{\otimes 3}, \nabla^4 S_t h(W + (U + U'\tilde{U})D) \rangle 1_{\{|D|_\infty \le \beta^{-1}\}}] \end{aligned}$$

が成り立つ. ここに, $\tilde{U} := 1 - 2U$ であり, U, U' は $[0, 1]$ 上の一様分布に従う独立な確率変数で (W, W') と独立であるようなものとする. W と Z が独立であることに注意すると, $W(t) := \sqrt{t} W + \sqrt{1 - t} Z$ と定めれば, 上の式は

$$\begin{aligned} \mathrm{E}[LS_t h(W)] &= \sqrt{t}\, \mathrm{E}[(\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} - R) \cdot \nabla h(W(t))] + t\, \mathrm{E}[\langle V(\beta^{-1}), \mathrm{Hess}\, h(W(t)) \rangle] \\ &\quad + \frac{t^2}{4} \mathrm{E}[(1 - U)\tilde{U} \langle \Lambda^{-1} D \otimes D^{\otimes 3}, \nabla^4 h(W(t) + \sqrt{t}(U + U'\tilde{U})D) \rangle 1_{\{|D|_\infty \le \beta^{-1}\}}] \\ &=: \sqrt{t} I + t II + \frac{t^2}{4} III \end{aligned} \tag{9.37}$$

と書き直せる.
Step 2. I を評価する. (9.27) と同様の議論によって, $A(t) := \{-\varepsilon \le \max_{1 \le j \le d}(\sqrt{t} W_j + \sqrt{1 - t} Z_j - y_j) \le \varepsilon\}$ とおくと $\nabla h(W(t)) = \nabla h(W(t)) 1_{A(t)}$ が成り立つことを示せるから,

$$I = \mathrm{E}[(\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} - R) \cdot \nabla h(W(t)) 1_{A(t)}]$$

と書き直せる. 従って,

$$|I| \le \mathrm{E}\left[\left(|\,\mathrm{E}[\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty + |R|_\infty\right) 1_{A(t)} \sum_{j=1}^d |\partial_j h(W(t))|\right]$$

が成り立つ. 故に, 補題 9.4 より

$$|I| \lesssim \varepsilon^{-1}\,\mathrm{E}\left[\left(|\,\mathrm{E}[\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty + |R|_\infty\right) 1_{A(t)}\right]$$

を得る. ここで, $\mathcal{G}$ と Z が独立であることに注意すると,

$$\begin{aligned}
&\mathrm{E}\left[\left(|\,\mathrm{E}[\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty + |R|_\infty\right) 1_{A(t)}\right] \\
&\le \mathrm{E}\left[|\,\mathrm{E}[\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty + |R|_\infty\right] \sup_{z\in\mathbb{R}^d} \mathrm{P}\left(-\varepsilon \le \max_{1\le j\le d}(\sqrt{1-t} Z_j - z_j) \le \varepsilon\right)
\end{aligned}$$

が成り立つ. Nazarov の不等式より

$$\sup_{z\in\mathbb{R}^d} \mathrm{P}\left(-\varepsilon \le \max_{1\le j\le d}(\sqrt{1-t} Z_j - z_j) \le \varepsilon\right) \lesssim \frac{\varepsilon\sqrt{\log d}}{\underline{\sigma}\sqrt{1-t}}$$

が成り立つので,

$$|I| \lesssim \frac{\sqrt{\log d}}{\underline{\sigma}\sqrt{1-t}}\,\mathrm{E}\left[|\,\mathrm{E}[\Lambda^{-1} D 1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty + |R|_\infty\right] \tag{9.38}$$

を得る.

Step 3. II を評価する. (9.27) と同様の議論によって,

$$II = \mathrm{E}[\langle V(\beta^{-1}), \mathrm{Hess}\, h(W(t))\rangle 1_{A(t)}]$$

と書き直せるから,

$$|II| \le \mathrm{E}\left[|V(\beta^{-1})|_\infty 1_{A(t)} \sum_{j,k=1}^d |\partial_{jk} h(W(t))|\right]$$

が成り立つ. 従って, 補題 9.6 の性質 (i)–(ii) より,

$$|II| \lesssim \varepsilon^{-2}(\log d)\,\mathrm{E}\left[|V(\beta^{-1})|_\infty 1_{A(t)}\right]$$

を得る. ここで, $\mathcal{G}$ と Z が独立であることに注意すると,

$$\mathrm{E}\left[|V(\beta^{-1})|_\infty 1_{A(t)}\right] \le \mathrm{E}\left[|V(\beta^{-1})|_\infty\right] \sup_{z\in\mathbb{R}^d} \mathrm{P}\left(-\varepsilon \le \max_{1\le j\le d}(\sqrt{1-t} Z_j - z_j) \le \varepsilon\right)$$

が成り立つ. 以上より,

$$|II| \lesssim \varepsilon^{-1} \frac{(\log d)^{3/2}}{\underline{\sigma}\sqrt{1-t}}\,\mathrm{E}\left[|V(\beta^{-1})|_\infty\right] \tag{9.39}$$

を得る.

Step 4. III を評価する. まず, $|U+U'\tilde{U}| \le U \vee (U+\tilde{U}) = U \vee (1-U) \le 1$ に注意する. いま, (9.27) と同様の議論によって,

$$\nabla^4 h(W(t)+\sqrt{t}(U+U'\tilde{U})D) \neq 0 \Rightarrow -\varepsilon \le \max_{1\le j\le d}(W_j(t)+\sqrt{t}(U+U'\tilde{U})D_j - y_j) \le \varepsilon$$

が成り立つから, 事象 $\{|D|_\infty \le \beta^{-1}\}$ 上では

$$\nabla^4 h(W(t)+\sqrt{t}(U+U'\tilde{U})D) \neq 0 \Rightarrow -\varepsilon-\beta^{-1} \le \max_{1\le j\le d}(W_j(t)-y_j) \le \varepsilon+\beta^{-1}$$

が成り立つ. 従って, $A'(t) := \{-\varepsilon-\beta^{-1} \le \max_{1\le j\le d}(W_j(t)-y_j) \le \varepsilon+\beta^{-1}\}$ とおくと $\nabla^4 h(W(t)+(U+U'\tilde{U})D)1_{\{|D|_\infty\le\beta^{-1}\}} = \nabla^4 h(W(t)+(U+U'\tilde{U})D)1_{\{|D|_\infty\le\beta^{-1}\}\cap A'(t)}$ が成り立つから,

$$III = \mathrm{E}[(1-U)\tilde{U}\langle \Lambda^{-1}D\otimes D^{\otimes 3}, \nabla^4 h(W(t)+(U+U'\tilde{U})D)\rangle 1_{\{|D|_\infty\le\beta^{-1}\}\cap A'(t)}]$$

が成り立つ. 次に, 補題 9.6 で与えられる $\mathbb{R}^d$ 上の非負値関数の族 $(U_{jklm})_{1\le j,k,l,m\le d}$ を考えると, 補題 9.6 の性質 (i) と (iii) より

$$|III| \lesssim \mathrm{E}\left[\sum_{j,k,l,m=1}^d |(\Lambda^{-1}D)_j D_k D_l D_m| U_{jklm}(W(t)) 1_{\{|D|_\infty\le\beta^{-1}\}\cap A'(t)}\right]$$

を得る. 従って,

$$|III| \lesssim \mathrm{E}\left[\max_{1\le j,k,l,m\le d}\mathrm{E}[|(\Lambda^{-1}D)_j D_k D_l D_m| 1_{\{|D|_\infty\le\beta^{-1}\}} \mid \mathcal{G}] 1_{A'(t)} \sum_{j,k,l,m=1}^d U_{jklm}(W(t))\right]$$

が成り立つ. 故に, 補題 9.6 の性質 (ii) より,

$$|III| \lesssim \varepsilon^{-4}(\log d)^3\,\mathrm{E}\left[\max_{1\le j,k,l,m\le d}\mathrm{E}[|(\Lambda^{-1}D)_j D_k D_l D_m| 1_{\{|D|_\infty\le\beta^{-1}\}} \mid \mathcal{G}] 1_{A'(t)}\right]$$

を得る. $\mathcal{G}$ と Z が独立であることに注意すると,

$$\begin{aligned}
&\mathrm{E}\left[\max_{1\le j,k,l,m\le d}\mathrm{E}[|(\Lambda^{-1}D)_j D_k D_l D_m| 1_{\{|D|_\infty\le\beta^{-1}\}} \mid \mathcal{G}] 1_{A'(t)}\right] \\
&\le \mathrm{E}\left[\max_{1\le j,k,l,m\le d}\mathrm{E}[|(\Lambda^{-1}D)_j D_k D_l D_m| 1_{\{|D|_\infty\le\beta^{-1}\}} \mid \mathcal{G}]\right] \\
&\qquad \times \sup_{z\in\mathbb{R}^d}\mathrm{P}\left(-\varepsilon-\beta^{-1} \le \max_{1\le j\le d}(\sqrt{1-t}Z_j - z_j) \le \varepsilon+\beta^{-1}\right)
\end{aligned}$$

が成り立つ. Nazarov の不等式より

$$\sup_{z\in\mathbb{R}^d}\mathrm{P}\left(-\varepsilon-\beta^{-1} \le \max_{1\le j\le d}(\sqrt{1-t}Z_j - z_j) \le \varepsilon+\beta^{-1}\right) \lesssim \frac{\varepsilon\sqrt{\log d}}{\underline{\sigma}\sqrt{1-t}}$$

が成り立つので,

$$|III| \lesssim \varepsilon^{-3}\frac{(\log d)^{7/2}}{\underline{\sigma}\sqrt{1-t}}\,\mathrm{E}\left[\max_{1\le j,k,l,m\le d}\mathrm{E}[|(\Lambda^{-1}D)_j D_k D_l D_m| 1_{\{|D|_\infty\le\beta^{-1}\}} \mid \mathcal{G}]\right] \tag{9.40}$$

を得る.

(9.36)–(9.40) と f が Stein 方程式 (9.25) の解であることに注意すると,

$$
\begin{aligned}
\Delta_\varepsilon(W,Z) \lesssim \frac{\sqrt{\log d}}{\underline{\sigma}} \mathrm{E}\left[|\,\mathrm{E}[\Lambda^{-1}D1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty + |R|_\infty\right] \\
+ \varepsilon^{-1}\frac{(\log d)^{3/2}}{\underline{\sigma}} \mathrm{E}\left[|\,\mathrm{E}[V(\beta^{-1}) \mid \mathcal{G}]|_\infty\right] \\
+ \varepsilon^{-3}\frac{(\log d)^{7/2}}{\underline{\sigma}} \mathrm{E}\left[\max_{1\le j,k,l,m\le d} \mathrm{E}[|(\Lambda^{-1}D)_j D_k D_l D_m| 1_{\{|D|_\infty \le \beta^{-1}\}} \mid \mathcal{G}]\right]
\end{aligned}
$$

を得る. これと (9.35) をあわせて示すべき不等式を得る. □

9.4.8　最大不等式

この小節では, 定理 9.8 で与えられる不等式の上界に現れる種々の量を $W = S_n$ の場合に評価する際に必要となる最大不等式を証明する.

定義 9.5 (Rademacher 確率変数)**.** 確率変数 ϵ が **Rademacher 確率変数**であるとは, $\mathrm{P}(\epsilon = 1) = \mathrm{P}(\epsilon = -1) = 1/2$ が成り立つことをいう.

補題 9.12 (van der Vaart & Wellner (1996), Lemma 2.2.7)**.** $\epsilon_1, \dots, \epsilon_n$ を独立な Rademecher 確率変数列, $a_1, \dots, a_n$ を実数列とする. このとき,

$$
\left\| \sum_{i=1}^n a_i \epsilon_i \right\|_{\psi_2} \le \sqrt{6 \sum_{i=1}^n a_i^2}
$$

が成り立つ.

証明. $\lambda \in \mathbb{R}$ を任意に 1 つ固定する. 各 $i = 1, \dots, n$ について,

$$
\mathrm{E}[e^{\lambda a_i \epsilon_i}] = \frac{e^{\lambda a_i} + e^{-\lambda a_i}}{2} = \sum_{p=0}^\infty \frac{(\lambda a_i)^{2p}}{(2p)!} \le 1 + \sum_{p=0}^\infty \frac{(\lambda^2 a_i^2/2)^p}{p!} = e^{\lambda^2 a_i^2/2}
$$

が成り立つから,

$$
\mathrm{E}\left[\exp\left(\lambda \sum_{i=1}^n a_i \epsilon_i\right)\right] = \prod_{i=1}^n \mathrm{E}[e^{\lambda a_i \epsilon_i}] \le \exp\left(\frac{\lambda^2}{2}\sum_{i=1}^n a_i^2\right)
$$

が成り立つ. 従って, 補題 3.1 より, 任意の $u \ge 0$ に対して,

$$
\mathrm{P}\left(\left|\sum_{i=1}^n a_i \epsilon_i\right| \ge u\right) \le 2\exp\left(-\frac{u^2}{2\sum_{i=1}^n a_i^2}\right)
$$

が成り立つ. この評価と命題 2.3 より示すべき主張を得る. □

補題 9.13 (対称化不等式)**.** $Y_1, \dots, Y_n$ を独立な d 次元確率変数列で $\mathrm{E}\,|Y_i| < \infty$ $(i = 1, \dots, n)$ を満たすものとする. このとき, 任意の $p \ge 1$ に対して

$$\left\| \left| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right|_\infty \right\|_p \le 2 \left\| \left| \sum_{i=1}^n \epsilon_i Y_i \right|_\infty \right\|_p$$

が成り立つ. ここに, $\epsilon_1, \dots, \epsilon_n$ は独立な Rademecher 確率変数列で, $(Y_1, \dots, Y_n)$ と独立なものである.

証明. $Y_1', \dots, Y_n'$ を $Y_1, \dots, Y_n$ の独立なコピーとする. $\mathcal{Y} = \sigma(Y_1, \dots, Y_n)$ とおくと,

$$\left\| \left| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right|_\infty \right\|_p = \left\| \left| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i' \mid \mathcal{Y}]) \right|_\infty \right\|_p = \left\| \mathrm{E}\left[\sum_{i=1}^n (Y_i - Y_i') \mid \mathcal{Y} \right] \right|_\infty \Bigg\|_p$$

が成り立つから, Jensen の不等式より

$$\left\| \left| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right|_\infty \right\|_p \le \left\| \left| \sum_{i=1}^n (Y_i - Y_i') \right|_\infty \right\|_p$$

が成り立つ. $Y_1 - Y_1', \dots, Y_n - Y_n'$ は $\epsilon_1(Y_1 - Y_1'), \dots, \epsilon_n(Y_n - Y_n')$ と同分布だから,

$$\begin{aligned}
\left\| \left| \sum_{i=1}^n (Y_i - Y_i') \right|_\infty \right\|_p &= \left\| \left| \sum_{i=1}^n \epsilon_i (Y_i - Y_i') \right|_\infty \right\|_p \\
&\le \left\| \left| \sum_{i=1}^n \epsilon_i Y_i \right|_\infty \right\|_p + \left\| \left| \sum_{i=1}^n \epsilon_i Y_i' \right|_\infty \right\|_p = 2 \left\| \left| \sum_{i=1}^n \epsilon_i Y_i \right|_\infty \right\|_p
\end{aligned}$$

を得る. □

補題 9.14 (Nemirovski の不等式)**.** $Y_1, \dots, Y_n$ を独立な d 次元確率変数列で $\mathrm{E}\,|Y_i| < \infty$ $(i = 1, \dots, n)$ を満たすものとする. このとき, ある普遍定数 $C > 0$ が存在して, 任意の $p \ge 1$ に対して

$$\left\| \left| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right|_\infty \right\|_p \le C \sqrt{p \log d} \left\| \sqrt{\max_{1 \le j \le d} \sum_{i=1}^n Y_{ij}^2} \right\|_p .$$

証明. まず, 対称化不等式より

$$\left\| \left| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right|_\infty \right\|_p \le 2 \left\| \left| \sum_{i=1}^n \epsilon_i Y_i \right|_\infty \right\|_p$$

が成り立つ. ここに, $\epsilon_1, \dots, \epsilon_n$ は独立な Rademecher 確率変数列で, $(Y_1, \dots, Y_n)$ と独立なものである. 次に, 補題 9.12 と命題 2.2, 2.4 より, 任意の $y_1, \dots, y_n \in \mathbb{R}^d$ に対して

$$\left\| \left| \sum_{i=1}^n \epsilon_i y_i \right|_\infty \right\|_p \lesssim \sqrt{p (\log d) \max_{1 \le j \le d} \sum_{i=1}^n y_{ij}^2}$$

が成り立つ. 従って,

$$\mathrm{E}\left[\left|\sum_{i=1}^{n}\epsilon_i Y_i\right|_{\infty}^{p}\right] = \mathrm{E}\left[\mathrm{E}\left[\left|\sum_{i=1}^{n}\epsilon_i Y_i\right|_{\infty}^{p} \mid Y_1,\dots,Y_n\right]\right] \lesssim \mathrm{E}\left[\left(\sqrt{p(\log d)\max_{1\le j\le d}\sum_{i=1}^{n}Y_{ij}^2}\right)^p\right]$$

を得る. □

補題 9.15 (cf. Chernozhukov, Chetverikov & Kato (2015), Lemma 9)**.** $V_1,\dots,V_n$ を独立な d 次元確率変数列とし, 任意の $i=1,\dots,n$ と $j=1,\dots,d$ に対して $V_{ij}\ge 0$ を満たすものとする. このとき, ある普遍定数 $C>0$ が存在して, 任意の $p\ge 1$ に対して

$$\left\|\max_{1\le j\le d}\sum_{i=1}^{n}V_{ij}\right\|_p \le C\left(\max_{1\le j\le d}\sum_{i=1}^{n}\mathrm{E}[V_{ij}] + p\left\|\max_{1\le i\le n}\max_{1\le j\le d}V_{ij}\right\|_p \log d\right).$$

証明.

$$I := \left\|\max_{1\le j\le d}\sum_{i=1}^{n}V_{ij}\right\|_p, \qquad B := \max_{1\le i\le n}\max_{1\le j\le d}V_{ij}$$

とおく. 三角不等式より

$$I \le \max_{1\le j\le d}\sum_{i=1}^{n}\mathrm{E}[V_{ij}] + \left\|\left|\sum_{i=1}^{n}(V_i-\mathrm{E}[V_i])\right|_{\infty}\right\|_p$$

が成り立つ. Nemirovski の不等式と Schwarz の不等式より

$$\begin{aligned}\left\|\left|\sum_{i=1}^{n}(V_i-\mathrm{E}[V_i])\right|_{\infty}\right\|_p &\lesssim \sqrt{p\log d}\left\|\sqrt{\max_{1\le j\le d}\sum_{i=1}^{n}V_{ij}^2}\right\|_p \\ &\le \sqrt{p\log d}\left\|\sqrt{B\max_{1\le j\le d}\sum_{i=1}^{n}V_{ij}}\right\|_p \le \sqrt{p\|B\|_p\log d}\sqrt{I}\end{aligned}$$

が成り立つので,

$$I \lesssim \max_{1\le j\le d}\sum_{i=1}^{n}\mathrm{E}[V_{ij}] + \sqrt{p\|B\|_p\log d}\sqrt{I}$$

を得る. 正の定数 a,b に対して, 実数 x が不等式 $x^2\le a+bx$ を満たすための必要十分条件は $(b-\sqrt{b^2+4a})/2\le x\le(b+\sqrt{b^2+4a})/2$ なので,

$$\sqrt{I} \lesssim \sqrt{p\|B\|_p\log d} + \sqrt{p\|B\|_p\log d + 4\max_{1\le j\le d}\sum_{i=1}^{n}\mathrm{E}[V_{ij}]} \lesssim \sqrt{\max_{1\le j\le d}\sum_{i=1}^{n}\mathrm{E}[V_{ij}] + p\|B\|_p\log d}$$

を得る. 両辺を二乗して示すべき不等式を得る. □

補題 9.16 (cf. Chernozhukov, Chetverikov & Kato (2015), Lemma 8)**.** $Y_1, \dots, Y_n$ を独立な d 次元確率変数列で $\mathrm{E}\,|Y_i|^2 < \infty$ $(i = 1, \dots, n)$ を満たすものとする. このとき, ある普遍定数 $C > 0$ が存在して, 任意の $p \geq 1$ に対して

$$\left\| \left| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right|_\infty \right\|_p \leq C \left(\sqrt{p \max_{1 \leq j \leq d} \sum_{i=1}^n \mathrm{E}[Y_{ij}^2]} \sqrt{\log d} + p \sqrt{\left\| \max_{1 \leq i \leq n} \max_{1 \leq j \leq d} Y_{ij}^2 \right\|_p \log d} \right).$$

証明. Nemirovski の不等式より

$$\left\| \left| \sum_{i=1}^n (Y_i - \mathrm{E}[Y_i]) \right|_\infty \right\|_p \lesssim \sqrt{p \log d} \left\| \sqrt{\max_{1 \leq j \leq d} \sum_{i=1}^n Y_{ij}^2} \right\|_p$$

が成り立つ. また, Jensen の不等式と補題 9.15 より,

$$\left\| \sqrt{\max_{1 \leq j \leq d} \sum_{i=1}^n Y_{ij}^2} \right\|_p \lesssim \sqrt{\max_{1 \leq j \leq d} \sum_{i=1}^n \mathrm{E}[Y_{ij}^2] + p \left\| \max_{1 \leq i \leq n} \max_{1 \leq j \leq d} Y_{ij}^2 \right\|_p \log d}$$

が成り立つ. $a, b \geq 0$ に対して $\sqrt{a + b} \leq \sqrt{a} + \sqrt{b}$ が成り立つことから, 示すべき不等式が得られる. □

9.4.9 定理 9.1 の証明

まず, 劣指数性を仮定せずに得られる中間的な結果を証明する.

補題 9.17. 定理 9.1 の仮定が条件 (i) を除いて成立するならば, ある普遍定数 $C > 0$ が存在して, 任意の $\varepsilon > 0$ に対して

$$\begin{aligned}
&\sup_{x \in \mathbb{R}^d} |\mathrm{P}(S_n \leq x) - \mathrm{P}(Z_n \leq x)| \\
&\qquad \leq \frac{C}{b} \left(\frac{(\log d)^{3/2}}{\varepsilon n} \sum_{i=1}^n \mathrm{E}[|X_i|_\infty^2 1_{\{|X_i|_\infty > \sqrt{n}\varepsilon / \log d\}}] + \frac{B_n^2 (\log d)^{7/2}}{\varepsilon^3 n} + \varepsilon \sqrt{\log d} \right)
\end{aligned}$$

が成り立つ.

証明の中で次の初等的な不等式を用いる.

補題 9.18 (Chebyshev の結合不等式)**.** $f : \mathbb{R} \to \mathbb{R}, g : \mathbb{R} \to \mathbb{R}$ を 2 つの非減少関数とすると, 任意の確率変数 Y に対して,

$$\mathrm{E}[f(Y)g(Y)] \geq \mathrm{E}[f(Y)]\,\mathrm{E}[g(Y)].$$

証明. Y^* を Y の独立なコピーとする. f, g がともに非減少であることから, $(f(Y) - f(Y^*))(g(Y) - g(Y^*)) \geq 0$ が成り立つ. 両辺の期待値を取って

$$0 \leq \mathrm{E}[(f(Y) - f(Y^*))(g(Y) - g(Y^*))] = 2\{\mathrm{E}[f(Y)g(Y)] - \mathrm{E}[f(Y)]\,\mathrm{E}[g(Y)]\}$$

を得る. □

補題 9.17 の証明. 5 ステップに分けて証明する.

Step 1. $\xi_i = X_i/\sqrt{n}$ $(i = 1, \dots, n)$ とおく. $\xi^* = (\xi_1^*, \dots, \xi_n^*)$ を $\xi = (\xi_1, \dots, \xi_n)$ の独立なコピーとし, I を $\{1, \dots, n\}$ 上の一様分布に従う確率変数で ξ, ξ^* と独立であるようなものとする. $\mathbb{R}^{d\times n}$ 値確率変数 $\xi' = (\xi_1', \dots, \xi_n')$ を

$$\xi_i' = \begin{cases} \xi_i^* & \text{if } I = i, \\ \xi_i & \text{otherwise,} \end{cases} \qquad i = 1, \dots, n,$$

で定めると, 命題 9.6 より (ξ, ξ') は exchangeable pair である. 従って, $S_n' := \sum_{i=1}^n \xi_i'$ とおけば, (S_n, S_n') も exchangeable pair である. さらに, $\mathcal{G} := \sigma(\xi)$ とおくと,

$$\mathrm{E}[S_n' - S_n \mid \mathcal{G}] = \mathrm{E}[\xi_I^* - \xi_I \mid \mathcal{G}] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[\xi_i^* - \xi_i \mid \mathcal{G}] = -\frac{1}{n}\sum_{i=1}^n \xi_i = -\frac{1}{n}S_n$$

が成り立つから, $(W, W') = (S_n, S_n')$ に対して条件 (9.30) が $\Lambda = n^{-1}I_d, R = 0$ として成り立つ. 従って, 定理 9.8 を $(W, W') = (S_n, S_n'), \Sigma = \mathrm{Cov}[W]$ として適用することができて,

$$\begin{aligned} &\sup_{x\in\mathbb{R}^d} |\mathrm{P}(S_n \le x) - \mathrm{P}(Z_n \le x)| \\ &\quad\lesssim \frac{1}{b}\Bigg(\sqrt{\log d}\,\mathrm{E}\left[|\mathrm{E}[nD1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty\right] + \varepsilon^{-1}(\log d)^{3/2}\,\mathrm{E}\left[|V(\beta^{-1})|_\infty\right] \\ &\quad + \varepsilon^{-3}(\log d)^{7/2}\,\mathrm{E}\left[\max_{1\le j,k,l,m\le d} \mathrm{E}[n|D_jD_kD_lD_m|1_{\{|D|_\infty \le \beta^{-1}\}} \mid \mathcal{G}]\right] + \varepsilon\sqrt{\log d}\Bigg) \end{aligned} \tag{9.41}$$

が成り立つ. ただし, $D := S_n' - S_n, \beta = \varepsilon^{-1}\log d$, および

$$V(\beta^{-1}) := \Sigma - \frac{n}{2}\mathrm{E}[DD^\top 1_{\{|D|_\infty \le \beta^{-1}\}} \mid \mathcal{G}]$$

である.

Step 2. (9.41) の右辺第 1 項を評価する. $Y_i := \xi_i' - \xi_i$ $(i = 1, \dots, n)$ とおくと,

$$\mathrm{E}[nD1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}] = \sum_{i=1}^n \mathrm{E}[Y_i 1_{\{|Y_i|_\infty > \beta^{-1}\}} \mid \mathcal{G}]$$

と書き直せるので, Jensen の不等式より

$$\mathrm{E}\left[|\mathrm{E}[nD1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty\right] \le \mathrm{E}\left|\sum_{i=1}^n Y_i 1_{\{|Y_i|_\infty > \beta^{-1}\}}\right|_\infty$$

が成り立つ. 従って, Nemirovski の不等式より

$$\sqrt{\log d}\,\mathrm{E}\left[|\mathrm{E}[nD1_{\{|D|_\infty > \beta^{-1}\}} \mid \mathcal{G}]|_\infty\right] \lesssim (\log d)\,\mathrm{E}\left[\sqrt{\max_{1\le j\le d}\sum_{i=1}^n Y_{ij}^2 1_{\{|Y_i|_\infty > \beta^{-1}\}}}\right]$$

$$\leq (\log d)\sqrt{\sum_{i=1}^n \mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty > \beta^{-1}\}}]}$$

を得る.

Step 3. (9.41) の右辺第 2 項を評価する. $\tilde{Y}_i := Y_i 1_{\{|Y_i|_\infty \leq \beta^{-1}\}}$ $(i = 1, \ldots, n)$ とおいて, $V(\beta^{-1})$ を

$$\begin{aligned} V(\beta^{-1}) &= \Sigma - \frac{1}{2}\sum_{i=1}^n \mathrm{E}[\tilde{Y}_i\tilde{Y}_i^\top \mid \mathcal{G}] \\ &= \left(\Sigma - \frac{1}{2}\sum_{i=1}^n \mathrm{E}[\tilde{Y}_i\tilde{Y}_i^\top]\right) - \frac{1}{2}\sum_{i=1}^n \left(\mathrm{E}[\tilde{Y}_i\tilde{Y}_i^\top \mid \mathcal{G}] - \mathrm{E}[\tilde{Y}_i\tilde{Y}_i^\top]\right) \\ &=: I + II \end{aligned}$$

と分解する. $\Sigma = \frac{1}{2}\sum_{i=1}^n \mathrm{E}[Y_iY_i^\top]$ と書き直せることから,

$$|I|_\infty = \frac{1}{2}\left|\sum_{i=1}^n \mathrm{E}[Y_iY_i^\top 1_{\{|Y_i|_\infty > \beta^{-1}\}}]\right|_\infty \leq \frac{1}{2}\sum_{i=1}^n \mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty > \beta^{-1}\}}]$$

を得る. 一方で, Jensen の不等式より,

$$\mathrm{E}\,|II|_\infty \leq \frac{1}{2}\mathrm{E}\left|\sum_{i=1}^n \left(\tilde{Y}_i\tilde{Y}_i^\top - \mathrm{E}[\tilde{Y}_i\tilde{Y}_i^\top]\right)\right|_\infty$$

が成り立つから, 補題 9.16 より

$$\begin{aligned} \mathrm{E}\,|II|_\infty &\lesssim \sqrt{\max_{1\leq j,k\leq d}\sum_{i=1}^n \mathrm{E}[\tilde{Y}_{ij}^2\tilde{Y}_{ik}^2]}\sqrt{\log(d^2)} + \sqrt{\mathrm{E}\left[\max_{1\leq i\leq n}\max_{1\leq j,k\leq d}\tilde{Y}_{ij}^2\tilde{Y}_{ik}^2\right]}\log(d^2) \\ &\lesssim \sqrt{\max_{1\leq j\leq d}\sum_{i=1}^n \mathrm{E}[Y_{ij}^4]}\sqrt{\log d} + \beta^{-2}\log d \lesssim \frac{B_n\sqrt{\log d}}{\sqrt{n}} + \frac{\varepsilon^2}{\log d} \end{aligned}$$

を得る. 以上より,

$$\varepsilon^{-1}(\log d)^{3/2}\,\mathrm{E}\left[|V(\beta^{-1})|_\infty\right] \lesssim \varepsilon^{-1}(\log d)^{3/2}\sum_{i=1}^n \mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty > \beta^{-1}\}}] + \frac{B_n \log^2 d}{\varepsilon\sqrt{n}} + \varepsilon\sqrt{\log d}$$

が成り立つ.

Step 4. (9.41) の右辺第 3 項を評価する.

$$\begin{aligned} \max_{1\leq j,k,l,m\leq d}\mathrm{E}[n|D_jD_kD_lD_m|1_{\{|D|_\infty\leq\beta^{-1}\}} \mid \mathcal{G}] &= \max_{1\leq j,k,l,m\leq d}\sum_{i=1}^n \mathrm{E}[|\tilde{Y}_{ij}\tilde{Y}_{ik}\tilde{Y}_{il}\tilde{Y}_{im}| \mid \mathcal{G}] \\ &= \max_{1\leq j\leq d}\sum_{i=1}^n \mathrm{E}[\tilde{Y}_{ij}^4 \mid \mathcal{G}] \end{aligned}$$

と書き直せるから, Jensen の不等式より

$$\mathrm{E}\left[\max_{1\le j,k,l,m\le d}\mathrm{E}[n|D_jD_kD_lD_m|1_{\{|D|_\infty\le\beta^{-1}\}}\mid\mathcal{G}]\right]\le\mathrm{E}\left[\max_{1\le j\le d}\sum_{i=1}^n\tilde{Y}_{ij}^4\right]$$

が成り立つ. 補題 9.15 より

$$\mathrm{E}\left[\max_{1\le j\le d}\sum_{i=1}^n\tilde{Y}_{ij}^4\right]\lesssim\max_{1\le j\le d}\sum_{i=1}^n\mathrm{E}[\tilde{Y}_{ij}^4]+\mathrm{E}\left[\max_{1\le i\le n}\max_{1\le j\le d}\tilde{Y}_{ij}^4\right]\log d\lesssim\frac{B_n^2}{n}+\frac{\varepsilon^4}{\log^3 d}$$

が成り立つから,

$$\varepsilon^{-3}(\log d)^{7/2}\,\mathrm{E}\left[\max_{1\le j,k,l,m\le d}\mathrm{E}[n|D_jD_kD_lD_m|1_{\{|D|_\infty\le\beta^{-1}\}}\mid\mathcal{G}]\right]\lesssim\frac{B_n^2(\log d)^{7/2}}{\varepsilon^3 n}+\varepsilon\sqrt{\log d}$$

を得る.

Step 5. Steps 2–4 で得た評価を (9.41) に挿入すると,

$$\begin{aligned}\sup_{x\in\mathbb{R}^d}&|\mathrm{P}(S_n\le x)-\mathrm{P}(Z_n\le x)|\\&\lesssim\frac{1}{b}\left(\sqrt{\sum_{i=1}^n\mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty>\beta^{-1}\}}]\log d}+\varepsilon^{-1}(\log d)^{3/2}\sum_{i=1}^n\mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty>\beta^{-1}\}}]\right.\\&\qquad\left.+\frac{B_n\log^2 d}{\varepsilon\sqrt{n}}+\frac{B_n^2(\log d)^{7/2}}{\varepsilon^3 n}+\varepsilon\sqrt{\log d}\right)\end{aligned}$$

を得る. 相加相乗平均不等式より

$$\sqrt{2\sum_{i=1}^n\mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty>\beta^{-1}\}}]\log d}\le\varepsilon^{-1}(\log d)^{3/2}\sum_{i=1}^n\mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty>\beta^{-1}\}}]+\varepsilon\sqrt{\log d},$$

$$2\frac{B_n\log^2 d}{\varepsilon\sqrt{n}}\le\frac{B_n^2(\log d)^{7/2}}{\varepsilon^3 n}+\varepsilon\sqrt{\log d}$$

が成り立つので,

$$\begin{aligned}\sup_{x\in\mathbb{R}^d}&|\mathrm{P}(S_n\le x)-\mathrm{P}(Z_n\le x)|\\&\lesssim\frac{1}{b}\left(\varepsilon^{-1}(\log d)^{3/2}\sum_{i=1}^n\mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty>\beta^{-1}\}}]+\frac{B_n^2(\log d)^{7/2}}{\varepsilon^3 n}+\varepsilon\sqrt{\log d}\right)\end{aligned}\tag{9.42}$$

を得る. さらに,

$$\sum_{i=1}^n\mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty>\beta^{-1}\}}]\lesssim\sum_{i=1}^n\mathrm{E}[|\xi_i|_\infty^2 1_{\{|\xi_i|_\infty>(2\beta)^{-1}\}}]+\sum_{i=1}^n\mathrm{E}[|\xi_i|_\infty^2 1_{\{|\xi_i^*|_\infty>(2\beta)^{-1}\}}]$$

であり, Chebyshev の結合不等式より

$$\mathrm{E}[|\xi_i|_\infty^2 1_{\{|\xi_i^*|_\infty>(2\beta)^{-1}\}}]=\mathrm{E}[|\xi_i|_\infty^2]\,\mathrm{E}[1_{\{|\xi_i|_\infty>(2\beta)^{-1}\}}]\le\mathrm{E}[|\xi_i|_\infty^2 1_{\{|\xi_i|_\infty>(2\beta)^{-1}\}}]$$

が成り立つから,

$$\sum_{i=1}^{n} \mathrm{E}[|Y_i|_\infty^2 1_{\{|Y_i|_\infty > \beta^{-1}\}}] \lesssim \sum_{i=1}^{n} \mathrm{E}[|\xi_i|_\infty^2 1_{\{|\xi_i|_\infty > (2\beta)^{-1}\}}] \tag{9.43}$$

を得る. (9.42) と (9.43) をあわせて ε を 2ε に置き換えれば示すべき不等式を得る. □

定理 9.1 の証明. この節の冒頭で述べたように, 定理 9.1 の仮定の下で (9.12) 式を示せばよい. (9.12) 式の左辺は 1 以下であることと, $b \le 1$ であることに注意すれば,

$$\frac{B_n^2 \log^5(dn)}{n} \le 1 \tag{9.44}$$

と仮定して一般性を失わない. 実際, (9.44) が成り立たない場合, $C \ge 1$ となるようにとれば (9.12) 式は成立する.

(9.44) を仮定した下で, 補題 9.17 を

$$\varepsilon = 2\left(\frac{B_n^2 \log^3(dn)}{n}\right)^{1/4}$$

として適用する. 定義より

$$\frac{B_n^2 (\log d)^{7/2}}{\varepsilon^3 n} + \varepsilon\sqrt{\log d} \lesssim \left(\frac{B_n^2 \log^5(dn)}{n}\right)^{1/4}$$

が直ちに従うから,

$$\frac{(\log d)^{3/2}}{\varepsilon n} \sum_{i=1}^{n} \mathrm{E}[|X_i|_\infty^2 1_{\{|X_i|_\infty > \sqrt{n}\varepsilon/\log d\}}] \lesssim \left(\frac{B_n^2 \log^5(dn)}{n}\right)^{1/4} \tag{9.45}$$

を示せば証明は完成する. 各 $i = 1, \ldots, n$ について, Schwarz の不等式より

$$\mathrm{E}[|X_i|_\infty^2 1_{\{|X_i|_\infty > \sqrt{n}\varepsilon/\log d\}}] \le \sqrt{\mathrm{E}[|X_i|_\infty^4]\, \mathrm{P}(|X_i|_\infty > \sqrt{n}\varepsilon/\log d)}$$

が成り立つ. 命題 2.2 と 2.4 より $\mathrm{E}[|X_i|_\infty^4] \lesssim (B_n \log d)^4$ が成り立つ. 一方で, (9.44) より

$$\frac{\sqrt{n}\varepsilon}{\log d} \ge 2\left(\frac{nB_n^2}{\log(dn)}\right)^{1/4} \ge 2B_n \log(dn)$$

が成り立つから,

$$\begin{aligned}
\mathrm{P}(|X_i|_\infty > \sqrt{n}\varepsilon/\log d) &\le \mathrm{P}(|X_i|_\infty > 2B_n \log(dn)) \\
&\le \sum_{j=1}^{d} \mathrm{P}(|X_{ij}| > 2B_n \log(dn)) \le 2d\exp(-2\log(dn)) = \frac{2}{dn^2}
\end{aligned}$$

が成り立つ. 以上より,

$$\begin{aligned}
&\frac{(\log d)^{3/2}}{\varepsilon n} \sum_{i=1}^{n} \mathrm{E}[|X_i|_\infty^2 1_{\{|X_i|_\infty > \sqrt{n}\varepsilon/\log d\}}] \\
&\lesssim \left(\frac{n \log^3(dn)}{B_n^2}\right)^{1/4} (B_n \log d)^2 \frac{1}{\sqrt{dn}} \lesssim \frac{B_n^{3/2} (\log(dn))^{3/4}}{n^{3/4}} = \left(\frac{B_n^2 \log(dn)}{n}\right)^{3/4}
\end{aligned}$$

を得る. 従って, (9.44) より (9.45) が従う. □

9.5　定理 9.2 の証明

補題 9.19. ある普遍定数 $C \geq 1$ が存在して,

$$\frac{1}{b}\left(\frac{B_n^2 \log^5(dn)}{n}\right)^{1/4} \leq 1 \tag{9.46}$$

ならば, 事象

$$|\widehat{\Sigma}_n - \Sigma|_\infty \leq C\sqrt{\frac{B_n^2 \log(dn)}{n}} \tag{9.47}$$

が $1 - 1/n$ 以上の確率で成り立つ.

証明. $\kappa_n = 2B_n \log(dn)$ とし, 各 $i = 1, \ldots, n$ と $j = 1, \ldots, d$ について $Y_{ij} := X_{ij} 1_{\{|X_{ij}| \leq \kappa_n\}}$ とおく. このとき, 任意の $t > 0$ に対して,

$$\mathrm{P}\left(|\widehat{\Sigma}_n - \Sigma|_\infty > t\right) \leq \mathrm{P}\left(\max_{1 \leq j,k \leq d} |R_{jk}| > t\right) + \mathrm{P}\left(\max_{i=1,\ldots,n} |X_i|_\infty > \kappa_n\right) \tag{9.48}$$

が成り立つ. ただし,

$$R_{jk} := \frac{1}{n}\sum_{i=1}^n Y_{ij}Y_{ik} - \bar{X}_{n,j}\bar{X}_{n,k} - \Sigma_{jk}$$

と定める. (9.48) の右辺第二項は, union bound と Markov の不等式によって,

$$\mathrm{P}\left(\max_{i=1,\ldots,n} |X_i|_\infty > \kappa_n\right) \leq \sum_{i=1}^n \sum_{j=1}^d \mathrm{P}\left(|X_{ij}| > \kappa_n\right) \leq \frac{2}{dn} \leq \frac{2}{3n}$$

と評価できる. t を適切に選ぶと (9.48) の右辺第一項も小さくなることを示すために, $|R_{jk}|$ を以下のように評価する:

$$\begin{aligned}
|R_{jk}| &\leq \left|\frac{1}{n}\sum_{i=1}^n (Y_{ij}Y_{ik} - \mathrm{E}[Y_{ij}Y_{ik}])\right| + \left|\frac{1}{n}\sum_{i=1}^n \mathrm{E}[Y_{ij}Y_{ik}] - \Sigma_{jk}\right| + |\bar{X}_{n,j}\bar{X}_{n,k}| \\
&=: I_{jk} + II_{jk} + III_{jk}.
\end{aligned}$$

II_{jk} は次のように評価できる:

$$\begin{aligned}
|II_{jk}| &= \left|\frac{1}{n}\sum_{i=1}^n (\mathrm{E}[X_{ij}Y_{ik}1_{\{|X_{ij}| > \kappa_n\}}] + \mathrm{E}[X_{ij}X_{ik}1_{\{|X_{ik}| > \kappa_n\}}])\right| \\
&\leq 2\max_i \max_{j,k} \mathrm{E}[|X_{ij}X_{ik}|1_{\{|X_{ij}| > \kappa_n\}}] \\
&\leq 2\max_i \max_j \sqrt{\mathrm{E}[X_{ij}^4]} \max_j \sqrt{\mathrm{P}(|X_{ij}| > \kappa_n)} \ (\because \text{Schwarz の不等式}) \\
&\lesssim \frac{B_n^2}{dn} \ (\because \text{命題 2.4, Markov の不等式}).
\end{aligned}$$

一方で, 任意の $p \geq 1$ に対して, 補題 9.16 より

$$\|I_{jk}\|_p \lesssim \frac{1}{n}\sqrt{p\sum_{i=1}^n \mathrm{E}[Y_{ij}^2Y_{ik}^2]} + \frac{p}{n}\sqrt{\left\|\max_i |Y_{ij}Y_{ik}|^2\right\|_p} \lesssim \sqrt{\frac{pB_n^2}{n}} + \frac{p\kappa_n^2}{n}$$

が成り立つ. また, 命題 3.1 と補題 3.3 より,

$$\left\|\bar{X}_{n,j}\right\|_{\psi_1} \lesssim \sqrt{\frac{B_n^2}{n}} + \frac{B_n}{n} \lesssim \sqrt{\frac{B_n^2}{n}}$$

が成り立つから, Schwarz の不等式と命題 2.4 より

$$\|III_{jk}\|_p \lesssim \frac{p^2B_n^2}{n}$$

が成り立つ. 特に, $p = 2\log(dn)$ に対して, (9.46) と $b \leq 1$ に注意すると,

$$\begin{aligned}
\frac{p\kappa_n^2}{n} &= 8\frac{B_n^2\log^3(dn)}{n} = 8\sqrt{\frac{B_n^2\log(dn)}{n}}\sqrt{\frac{B_n^2\log^5(dn)}{n}} \leq 8\sqrt{\frac{B_n^2\log(dn)}{n}}, \\
\frac{p^2B_n^2}{n} &= 4\sqrt{\frac{B_n^2\log(dn)}{n}}\sqrt{\frac{B_n^2\log^3(dn)}{n}} \leq \sqrt{\frac{B_n^2\log(dn)}{n}}
\end{aligned}$$

が成り立つから,

$$\|I_{jk}\|_p + \|III_{jk}\|_p \lesssim \sqrt{\frac{B_n^2\log(dn)}{n}}$$

を得る. 以上より, ある普遍定数 $C_1 > 0$ が存在して,

$$\|R_{jk}\|_p \leq C_1\sqrt{\frac{B_n^2\log(dn)}{n}}$$

が成り立つことがわかる. 従って, union bound と Markov の不等式より,

$$\mathrm{P}\left(\max_{j,k}|R_{jk}| > eC_1\sqrt{\frac{B_n^2\log(dn)}{n}}\right) \leq d^2e^{-p} = \frac{1}{n^2} \leq \frac{1}{3n}$$

を得る. 故に, (9.48) において $t = eC_1\sqrt{B_n^2\log(dn)/n}$ と選べば,

$$\mathrm{P}\left(|\widehat{\Sigma}_n - \Sigma|_\infty > t\right) \leq \frac{1}{3n} + \frac{2}{3n} = \frac{1}{n}$$

が成り立つ. よって, $C = 1 \vee eC_1$ とすれば示すべき主張を得る. □

補題 9.20. ある普遍定数 $C > 0$ が存在して

$$\mathrm{P}\left(\sup_{A\in\mathcal{R}_d}|\mathrm{P}(S_n^* \in A \mid X) - \mathrm{P}(Z_n \in A)| > \frac{C}{b}\left(\frac{B_n^2\log^5(dn)}{n}\right)^{1/4}\right) \leq \frac{1}{n}$$

が成り立つ.

証明.

$$\rho^* := \sup_{A\in\mathcal{R}_d} |\mathrm{P}\left(S_n^* \in A \mid X\right) - \mathrm{P}\left(Z_n \in A\right)|$$

とおく. $\rho^* \le 1$ であることに注意すると, (9.46) を仮定して一般性を失わない. いま, 定理 9.7 より, ある普遍定数 $C_1 > 0$ が存在して

$$\sup_{A\in\mathcal{R}_d} |\mathrm{P}\left(Z_n \in A\right) - \mathrm{P}\left(S_n^* \in A \mid X\right)| \le C_1\sqrt{\frac{|\Sigma - \widehat{\Sigma}_n|_\infty (\log d)^2}{b^2}}$$

が成り立つ. 従って, 補題 9.19 に現れる普遍定数を C_2 として, $C = C_1\sqrt{C_2}$ とすれば示すべき主張を得る. □

定理 9.2 の証明. (9.3) の左辺は 1 以下であることに注意すると, (9.46) を仮定して一般性を失わない. 定理 9.1 の普遍定数を C_1, 補題 9.20 の普遍定数を C_2 とし,

$$\Delta := (C_1 \vee C_2)\frac{1}{b}\left(\frac{B_n^2 \log^5(dn)}{n}\right)^{1/4}$$

とおく. 定義より,

$$\sup_{A\in\mathcal{R}_d} |\,\mathrm{P}(S_n \in A) - \mathrm{P}(Z_n \in A)| \le \Delta \tag{9.49}$$

であり, また

$$\mathcal{E}_n := \left\{\sup_{A\in\mathcal{R}_d} |\mathrm{P}\left(S_n^* \in A \mid X\right) - \mathrm{P}\left(Z_n \in A\right)| \le \Delta\right\}$$

とおくと $\mathrm{P}(\mathcal{E}_n) \ge 1 - 1/n$ が成り立つ.

次に, $T_n := \max_{1\le j\le d} |S_{n,j}|, T_n^* := \max_{1\le j\le d} |S_{n,j}^*|$ とおく. (9.49) から特に,

$$\sup_{t\in\mathbb{R}} P(T_n = t) \le \Delta \tag{9.50}$$

であることに注意する. さらに, 各 $\gamma \in (0,1)$ について, T_n の分布の $(1-\gamma)$-分位点を c_γ とする. 便宜上 $\gamma \le 0$ に対しては $c_\gamma := -\infty, \gamma \ge 1$ に対しては $c_\gamma := \infty$ とそれぞれおく. このとき任意の $\gamma \in \mathbb{R}$ に対して $\mathrm{P}(T_n < c_\gamma) \le 1-\gamma \le \mathrm{P}(T_n \le c_\gamma)$ が成り立つことに注意する. 従って, 事象 $\mathcal{E}_n$ 上で

$$1 - \alpha + \Delta \le \mathrm{P}(T_n \le c_{\alpha-\Delta}) \le \mathrm{P}(T_n^* \le c_{\alpha-\Delta} \mid X) + \Delta$$

が成り立つ. 分位点の定義から, これは $\mathcal{E}_n$ 上で $c_\alpha^* \le c_{\alpha-\Delta}$ が成り立つことを意味する. 一方で, $\mathcal{E}_n$ 上で

$$\mathrm{P}(T_n^* \le c_{\alpha+3\Delta} \mid X) \le \mathrm{P}(T_n \le c_{\alpha+3\Delta}) + \Delta$$

も成り立つ. ここで, (9.50) より

$$\mathrm{P}(T_n \le c_{\alpha+3\Delta}) \le \mathrm{P}(T_n < c_{\alpha+3\Delta}) + \Delta \le 1 - \alpha - 2\Delta$$

となることに注意すると, $\mathcal{E}_n$ 上で

$$P(T_n^* \le c_{\alpha+3\Delta} \mid X) \le 1-\alpha-\Delta < 1-\alpha$$

が成り立つことが従う. 従って $\mathcal{E}_n$ 上で $c_\alpha^* > c_{\alpha+3\Delta}$ が成り立つ. 以上より,

$$\mathrm{P}(c_{\alpha+3\Delta} < c_\alpha^* \le c_{\alpha-\Delta}) \ge \mathrm{P}(\mathcal{E}_n) \ge 1-1/n$$

であるから,

$$\mathrm{P}(T_n > c_\alpha^*) \le \mathrm{P}(T_n > c_{\alpha+3\Delta}) + \frac{1}{n^2} \le \alpha + 3\Delta + \frac{1}{n}$$

が成り立ち, また, (9.50) に注意すると

$$\mathrm{P}(T_n > c_\alpha^*) \ge \mathrm{P}(T_n > c_{\alpha-\Delta}) - \frac{1}{n} \ge \mathrm{P}(T_n \ge c_{\alpha-\Delta}) - \Delta - \frac{1}{n} \ge \alpha - 2\Delta - \frac{1}{n}$$

が成り立つ. 2 式をあわせて示すべき不等式を得る. □

9.6　定理 9.3 の証明

補題 9.21 (Le Cam (1986), Chapter 15, Lemma 2)**.** V, W を 2 つの確率変数である定数 $r_1, r_2 > 0$ に対して $\mathrm{P}(|V-W| > r_1) \le r_2$ を満たすものとする. このとき,

$$\sup_{t\in\mathbb{R}} |\,\mathrm{P}(V \le t) - \mathrm{P}(W \le t)| \le \sup_{t\in\mathbb{R}} \mathrm{P}(t < W \le t+r_1) + r_2.$$

証明. 任意に $t \in \mathbb{R}$ に対して

$$\begin{aligned}\mathrm{P}(V \le t) &\le \mathrm{P}(V \le t, |V-W| \le r_1) + r_2 \le \mathrm{P}(W \le t+r_1) + r_2 \\ &\le \mathrm{P}(W \le t) + \mathrm{P}(t < W \le t+r_1) + r_2\end{aligned}$$

および

$$\mathrm{P}(W \le t) \le \mathrm{P}(W \le t-r_1) + \mathrm{P}(t-r_1 < W \le t) \le \mathrm{P}(V \le t) + r_2 + \mathrm{P}(t-r_1 < W \le t)$$

が成り立つことから従う. □

補題 9.22. 任意の $x \ge 0$ に対して, $1-\Phi(x) \le e^{-x^2/2}$ が成り立つ.

証明. $Z \sim N(0,1)$ とすると, 任意の $\lambda \in \mathbb{R}$ に対して $\mathrm{E}[e^{\lambda Z}] = \exp(\lambda^2/2)$ が成り立つから, 補題 3.1 を $b=0, v=1$ で適用して示すべき不等式を得る. □

定理 9.3 の証明. $\sigma_j := \sqrt{\mathrm{Var}[S_{n,j}]}$ $(j=1,\dots,d)$ とおく. また, 補題 9.19 に現れる普遍定数を C_1 とする. 定理 9.2 の証明と同様の理由で,

$$\frac{1}{b}\left(\frac{B_n^2 \log^5(dn)}{n}\right)^{1/4} \le \frac{1}{2C_1} \tag{9.51}$$

と仮定して一般性を失わない. このとき, $C_1 \geq 1$ より (9.46) も成り立つことに注意する. 従って,

$$C_1\sqrt{\frac{B_n^2\log(dn)}{n}} \leq C_1\left(\frac{B_n^2\log^5(dn)}{n}\right)^{1/4} \leq \frac{b}{2}$$

が成り立つ. 故に, (9.47) が成立するという事象を $\mathcal{E}_n$ とすれば, $\mathcal{E}_n$ 上では任意の $j = 1, \dots, d$ に対して

$$\hat{\sigma}_{n,j}^2 \geq \sigma_j^2 - |\widehat{\Sigma}_n - \Sigma|_\infty \geq \sigma_j^2/2 \geq b/2$$

が成り立つ.

以下, 記号の簡単のために

$$\Delta_n := \frac{1}{b}\left(\frac{B_n^2\log^5(dn)}{n}\right)^{1/4}$$

とおく. 定理 9.1 より

$$\sup_{x\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\leq j\leq d}\frac{|S_{n,j}|}{\sigma_j} \leq x\right) - \mathrm{P}\left(\max_{1\leq j\leq d}\frac{|Z_{n,j}|}{\sigma_j} \leq x\right)\right| \lesssim \Delta_n \tag{9.52}$$

が成り立つので, 定理 9.2 の証明と同様の議論によって, 次の 2 つの不等式を示せば証明は完成する.

$$\sup_{x\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\leq j\leq d}\frac{|S_{n,j}|}{\hat{\sigma}_{n,j}} \leq x\right) - \mathrm{P}\left(\max_{1\leq j\leq d}\frac{|S_{n,j}|}{\sigma_j} \leq x\right)\right| \lesssim \Delta_n, \tag{9.53}$$

$$\mathrm{P}\left(\sup_{x\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\leq j\leq d}\frac{|S_{n,j}^*|}{\hat{\sigma}_{n,j}} \leq x \mid X\right) - \mathrm{P}\left(\max_{1\leq j\leq d}\frac{|Z_{n,j}|}{\sigma_j} \leq x\right)\right| > c\Delta_n\right) \leq \frac{1}{n}. \tag{9.54}$$

ただし, $c > 0$ はある普遍定数を表す.

<u>(9.53) の証明.</u>

$$r_1 = \frac{2}{b}\left(\frac{B_n^2\log^3(dn)}{n}\right)^{1/4}$$

とおく. (9.52) と Nazarov の不等式より,

$$\sup_{x\in\mathbb{R}}\mathrm{P}\left(x < \max_{1\leq j\leq d}\frac{|S_{n,j}|}{\sigma_j} \leq x + r_1\right) \lesssim \Delta_n + r_1\sqrt{\log d} \lesssim \Delta_n$$

が成り立つので, 補題 9.21 より

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\leq j\leq d}\frac{|S_{n,j}|}{\hat{\sigma}_{n,j}} \leq x\right) - \mathrm{P}\left(\max_{1\leq j\leq d}\frac{|S_{n,j}|}{\sigma_j} \leq x\right)\right| \\ &\lesssim \mathrm{P}\left(\left|\max_{1\leq j\leq d}\frac{|S_{n,j}|}{\hat{\sigma}_{n,j}} - \max_{1\leq j\leq d}\frac{|S_{n,j}|}{\sigma_j}\right| > r_1\right) + \Delta_n\end{aligned} \tag{9.55}$$

を得る.

$$\left|\max_{1\leq j\leq d}\frac{|S_{n,j}|}{\hat{\sigma}_{n,j}} - \max_{1\leq j\leq d}\frac{|S_{n,j}|}{\sigma_j}\right| \leq \max_{1\leq j\leq d}\frac{|S_{n,j}|}{\sigma_j}\max_{1\leq j\leq d}\frac{|\hat{\sigma}_{n,j}^2 - \sigma_j^2|}{\hat{\sigma}_{n,j}(\hat{\sigma}_{n,j} + \sigma_j)}$$

と評価できるから,

$$\mathrm{P}\left(\left|\max_{1\le j\le d}\frac{|S_{n,j}|}{\hat{\sigma}_{n,j}}-\max_{1\le j\le d}\frac{|S_{n,j}|}{\sigma_j}\right|>r_1\right)\le \mathrm{P}\left(\max_{1\le j\le d}\frac{|S_{n,j}|}{\sigma_j}\cdot\frac{2C_1}{b^2}\sqrt{\frac{B_n^2\log(dn)}{n}}>r_1\right)+\mathrm{P}(\mathcal{E}_n^c) \tag{9.56}$$

が成り立つ. ここで, 補題 9.22 より

$$\mathrm{P}\left(\max_{1\le j\le d}\frac{|Z_{n,j}|}{\sigma_j}>\sqrt{2\log(dn)}\right)\le\sum_{j=1}^d\mathrm{P}\left(\frac{|Z_{n,j}|}{\sigma_j}>\sqrt{2\log(dn)}\right)\le 2d\exp(-\log(dn))=\frac{2}{n}$$

が成り立つから, (9.52) とあわせて

$$\mathrm{P}\left(\max_{1\le j\le d}\frac{|S_{n,j}|}{\sigma_j}>\sqrt{2\log(dn)}\right)\lesssim\Delta_n \tag{9.57}$$

を得る. このことと, (9.51) より

$$\sqrt{2\log(dn)}\frac{2C_1}{b^2}\sqrt{\frac{B_n^2\log(dn)}{n}}\le\frac{\sqrt{2}}{b}\left(\frac{B_n^2\log^2(dn)}{n}\right)^{1/4}<r_1$$

となることに注意すると,

$$\mathrm{P}\left(\max_{1\le j\le d}\frac{|S_{n,j}|}{\sigma_j}\cdot\frac{2C_1}{b^2}\sqrt{\frac{B_n^2\log(dn)}{n}}>r_1\right)\lesssim\Delta_n$$

が成り立つ. これと (9.55)–(9.56) および $\mathrm{P}(\mathcal{E}_n^c)\le 1/n$ をあわせて (9.53) を得る.

(9.54) の証明. 定理 9.7 より, ある普遍定数 $C_2>0$ が存在して

$$\sup_{x\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\le j\le d}\frac{|S^*_{n,j}|}{\hat{\sigma}_{n,j}}\le x\mid X\right)-\mathrm{P}\left(\max_{1\le j\le d}\frac{|Z_{n,j}|}{\sigma_j}\le x\right)\right|\le C_2\sqrt{(\log d)^2\max_{1\le j,k\le d}\left|\frac{\widehat{\Sigma}_{n,jk}}{\hat{\sigma}_{n,j}\hat{\sigma}_{n,k}}-\frac{\Sigma_{jk}}{\sigma_j\sigma_k}\right|}$$

が成り立つ. ここで,

$$\begin{aligned}
&\max_{1\le j,k\le d}\left|\frac{\widehat{\Sigma}_{n,jk}}{\hat{\sigma}_{n,j}\hat{\sigma}_{n,k}}-\frac{\Sigma_{n,jk}}{\sigma_j\sigma_k}\right|\\
&\max_{1\le j,k\le d}\left|\frac{\widehat{\Sigma}_{n,jk}}{\hat{\sigma}_{n,j}\hat{\sigma}_{n,k}}-\frac{\widehat{\Sigma}_{n,jk}}{\sigma_j\sigma_k}\right|+\max_{1\le j,k\le d}\left|\frac{\widehat{\Sigma}_{n,jk}}{\sigma_j\sigma_k}-\frac{\Sigma_{n,jk}}{\sigma_j\sigma_k}\right|\\
&\le\max_{1\le j,k\le d}\frac{|\widehat{\Sigma}_{n,jk}|}{\sigma_j\sigma_k}\frac{|\widehat{\sigma}_{n,j}^2\widehat{\sigma}_{n,k}^2-\sigma_j^2\sigma_k^2|}{\hat{\sigma}_{n,j}\hat{\sigma}_{n,k}(\widehat{\sigma}_{n,j}\widehat{\sigma}_{n,k}+\sigma_j\sigma_k)}+\frac{1}{b^2}|\widehat{\Sigma}_n-\Sigma|_\infty\\
&\le\max_{1\le j,k\le d}\frac{|\widehat{\Sigma}_{n,jk}|}{\sigma_j\sigma_k}\left(\frac{|\widehat{\sigma}_{n,j}^2-\sigma_j^2|}{\hat{\sigma}_{n,j}^2}+\frac{\sigma_j|\widehat{\sigma}_{n,k}^2-\sigma_k^2|}{\hat{\sigma}_{n,j}\hat{\sigma}_{n,k}\sigma_k}\right)+\frac{1}{b^2}|\widehat{\Sigma}_n-\Sigma|_\infty
\end{aligned}$$

および

$$\max_{1\le j,k\le d}\frac{|\widehat{\Sigma}_{n,jk}|}{\sigma_j\sigma_k}\le\max_{1\le j,k\le d}\frac{|\Sigma_{jk}|}{\sigma_j\sigma_k}+\frac{|\widehat{\Sigma}_n-\Sigma|_\infty}{b^2}\le 1+\frac{|\widehat{\Sigma}_n-\Sigma|_\infty}{b^2}$$

に注意すると, ある普遍定数 $C_3>0$ が存在して, $\mathcal{E}_n$ 上では

$$\max_{1\le j,k\le d}\left|\frac{\widehat{\Sigma}_{n,jk}}{\hat{\sigma}_{n,j}\hat{\sigma}_{n,k}}-\frac{\Sigma_{n,jk}}{\sigma_j\sigma_k}\right|\le\frac{C_3}{b^2}\sqrt{\frac{B_n^2\log(dn)}{n}}$$

が成り立つ. 従って, $c=C_2\sqrt{C_3}$ とすれば (9.54) が成り立つ. □

9.7　9.3 節の結果の証明

この節では, 2 つの確率変数 ξ,η に対して, T のみに依存する正の定数 C が存在して $\xi\le C\eta$ が成り立つことを, 記号 $\xi\lesssim\eta$ で表す.

9.7.1　定理 9.4 の証明

補題 9.23. N を正整数とし, 各 $i=1,\dots,N$ について可測関数 $f_i:[0,T]\to\mathbb{R}^r$ で $\int_0^T|f(t)|^2dt<\infty$ を満たすものが与えられているとする. このとき,

$$\left(\int_0^T f_1(t)^\top dW_t,\dots,\int_0^T f_N(t)^\top dW_t\right)^\top$$

は平均 0, 共分散行列 $V:=(\int_0^T f_i(t)\cdot f_j(t)dt)_{1\le i,j\le N}$ をもつ N 次元正規分布に従う.

演習問題 7. 補題 9.23 を示せ.

補題 9.24 (Isserlis の公式)**.** Z を平均 0 の 4 次元正規確率変数とすると,

$$\mathrm{E}[Z_1Z_2Z_3Z_4]=\mathrm{E}[Z_1Z_2]\,\mathrm{E}[Z_3Z_4]+\mathrm{E}[Z_1Z_3]\,\mathrm{E}[Z_2Z_4]+\mathrm{E}[Z_1Z_4]\,\mathrm{E}[Z_2Z_3].$$

証明. Stein の等式 (補題 9.9) を $j=1, f(z)=z_2z_3z_4$ として適用すればよい. □

定理 9.4 の証明. まず, $s\in[0,T]$ に対して $\zeta_s\sim\mathrm{N}(0,c_s)$ とすると, Isserlis の公式より $\Sigma=\int_0^T\mathrm{Cov}[\mathrm{vec}(\zeta_s\zeta_s^\top)]ds$ と書き直せるから, Σ は半正定値対称行列である. 次に, $b=0$ であり, σ が非ランダムであることから, 補題 9.23 より, $\Delta_1^nX,\dots,\Delta_n^nX$ は独立であり, 各 $i=1,\dots,n$ について $\Delta_i^nX\sim\mathrm{N}(0,\int_{t_{i-1}}^{t_i}c_sds)$ となることに注意しておく. 以下では (9.8) を 2 ステップに分けて証明する.

Step 1. まず, T のみに依存する定数 $c_1>0$ が存在して, 任意の $j,k,l,m\in\{1,\dots,d\}$ に対して,

$$\left|n\,\mathrm{Cov}\left[\widehat{[X^j,X^k]}_T,\widehat{[X^l,X^m]}_T\right]-\Sigma_{(j-1)d+k,(l-1)d+m}\right|\le c_1\frac{K^4}{n^\gamma}\tag{9.58}$$

が成り立つことを示す. 独立性と Isserlis の公式より,

$$n\,\mathrm{Cov}\left[\widehat{[X^j,X^k]}_T,\widehat{[X^l,X^m]}_T\right]$$

$$= n \sum_{i=1}^{n} \mathrm{Cov}\left[\Delta_i^n X^j \Delta_i^n X^k, \Delta_i^n X^l \Delta_i^n X^m\right]$$
$$= n \sum_{i=1}^{n} \left(\left(\int_{t_{i-1}}^{t_i} c_t^{jl} dt \right) \left(\int_{t_{i-1}}^{t_i} c_t^{km} dt \right) + \left(\int_{t_{i-1}}^{t_i} c_t^{jm} dt \right) \left(\int_{t_{i-1}}^{t_i} c_t^{kl} dt \right) \right)$$

が成り立つので,

$$\left| n\,\mathrm{Cov}\left[\widehat{[X^j, X^k]}_T, \widehat{[X^l, X^m]}_T\right] - \Sigma_{(j-1)d+k,(l-1)d+m} \right|$$
$$\leq \sum_{i=1}^{n} \left\{ \left(\int_{t_{i-1}}^{t_i} c_t^{jl} \left(n \int_{t_{i-1}}^{t_i} |c_s^{km} - c_t^{km}| ds \right) dt + \int_{t_{i-1}}^{t_i} |c_t^{jm}| \left(n \int_{t_{i-1}}^{t_i} |c_s^{kl} - c_t^{kl}| ds \right) dt \right) \right\}$$
$$\lesssim K^4 \left(\frac{T}{n} \right)^{\gamma} \leq \frac{(1 \vee T)^{\gamma} K^4}{n^{\gamma}} \lesssim \frac{K^4}{n^{\gamma}}$$

となる.

Step 2. 次に, (9.8) を示す. 上式の左辺は常に 1 以下なので, (9.58) に現れる定数 c_1 に対して,

$$c_1 \frac{K^4 \log^2 d}{v n^{\gamma}} \leq \frac{1}{2}$$

が成り立つと仮定して一般性を失わない. このとき, (9.58) より $\min_{j,k} \mathrm{Var}[\sqrt{n}\widehat{[X^j, X^k]}_T] \geq v/2$ が成り立つ. また, 命題 2.5 より $\max_{i,j,k} \|n \Delta_i^n X^k \Delta_i^n X^k\|_{\psi_1} \lesssim K^2$ が成り立つ. 従って, 定理 9.1 を $X_i = n\,\mathrm{vec}(\Delta_i^n X (\Delta_i^n X)^\top - \mathrm{E}[\Delta_i^n X (\Delta_i^n X)^\top]), B_n \lesssim K^2, b = \sqrt{v/2}$ として適用することができて,

$$\sup_{A \in \mathcal{R}_{d^2}} |\mathrm{P}(S_n \in A) - \mathrm{P}(\tilde{Z}_n \in A)| \lesssim \left(\frac{K^8 \log^5 (dn)}{v^2 n} \right)^{1/4}$$

を得る. ここに, $\tilde{Z}_n \sim \mathrm{N}(0, \mathrm{Cov}[S_n])$ である. 一方で, 定理 9.7 と (9.58) より,

$$\sup_{A \in \mathcal{R}_{d^2}} |\mathrm{P}(\tilde{Z}_n \in A) - \mathrm{P}(Z_n \in A)| \lesssim \sqrt{\frac{K^4 \log^2 d}{v n^{\gamma}}}$$

を得る. 以上をあわせて示すべき不等式を得る. □

9.7.2　命題 9.1 の証明

(9.58) と同様の議論によって,

$$\left| n \sum_{i=1}^{n} \mathrm{E}[\Delta_i^n X^j \Delta_i^n X^k \Delta_i^n X^l \Delta_i^n X^m] - \int_0^T \left(c_t^{jk} c_t^{lm} + c_t^{jl} c_t^{km} + c_t^{jm} c_t^{kl} \right) dt \right| \lesssim \frac{K^4}{n^{\gamma}},$$
$$\left| n \sum_{i=1}^{n-1} \mathrm{E}[\Delta_i^n X^j \Delta_i^n X^k \Delta_{i+1}^n X^l \Delta_{i+1}^n X^m] - \int_0^T c_t^{jk} c_t^{lm} dt \right| \lesssim \frac{K^4}{n^{\gamma}}$$

を得る. 従って,

$$
\begin{aligned}
&\left\|\left|n\sum_{i=1}^{n-1}\left(\Delta_i^n X^j\Delta_i^n X^k\Delta_i^n X^l\Delta_i^n X^m - \mathrm{E}[\Delta_i^n X^j\Delta_i^n X^k\Delta_i^n X^l\Delta_i^n X^m]\right)\right|\right\|_p \\
&\qquad\lesssim K^4\left(\sqrt{\frac{p}{n}}+\frac{p^3\log^2 n}{n}\right),
\end{aligned}
\tag{9.59}
$$

および

$$
\begin{aligned}
&\mathrm{E}\left[\left|n\sum_{i=1}^{n-1}\left(\Delta_i^n X^j\Delta_i^n X^k\Delta_{i+1}^n X^l\Delta_{i+1}^n X^m - \mathrm{E}[\Delta_i^n X^j\Delta_i^n X^k\Delta_{i+1}^n X^l\Delta_{i+1}^n X^m]\right)\right|\right] \\
&\qquad\lesssim K^4\left(\sqrt{\frac{p}{n}}+\frac{p^3\log^2 n}{n}\right)
\end{aligned}
\tag{9.60}
$$

が成り立つことを示せばよい.

ここでは (9.60) のみ示す. (9.59) も同様のより簡単な議論で証明できる.

$$
\xi_i^{jklm} := n\Delta_i^n X^j\Delta_i^n X^k\Delta_{i+1}^n X^l\Delta_{i+1}^n X^m
$$

とおく. $(\xi_{2i}^{jklm})_{i=1}^{\lfloor (n-1)/2\rfloor}$ は独立であるから,

$$
\begin{aligned}
&\left\|\left|\sum_{i=1}^{\lfloor (n-1)/2\rfloor}\left(\xi_{2i}^{jklm}-\mathrm{E}[\xi_{2i}^{jklm}]\right)\right|\right\|_p \\
&\lesssim \sqrt{p\sum_{i=1}^{\lfloor (n-1)/2\rfloor}\mathrm{E}[|\xi_{2i}^{jklm}|^2]}+p\sqrt{\left\|\max_{1\le i\le\lfloor (n-1)/2\rfloor}|\xi_{2i}^{jklm}|^2\right\|_p} \\
&\hspace{20em}(\because\text{補題 } 9.16) \\
&\lesssim \sqrt{p\frac{K^8}{n}}+p^3\sqrt{\frac{K^8\log^4 n}{n^2}}\log d \quad (\because \text{命題 } 2.2, 2.4)
\end{aligned}
$$

を得る. 同様の議論によって,

$$
\left\|\left|\sum_{i=1}^{\lfloor n/2\rfloor}\left(\xi_{2i-1}^{jklm}-\mathrm{E}[\xi_{2i-1}^{jklm}]\right)\right|\right\|_p \lesssim \sqrt{p\frac{K^8}{n}}+p^3\sqrt{\frac{K^8\log^4 n}{n^2}}
$$

も得られるので, 三角不等式より (9.60) が従う. □

9.7.3 定理 9.5 の証明

補題 9.25. ある普遍定数 $C\ge 1$ が存在して, 定理 9.5 の仮定の下で, 事象

$$
|\widehat{\Sigma}_n-\Sigma|_\infty \le CK^4\left(\sqrt{\frac{\log(dn)}{n}}+\frac{1}{n^\gamma}\right)
\tag{9.61}
$$

が $1-1/n^4$ 以上の確率で成り立つ.

証明. (9.10) の下で, $p = 4\log(dn)$ に対して,

$$\frac{p^3 \log^2 n}{n} = 64\sqrt{\frac{\log(dn)}{n}}\sqrt{\frac{\log^5(dn)\log^4 n}{n}} \leq 64\sqrt{\frac{\log(dn)}{n}}$$

となるから, 命題 9.1 より, ある普遍定数 $C_1 > 0$ が存在して

$$\max_{j,k,l,m} \left\| \widehat{\Sigma}_{(j-1)d+k,(l-1)d+m} - \Sigma_{(j-1)d+k,(l-1)d+m} \right\|_p \leq C_1 K^4 \left(\sqrt{\frac{\log(dn)}{n}} + \frac{1}{n^\gamma} \right)$$

が成り立つ. 従って, $C = 1 \vee eC_1$ とすれば, 補題 9.19 の証明と同様の議論によって示すべき主張を得る. □

定理 9.5 の証明. 補題 9.25 に注意すれば, 定理 9.2–9.3 の証明とパラレルな議論で証明できる. 詳細は省略する. □

参考文献

[1] Aït-Sahalia, Yacine & Jean Jacod, High-frequency financial econometrics.: Princeton University Press, 2014.

[2] Aït-Sahalia, Yacinei & Dacheng Xiu, 'Using principal component analysis to estimate a high dimensional factor model with high-frequency data.' J. Econometrics, **201**, pp. 384–399, 2017.

[3] Barlow, M. T. & M. Yor, 'Semi-martingale inequalities via the Garsia-Rodemich-Rumsey lemma and application to local times.' J. Funct. Anal., **49**, pp. 198–229, 1982.

[4] Barndorff-Nielsen, Ole E. & Neil Shephard, 'Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics.' Econometrica, **72** (3), pp. 885–925, 2004.

[5] Belloni, Alexandre, Victor Chernozhukov, Denis Chetverikov, Christian Hansen, & Kengo Kato, 'High-dimensional econometrics and regularized GMM.' 2018, Working paper. arXiv: 1806.01888.

[6] Birnbaum, Z. W., 'An Inequality for Mill's Ratio.' Ann. Math. Statist., **13** (2), pp. 245–246, 1942.

[7] Bonis, Thomas, 'Stein's method for normal approximation in Wasserstein distances with application to the multivariate Central Limit Theorem.' Probab. Theory Related Fields, **178**, pp. 827–860, 2020.

[8] Boucheron, Stéphane, G'abor Lugosi, & Pascal Massart, Concentration inequalities: A nonasymptotic theory of independence.: Oxford University Press, 2013.

[9] Bühlmann, Peter & Sara van de Geer, Statistics for High-Dimensional Data.: Springer, 2011.

[10] Burkholder, D. L., 'Distribution Function Inequalities for Martingales.' Ann. Probab., **1** (1), pp. 19–42, 1973.

[11] Carlen, Eric & Paul Kree, 'L^p Estimates on Iterated Stochastic Integrals.' Ann. Probab., **19** (1), pp. 354–368, 1991.

[12] Chatterjee, Sourav, Superconcentration and Related Topics.: Springer, 2014.

[13] Chen, Louis H.Y., Qi-Man Shao, & Larry Goldstein, Normal Approximation by Stein's Method.: Springer, 2011.

[14] Chernozhukov, Victor, Denis Chetverikov, & Kengo Kato, 'Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors.' Ann. Statist., **41** (6), pp. 2786–2819, 2013.

[15] ——— , 'Comparison and anti-concentration bounds for maxima of Gaussian random vectors.' Probab. Theory Related Fields, **162**, pp. 47–70, 2015.

[16] ——— , 'Central Limit Theorems and Bootstrap in High Dimensions.' Ann. Probab., **45** (4), pp. 2309–2353, 2017.

[17] Chernozhukov, Victor, Denis Chetverikov, Kengo Kato, & Yuta Koike, 'Improved central limit

theorem and bootstrap approximation in high dimensions.' Ann. Statist., **50** (5), pp. 2562–2586, 2022.

[18] ——— , 'High-dimensional data bootstrap.' Annu. Rev. Stat. Appl., **10**, pp. 427–449, 2023.

[19] Davis, Burgess, 'On the L^p norms of stochastic integrals and other martingales.' Duke Math. J., **43** (4), pp. 697–704, 1976.

[20] Deng, Hang & Cun-Hui Zhang, 'Beyond Gaussian Approximation: Bootstrap for Maxima of Sums of Independent Random Vectors.' Ann. Statist., **48** (6), pp. 3643–3671, 2020.

[21] Duchi, John, Stephen Gould, & Daphne Koller, 'Projected Subgradient Methods for Learning Sparse Gaussians.' In Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence., pp. 153–160: AUAI Press, 2008.

[22] Fan, Jianqing, Alex Furger, & Dacheng Xiu, 'Incorporating Global Industrial Classification Standard Into Portfolio Allocation: A Simple Factor-Based Large Covariance Matrix Estimator With High-Frequency Data.' J. Bus. Econom. Statist., **34** (4), pp. 489–503, 2016.

[23] Fang, Xiao & Yuta Koike, 'High-dimensional central limit theorems by Stein's method.' Ann. Appl. Probab., **31** (4), pp. 1660–1686, 2021.

[24] ——— , 'New error bounds in multivariate normal approximations via exchangeable pairs with applications to Wishart matrices and fourth moment theorems.' Ann. Appl. Probab., **32** (1), pp. 602–631, 2022.

[25] van Handel, Ramon, 'Structured Random Matrices.' In Carlen, Eric, Mokshay Madiman, & Elisabeth M. Werner (eds.) Convexity and Concentration.: Springer, 2017.

[26] Horn, Roger A. & Charles R. Johnson, Matrix analysis.: Cambridge University Press, 2nd edition, 2013.

[27] Hounyo, Ulrich, 'Bootstrapping integrated covariance matrix estimators in noisy jump-diffusion models with non-synchronous trading.' J. Econometrics, **197**, pp. 130–152, 2017.

[28] Ikeda, Nobuyuki & Shinzo Watanabe, Stochastic differential equations and diffusion processes.: North-Holland, 2nd edition, 1989.

[29] Jacod, Jean & Philip Protter, Discretization of Processes.: Springer, 2012.

[30] Jacod, Jean & Albert N. Shiryaev, Limit theorems for stochastic processes.: Springer, 2nd edition, 2003.

[31] Janková, Jana & Sara van de Geer, 'Inference in high-dimensional graphical models.' In Handbook of Graphical Models.: CRC Press, Chapter 14, pp. 325–351, 2018.

[32] Karatzas, Ioannis & Steven E. Shreve, Brownian motion and stochastic calculus.: Springer, 2nd edition, 1998.

[33] 風巻 紀彦, 『マルチンゲールと確率積分』, 科学技術出版, 2003.

[34] Klaassen, Sven, Jannis Kück, Martin Spindler, & Victor Chernozhukov, 'Uniform Inference in High-Dimensional Gaussian Graphical Models.' Biometrika, **110** (1), pp. 51–68, 2023.

[35] Koike, Yuta, 'Mixed-normal limit theorems for multiple Skorohod integrals in high-

dimensions, with application to realized covariance.' Electron. J. Stat., **13** (1), pp. 1443–1522, 2019.

[36] Kuchibhotla, Arun Kumar & Abhishek Chakrabortty, 'Moving Beyond Sub-Gaussianity in High Dimensional Statistics: Applications in Covariance Estimation and Linear Regression.' Inf. Inference, **11**, pp. 1389–1456, 2022.

[37] Le Cam, Lucien, Asymptotic Methods in Statistical Decision Theory.: Springer, 1986.

[38] Ledoux, Michel & Michel Talagrand, Probability in Banach Spaces.: Springer, 1991.

[39] Magnus, Jan R. & Heinz Neudecker, Matrix differential calculus with applications in statistics and econometrics.: Wiley, 3rd edition, 2019.

[40] 長井 英生,『確率微分方程式』, 共立出版, 1999.

[41] Nourdin, Ivan & Giovanni Peccati, Normal Approximations with Malliavin Calculus: From Stein's Method to Universality.: Cambridge University Press, 2012.

[42] Pelger, Markus, 'Large-dimensional factor modeling based on high-frequency observations.' J. Econometrics, **208**, pp. 23–42, 2019.

[43] Revuz, Daniel & Marc Yor, Continuous martingales and Brownian motion.: Springer, 3rd edition, 1999.

[44] Rothman, Adam J., Peter J. Bickel, Elizaveta Levina, & Ji Zhu, 'Sparse permutation invariant covariance estimation.' Electron. J. Stat., **2**, pp. 494–515, 2008.

[45] Rudin, Walter, Real and complex analysis.: McGraw-Hill, 3rd edition, 1987.

[46] 高木 貞治,『解析概論 (改訂第三版)』, 岩波書店, 1983.

[47] Tanguy, Kevin, 'Some superconcentration inequalities for extrema of stationary Gaussian processes.' Statist. Probab. Lett., **106**, pp. 239–246, 2015.

[48] van der Vaart, Aad W. & Jon A. Wellner, Weak Convergence and Empirical Processes.: Springer, 1996.

[49] Vershynin, Roman, High-Dimensional Probability.: Cambridge University Press, 2018.

[50] Wainwright, Martin J., High-Dimensional Statistics.: Cambridge University Press, 2019.

[51] Wasserman, Larry, 'Stein's Method and The Bootstrap in Low and High Dimensions: A Tutorial.' 2014, Available at `http://www.cs.cmu.edu/~aarti/SMLRG/Stein.pdf`.