第2章　劣 Gauss 型確率変数列と最大不等式

1　正規分布の裾確率と積率母関数

定義 2.1 実数値確率変数 $X \in \mathbb{R}$ は正規分布 (Gauss 分布) に従うとは, X が測度空間 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の Lebesgue 測度 ($\mathbb{R}$ 上の Lebeague 測度) に関して次の確率密度関数 p を持つときをいう.

$$p(x) = \frac{1}{\sqrt{2\pi^2\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (x \in \mathbb{R}).$$

ただし, $\mu = \mathbb{E}[X]$, $\sigma^2 = \mathbb{V}[X]\,(\sigma > 0)$ である. これを $X \sim N(\mu, \sigma^2)$ と記す.

注意 2.2 Z と確率変数とする. $\mathbb{E}[Z] = 0$ と $\mathbb{V}[Z] = 1$ となる正規分布 $\mathsf{N}(0, 1)$ を標準正規分布という. $X = \sigma Z + \mu$ とおけば, $X \sim \mathsf{N}(\mu, \sigma^2)$ となる. 明らかに, 分布の台 $\{x \in \mathbb{R}; p(x) > 0\}$ は有界ではない. しかし, 次の意味で p はほとんど有界な台を持つと考えてよい.

$$\mathbb{P}\big(|X-\mu| \le 3\sigma\big) \simeq 0.997.$$

これは, $|x| \to \infty$ のとき, p の裾が急激に減少することを意味している.

命題 2.3 (Milll の不等式) $X \sim \mathsf{N}(\mu, \sigma^2)$ とする. このとき, $\forall t > 0$ に対して,

$$\mathbb{P}\big(X - \mu > t\big) \le \frac{\sigma}{t\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

対称性より

$$\mathbb{P}\big(X - \mu < -t\big) \le \frac{\sigma}{t\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

と

$$\mathbb{P}\big(|X - \mu| > t\big) \le \frac{\sigma}{t}\sqrt{\frac{2}{\sigma}} \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

証明　一般性を失うことなく, $\mu = 0$ と $\sigma = 1$ として示せばよいことに注意せよ. $Z \sim \mathsf{N}(0, 1)$ に対して,

$$\begin{aligned}
\mathbb{P}(Z > t) &= \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp\left(-\frac{z^2}{2}\right) dz \\
&\leq \frac{1}{\sqrt{2\pi}} \int_t^\infty \frac{z}{t} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \frac{1}{t\sqrt{2\pi}} \int_t^\infty \left\{-\frac{\partial}{\partial z} \exp\left(-\frac{z^2}{2}\right)\right\} dz \\
&= \frac{1}{t\sqrt{2\pi}} \left(-\frac{t^2}{2}\right).
\end{aligned}$$

2 番目の不等式は, p が偶関数であることからわかる. 3 番目の不等式は

$$\begin{aligned}
\mathbb{P}(|Z| > t) &= \mathbb{P}\big(\{Z > t\} \cup \{Z < -t\}\big) \\
&\leq \mathbb{P}(Z > t) + \mathbb{P}(Z < -t) \qquad (\because \text{ユニオン・バウンド}) \\
&= 2\mathbb{P}(Z > t). \qquad (\because \text{標準正規分布の確率密度関数の対称性})
\end{aligned}$$

□

正規分布に従う確率変数 Z の裾確率は指数関数的に 0 に近づくこと[1]は, 積率母関数 (MGF)

$$M : s \mapsto M(s) := \mathbb{E}[\exp\{sZ\}]$$

を通してもわかる.

まず, Z の積率母関数を求めておく. $s \in \mathbb{R}$ に対して,

$$\begin{aligned}
M_Z(s) &:= \mathbb{E}[\exp\{sZ\}] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp\{sz\} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{-\frac{(z-s)^2}{2} + \frac{s^2}{2}\right\} dz \\
&= \exp\left(\frac{s^2}{2}\right). \qquad \left(\because \tfrac{1}{\sqrt{2\pi}} \textstyle\int_{-\infty}^\infty \exp\left\{-\tfrac{(z-s)^2}{2}\right\} dz = 1\right)
\end{aligned}$$

[1]なぜ?

したがって, 一般の正規分布に対しては, $X \sim N(\mu, \sigma^2)$ する. $s \in \mathbb{R}$ に対して,

$$
\begin{aligned}
M_X(s) &:= \mathbb{E}\big[\exp\{sX\}\big] = \mathbb{E}\big[\exp\big\{s\sigma Z + s\mu\big\}\big] = \exp\{s\mu\}M_Z(s\sigma) \\
&= \exp\left\{\mu s + \frac{\sigma^2}{2}s^2\right\}.
\end{aligned}
$$

2 劣 Gauss 型確率変数列とその裾確率評価

独立な確率変数の標本平均の裾確率を制御してみよう. $X_1, X_2, \ldots, X_n \sim$ i.i.d. $\mathsf{N}(\mu, \sigma^2)$ とし,

$$
\overline{X} := \frac{1}{n}\sum_{j=1}^{n} X_j \sim \mathsf{N}\left(\mu, \frac{\sigma^2}{n}\right)
$$

を考える. 以下で求める補題 2.10 を用いると $\forall t > 0$ に対して,

$$
\mathbb{P}\big(|\overline{X} - \mu| > t\big) \leq 2\exp\left(-\frac{nt^2}{2\sigma^2}\right) \tag{2.1}
$$

となる. すると, 小さな $\delta > 0$ に対して,

$$
\mathbb{P}\left(\overline{X} - \sigma\sqrt{\frac{2\log(2/\delta)}{n}} \leq \mu \leq \overline{X} + \sigma\sqrt{\frac{2\log(2/\delta)}{n}}\right) > 1 - \delta \tag{2.2}
$$

を得る.

なぜならば,

$$
\begin{aligned}
&2\exp\left(-\frac{nt^2}{2\sigma^2}\right) = \delta \iff -\frac{nt^2}{2\sigma^2} = \log\left(\frac{\delta}{2}\right) \iff t^2 = \frac{2\sigma^2}{n}\log\left(\frac{2}{\delta}\right) \\
&\iff t = \pm\sigma\sqrt{\frac{2\log(2/\delta)}{n}}
\end{aligned}
$$

を得る. これを (2.1) に代入すれば,

$$
\mathbb{P}\left(\big|\overline{X} - \mu\big| > \sigma\sqrt{\frac{2\log(2/\delta)}{n}}\right) \leq \delta
$$

となる. これより

$$
\mathbb{P}\left(\big|\overline{X} - \mu\big| \leq \sigma\sqrt{\frac{2\log(2/\delta)}{n}}\right) > 1 - \delta \iff (2.2)
$$

となる.

以下で定義をする劣 Gauss 型確率変数に対しても同じような信頼区間を構成することができる.

定義 2.4 確率変数 X が代行分散 (variance proxy)σ^2 $(\sigma > 0)$ を持つ劣 Gauss 型確率変数であるとは, 次の (1), (2) をみたすときをいう.
(1) $\mathbb{E}[X] = 0$.
(2) X の積率母関数が以下をみたす. $\forall s \in \mathbb{R}$ に対して,

$$\mathbb{E}\big[\exp\{sX\}\big] \leq \exp\left(\frac{\sigma^2}{2}s^2\right).$$

これを $X \sim \mathsf{subG}(\sigma^2)$ と記す.

注意 2.5 $\mathsf{subG}(\sigma^2)$ は分布よりも分布族を表す記号である. したがって, $X \sim \mathsf{subG}(\sigma^2)$ は記号の乱用であることに注意せよ.

定義 2.6 確率ベクトル $\boldsymbol{X} \in \mathbb{R}^d$ が代行分散 σ^2 を持つ劣 Gauss 型分布であるとは, $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{0}$ で, 任意の $\boldsymbol{u} \in \mathbb{S}^{d-1}$ に対して, $\boldsymbol{u}^\top \boldsymbol{X} \sim \mathsf{subG}(\sigma^2)$ となるときをいう. この場合に, $\boldsymbol{X} \sim \mathsf{subG}_d(\sigma^2)$ と記す. ただし,

$$\mathbb{S}^{d-1} := \{\boldsymbol{x} \in \mathbb{R}^d; |\boldsymbol{x}|_2 = 1\}, \qquad |\boldsymbol{x}|_2 = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$$

である.

注意 2.7 $\boldsymbol{X} \sim \mathsf{subG}_d(\sigma^2)$ のとき, $\forall \boldsymbol{v} \in \mathbb{R}^d$ で, $|\boldsymbol{v}|_2 \leq 1$ なるものに対して, $\boldsymbol{v}^\top \boldsymbol{X} \sim \mathsf{subG}(\sigma^2)$ である.

実際, $\boldsymbol{u} = \boldsymbol{v}/|\boldsymbol{v}|_2 \in \mathbb{S}^{d-1}$ に対して,

$$\begin{aligned}
\mathbb{E}[\exp(s\boldsymbol{v}^\top \boldsymbol{X}\})] &= \mathbb{E}[\exp(s|\boldsymbol{v}|_2 \boldsymbol{u}^\top \boldsymbol{X})] \\
&\leq \exp\left(\frac{\sigma^2 |\boldsymbol{v}|_2^2}{2} s^2\right) \\
&\leq \exp\left(\frac{\sigma^2}{2} s^2\right)
\end{aligned}$$

よりわかる. □

定義 2.8 ランダム行列 $\boldsymbol{X} \in \mathbb{R}^{d\times T}$ が代行分散 σ^2 を持つ劣 Gauss 型分布であるとは, $\forall \boldsymbol{u} \in \mathbb{S}^{d-1}$, $\boldsymbol{v} \in \mathbb{S}^{T-1}$ に対して, $\boldsymbol{u}^\top \boldsymbol{X}\boldsymbol{v} \sim \mathsf{subG}(\sigma^2)$ のときをいう.

補題 2.9 (Markov 不等式) X を確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の非負値確率変数で $\mathbb{P}$ 可積分とする. このとき, $\forall a > 0$ に対して,

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a}$$

が成り立つ.

証明 $A = (a, \infty)$ とする. このとき,

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X\mathbb{1}_A(X)] + \mathbb{E}[X\mathbb{1}_{A^c}(X)] \\ &\geq \mathbb{E}[X\mathbb{1}_A(X)] \geq a\mathbb{E}[\mathbb{1}_A(X)] = a\mathbb{P}(X > a).\end{aligned}$$

よって,

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}[X]}{a}$$

を得る. □

補題 2.10 $X \sim \mathsf{subG}(\sigma^2)$ とする. このとき, $\forall t > 0$ に対して,

$$\mathbb{P}(X > t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right); \qquad \mathbb{P}(X < -t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

証明 Chernoff 限界と呼ばれるテクニックを用いる. $\forall s > 0$ に対して, Markov の不等式と $X \sim \mathsf{subG}(\sigma^2)$ であることより,

$$\mathbb{P}(X > t) = \mathbb{P}\big(e^{sX} > e^{st}\big) \leq \frac{\mathbb{E}[e^{sX}]}{e^{st}} \leq \exp\left(\frac{\sigma^2}{2}s^2 - st\right). \tag{2.3}$$

ここで, $\phi(s) := (\sigma^2/2)s^2 - st\,(s > 0)$ とおくと

$$\phi(s) \geq \inf_{s>0} \phi(s) = -\frac{t^2}{2\sigma^2}$$

となるので, 不等式 (2.3) は任意の $s > 0$ で成立しているので,

$$\mathbb{P}(X > t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

を得る. $\square$

次に積率の評価について述べる. $Z \sim \mathsf{N}(0, \sigma^2)\,(\sigma > 0)$ の絶対値の積率は次で与えられる. $k \in \mathbb{N}$ に対して,

$$\mathbb{E}[|Z|^k] = \frac{1}{\sqrt{2\pi}}(2\sigma^2)^{k/2}\Gamma\left(\frac{k+1}{2}\right).$$

ただし, $\Gamma(\,\cdot\,)$ はガンマ関数で

$$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dt \qquad (t > 0).$$

次の補題は, 補題 2.10 の裾確率の評価で, $X \sim \mathsf{subG}(\sigma^2)$ の絶対値の積率は $Z \sim N(0, \sigma^2)$ の絶対値の積率の定数倍で評価できることを示している.

補題 2.11 確率変数 X は

$$\mathbb{P}\big(|X| > t\big) \le 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$$

をみたすとする. このとき, $\forall k \ge 1\,(k \in \mathbb{N})$ に対して,

$$\mathbb{E}\big[|X|^k\big] \le (2\sigma^2)^{k/2}k\Gamma\left(\frac{k}{2}\right).$$

特に,

$$\big(\mathbb{E}\big[|X|^k\big]\big)^{1/k} \le \sigma e^{1/e}\sqrt{k}, \qquad (k \ge 2)$$

と

$$\mathbb{E}\big[|X|\big] \le \sigma\sqrt{2\pi}$$

である.

証明 命題 2.42 と補題 2.10 より, $t > 0$ に対して,

$$\begin{aligned}
\mathbb{E}\big[|X|^k\big] &= \int_0^\infty \mathbb{P}(|X| > t^{1/k})\,dt \qquad (\because \text{命題 } 2.42)\\
&\le 2\int_0^\infty \exp\left(-\frac{t^{2/k}}{2\sigma^2}\right)dt \qquad (\because \text{命題 } 2.10)\\
&= (2\sigma^2)^{k/2}k\int_0^\infty e^{-u}u^{(k/2)-1}\,du \qquad \left(\because u = \tfrac{t^{2/k}}{2\sigma^2} \text{ と変換}\right)\\
&= (2\sigma^2)^{k/2}k\Gamma\left(\frac{k}{2}\right).
\end{aligned}$$

2 番目の主張は, $\forall k \geq 2$ に対して,

$$\Gamma\left(\frac{k}{2}\right) \leq \left(\frac{k}{2}\right)^{k/2}; \qquad k^{1/k} \leq e^{1/e}$$

を用いると

$$\left\{(2\sigma^2)^{k/2} k\Gamma\left(\frac{k}{2}\right)\right\}^{1/k} \leq k^{1/k}\sqrt{\frac{2\sigma^2 k}{2}} \leq e^{1/e}\sigma\sqrt{k}.$$

さらに, $k = 1$ に対しては

$$\sqrt{2}\Gamma\left(\frac{1}{2}\right) = \sqrt{2\pi}$$

よりわかる. □

補題 2.12 X を確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の確率変数で以下をみたすものとする.
(1) $\mathbb{E}[X] = 0$.
(2) $\forall t > 0$ に対して,

$$\mathbb{P}(X > t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right); \qquad \mathbb{P}(X < -t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

ただし, $\sigma > 0$. このとき, $\forall s > 0$ に対して,

$$\mathbb{E}\left[e^{sX}\right] \leq e^{4\sigma^2 s^2}.$$

証明　指数関数に対する Taylor 展開を用いる.

$$
\begin{aligned}
\mathbb{E}[e^{sX}] &\leq 1+\sum_{k=0}^{\infty}\frac{s^k\mathbb{E}[|X|^k]}{k!} \\
&\leq 1+\sum_{k=2}^{\infty}\frac{(2\sigma^2s^2)^{k/2}k\Gamma(k/2)}{k!} \qquad (\because \text{補題 } 2.11) \\
&= 1+\sum_{k=1}^{\infty}\frac{(2\sigma^2s^2)^{k}(2k)\Gamma(k)}{(2k)!}+\sum_{k=1}^{\infty}\frac{(2\sigma^2s^2)^{k+1/2}(2k+1)\Gamma(k+1/2)}{(2k+1)!} \\
&\leq 1+\left(2+\sqrt{2\sigma^2s^2}\right)\sum_{k=1}^{\infty}\frac{(2\sigma^2s^2)^{k}k!}{(2k)!} \\
&\qquad \left(\because 2(k!)^2 \leq (2k)!\right) \\
&= 1+\left(1+\sqrt{\frac{\sigma^2s^2}{2}}\right)\left(e^{2\sigma^2s^2}-1\right) \\
&= e^{2\sigma^2s^2}+\sqrt{\frac{\sigma^2s^2}{2}}\left\{e^{2\sigma^2s^2}-1\right\} \\
&\leq e^{4\sigma^2s^2}.
\end{aligned}
$$

□

注意 2.13 最後の不等式について. $x>0$ に対して,

$$
e^{2x} \geq e^x+\sqrt{\frac{x}{4}}(e^x-1) \iff e^x \geq 1+\sqrt{\frac{x}{4}}(1-e^{-x}) \iff e^x+e^{-x}\sqrt{\frac{x}{4}} \geq 1+\sqrt{\frac{x}{4}}.
$$

最後の不等式は自明. □

2.1　劣 Gauss 確率変数列の和

$n\in\mathbb{N}$ とし, $X_1, X_2, \ldots, X_n$ は $\mathsf{N}(0, \sigma^2)$ $(\sigma>0)$ に独立同一に従う確率変数列とする. このとき, $\boldsymbol{a}=(a_1, a_2, \ldots, a_n)^\top\in\mathbb{R}^n$ に対して,

$$
\sum_{j=1}^{n}a_jX_j \sim N(0, |\boldsymbol{a}|_2^2\sigma^2)
$$

である. ただし, $|\boldsymbol{a}|_2=\sqrt{\boldsymbol{a}^\top\boldsymbol{a}}$ である. 分布の裾確率だけに注目すれば, 劣 Gauss 型確率変数も同じような性質を持つ.

命題 2.14 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ は独立で, 各 $X_j\,(j = 1, 2, \ldots, n)$ は劣 Gauss 型確率変数列で, 代行分散 $\sigma^2\,(\sigma > 0)$ を持つとする. このとき, 確率ベクトル $\boldsymbol{X}$ は代行分散 σ^2 の劣 Gauss である.

証明 $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{S}^{n-1}$ に対して,

$$
\begin{aligned}
\mathbb{E}[\exp\{s\boldsymbol{u}^\top \boldsymbol{X}\}] &= \prod_{j=1}^n \mathbb{E}[e^{su_jX_j}] \le \prod_{j=1}^n \exp\left(\frac{\sigma^2 s^2 u_j^2}{2}\right) \\
&= \exp\left(\frac{1}{2}\sigma^2 s^2 |\boldsymbol{u}|_2^2\right) = \exp\left(\frac{1}{2}s^2\sigma^2\right).
\end{aligned}
$$

□

Chernoff 限界を用いると, 次の系を得る.

系 2.15 $X_1, X_2, \ldots, X_n$ は独立な確率変数列で, $X_j \sim \mathsf{subG}(\sigma^2)\,(j = 1, 2, \ldots, n)$ とする. このとき, $\forall \boldsymbol{a} = (a_1, a_2, \ldots, a_n)^\top \in \mathbb{R}^n$ と $t > 0$ に対して,

$$
\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^n a_j X_j > t\right) &\le \exp\left(-\frac{t^2}{2\sigma^2|\boldsymbol{a}|_2^2}\right), \\
\mathbb{P}\left(\sum_{j=1}^n a_j X_j < -t\right) &\le \exp\left(-\frac{t^2}{2\sigma^2|\boldsymbol{a}|_2^2}\right)
\end{aligned}
$$

が成立する. 特に, $a_j = 1/n\,(j = 1, 2, \ldots, n)$ とすれば,

$$
\begin{aligned}
\mathbb{P}(\overline{X} > t) &\le \exp\left(-\frac{nt^2}{2\sigma^2}\right), \\
\mathbb{P}(\overline{X} < -t) &\le \exp\left(-\frac{nt^2}{2\sigma^2}\right)
\end{aligned}
$$

を得る. これは正規分布と同じである.

証明 命題 2.14 を用いる. $\forall s > 0$ に対して,

$$
\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^n a_j X_j > t\right) &= \mathbb{P}\left(\exp\left(s\sum_{j=1}^n a_j X_j\right) > e^{st}\right) \\
&\le \exp\left(\frac{\sigma^2|\boldsymbol{a}|_2^2}{2}s^2 - st\right).
\end{aligned}
$$

よって,

$$\mathbb{P}\left(\sum_{j=1}^{n} a_j X_j > t\right) = \inf_{s>0}\left\{\exp\left(\frac{\sigma^2 |\boldsymbol{a}|_2^2}{2}s^2 - st\right)\right\} = \exp\left(-\frac{t^2}{2\sigma^2|\boldsymbol{a}|_2^2}\right).$$

残りの 3 つの不等式も同様に示すことができる. □

補題 2.16 (Hoeffding の補題) X を確率変数で $\mathbb{E}[X]=0$ とし, ほとんど確実に $X \in [a, b]$ とする. ただし, $a < b$ である. このとき, $\forall s \in \mathbb{R}$ に対して,

$$\mathbb{E}[e^{sX}] \le \exp\left(\frac{s^2}{8}(b-a)^2\right)$$

が成立する. 特に, $X \sim \mathsf{subG}\left(\frac{(b-a)^2}{4}\right)$ である.

証明 $\psi(s) := \log \mathbb{E}\left[e^{sX}\right]$ $(sred \in \mathbb{R}black)$ とする. このとき,

$$\dot{\psi}(s) := \frac{d}{ds}\psi(s) = \frac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}[e^{sX}]}; \qquad \ddot{\psi}(s) := \frac{d^2}{ds^2}\psi(s) = \frac{\mathbb{E}[X^2 e^{sX}]}{\mathbb{E}[e^{sX}]} - \left\{\frac{\mathbb{E}[Xe^{sX}]}{\mathbb{E}[e^{sX}]}\right\}^2.$$

しかし, ほとんど確実に $X \in [a, b]$ なので,

$$\mathbb{V}[X] \;=\; \mathbb{V}\left[X - \frac{a+b}{2}\right] \le \mathbb{E}\left[\left(X - \frac{a+b}{2}\right)^2\right] \le \frac{(b-a)^2}{4}$$

となる. このこととキュムラント関数の性質から

$$\ddot{\psi}(s) \le \frac{(b-a)^2}{4}$$

である. ここで, 微積分学の基本定理, $\psi(0) = \log 1 = 0$ および $\psi'(0) = \mathbb{E}[X] = 0$ に注意すれば

$$\begin{aligned}\psi(s) \;&=\; \int_0^s \int_0^u \ddot{\psi}(r)\,dr\,du \le \int_0^s \int_0^u \frac{(b-a)^2}{4}\,dr\,du = \int_0^s \frac{(b-a)^2}{4}u\,du \\ &=\; \frac{(b-a)^2}{8}s^2\end{aligned}$$

を得る. □

命題 2.17 (Hoeffding の不等式) $X_1, X_2, \ldots, X_n$ は独立な確率変数列で, ほとんど確実に $X_j \in [a_j, b_j]\,(a_j < b_j;\, j = 1, 2, \ldots, n)$ とする.

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

としたとき, $\forall t > 0$ に対して,

$$\begin{aligned}\mathbb{P}\big(\overline{X} - \mathbb{E}[\overline{X}] > t\big) &\leq \exp\bigg(-\frac{2n^2t^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\bigg),\\ \mathbb{P}\big(\overline{X} - \mathbb{E}[\overline{X}] < -t\big) &\leq \exp\bigg(-\frac{2n^2t^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\bigg)\end{aligned}$$

を得る.

証明

$Y_j = X_j - \mathbb{E}[X_j]\,(j = 1, 2, \ldots, n)$ とおく. すると $Y_j \in \big[a_j - \mathbb{E}[X_j], b_j - \mathbb{E}[X_j]\big]$ かつ $\mathbb{E}[Y_j] = 0$ となる. ここで $Y_j\,(j = 1, 2, \ldots, n)$ に対して補題 2.16 を適用すると

$$\mathbb{E}\big[e^{sY_j}\big] \leq \exp\bigg(\frac{s^2}{8}(b_j - a_j)^2\bigg) \tag{2.4}$$

を得る. よって

$$\begin{aligned}\mathbb{P}\big(\overline{X} - \mathbb{E}[\overline{X}] > t\big) &= \mathbb{P}\bigg(\sum_{j=1}^{n} Y_j > nt\bigg) = \mathbb{P}\bigg(\exp\bigg(s\sum_{j=1}^{n} Y_j\bigg) > e^{nst}\bigg)\\ &\leq \mathbb{E}\bigg[\exp\bigg(s\sum_{j=1}^{n} Y_j\bigg)\bigg]e^{-nst} \qquad (\because \text{Markov の不等式})\\ &= \Pi_{j=1}^{n}\mathbb{E}[\exp\big(sY_j\big)]e^{-nst}\\ &\leq \Pi_{j=1}^{n}\exp\bigg(\frac{s^2}{8}(b_j - a_j)^2\bigg)e^{-nst} \qquad (\because (2.4))\\ &= \exp\bigg(\frac{s^2}{8}\sum_{j=1}^{n}(b_j - a_j)^2 - nst\bigg)\\ &= \exp\bigg\{\frac{\sum_{j=1}^{n}(b_j - a_j)^2}{8}\bigg(s - \frac{4nt}{\sum_{j=1}^{n}(b_j - a_j)^2}\bigg)^2 - \frac{2n^2t^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\bigg\}\end{aligned}$$

を得る．ここで

$$s = \frac{4nt}{\sum_{j=1}^{n}(b_j - a_j)^2}$$

とおくと

$$\mathbb{P}\big(\overline{X} - \mathbb{E}[\overline{X}] > t\big) \le \exp\left(-\frac{2n^2t^2}{\sum_{j=1}^{n}(b_j - a_j)^2}\right)$$

を得る． □

例 2.18 X は Rademacher 確率変数とする．すなわち，

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}.$$

このとき，$\forall s \in \mathbb{R}$ に対して，

$$\mathbb{E}\big[e^{sX}\big] = \frac{e^s + e^{-s}}{2} = \cosh(s) \le \exp\left(\frac{s^2}{2}\right).$$

さらに，上式の「2」は最良であることがわかる．実際，$a = -1$, $b = 1$, $\mathbb{E}[X] = 0$ なので，Hoeffding の補題より

$$\mathbb{E}\big[e^{sX}\big] \le \exp\left(\frac{s^2}{2}\right)$$

である． □

注意 2.19 Hoeffding の不等式は一般的なものである．しかし，一般性を得るための代償を払っている．確率変数の分散が非常に小さいとき，指数関数の裾確率の限界も小さくなるであろうが，Hoeffding の不等式は分布の台にのみ依存するので，小さな分散の影響を受けない．このことから，裾確率の評価について，更なる精緻化が必要になることがわかる．

□

3　劣指数型確率変数

中心化された確率変数が劣 Gauss 型分布ではないとき, 何がいえるだろうか? 典型的な例は, 母数 1 の両側指数分布 (Laplace) 分布[2] である. この分布を Lap(1) と記す. $X \sim \mathsf{Lap}(1)$ としたとき, $t \geq 0$ に対して,

$$\mathbb{P}(|X| > t) \quad = \quad 1 - \mathbb{P}(|X| \leq t) = 1 - \int_{-t}^{t} \frac{e^{-|s|}}{2}\, ds = 1 - \int_{0}^{t} e^{-|s|} = e^{-t}.$$

この分布の裾確率は正規分布のように急激に減少しない. このような裾は正規分布より重いという.

確率変数 X の分布の裾確率の挙動は X の積率母関数でとらえることができる. 実際,

$$\mathbb{E}[e^{sX}] = \frac{1}{1-s^2} \qquad (|s| < 1)$$

である. 劣ガウス型分布に対して示した限界のいくつかを再現するために, 積率母関数に対するある種の弱い制約だけが必要である. $X \sim \mathsf{Lap}(1)$ に対して,

$$\mathbb{E}[e^{sX}] \leq e^{2s^2} \qquad \left(|s| < \frac{1}{2}\right)$$

である. Laplace 分布の積率母関数は, 原点の近傍では, 正規分布の積率母関数でおさえられる. Laplace 分布と同じぐらいの重さの裾を持つ分布分布は同様の性質を持つことがわかる.

補題 2.20 X は中心化された確率変数で, ある $\lambda > 0$ が存在して, $\forall t > 0$ に対して,

$$\mathbb{P}(|X| \geq t) \leq 2\exp\left(-\frac{2t}{\lambda}\right)$$

をみたすとする. このとき, $\forall k \in \mathbb{N}$ に対して,

$$\mathbb{E}\left[|X|^k\right] \leq \lambda^k k!$$

[2] Lebesgue 測度に関する確率密度関数が

$$f(x) = \frac{1}{2}e^{-|x|} \qquad (x \in \mathbb{R})$$

を持つものである.

が成り立つ．さらに,

$$(\mathbb{E}[|X|^k])^{1/k} \leq 2\lambda k$$

と

$$\mathbb{E}[e^{sX}] \leq \exp\left(2s^2\lambda^2\right) \qquad \left(\forall |s| \leq \frac{1}{2\lambda}\right)$$

が成立する．

証明　$k \in \mathbb{N}$ と $t > 0$ に対して,

$$\begin{aligned}
\mathbb{E}\big[|X|^k\big] &= \int_0^\infty \mathbb{P}\big(|X| > t^{1/k}\big)\, dt \qquad (\because \text{命題 } 2.42) \\
&\leq \int_0^\infty 2\exp\left(-\frac{2t^{1/k}}{\lambda}\right) dt \qquad (\because \text{仮定}) \\
&= 2\left(\frac{\lambda}{2}\right)^k k \int_0^\infty e^{-u} u^{k-1}\, du \qquad \left(u = \tfrac{2t^{1/k}}{\lambda} \text{ と変換}\right) \\
&\leq \lambda^k k\Gamma(k) = \lambda^k k!.
\end{aligned}$$

2 番目の主張は, $k \geq 1$ に対して,

$$\Gamma(k) \leq k^k; \qquad k^{1/k} \leq e^{1/e} \leq 2$$

に注意すれば,

$$(\lambda^k k\Gamma(k))^{1/k} \leq 2\lambda k$$

よりわかる．

最後に, X の積率母関数を評価するために, 指数関数の Taylor 展開を用いる．優収束定理より, $|s| \leq 1/(2\lambda)$ に対して,

$$\begin{aligned}
\mathbb{E}[e^{sX}] &\leq 1 + \sum_{k=2}^\infty \frac{|s|^k \mathbb{E}[|X|^k]}{k!} \qquad (\because \text{一番目の主張より}) \\
&\leq 1 + \sum_{k=2}^\infty (|s|\lambda)^k = 1 + s^2\lambda^2 \sum_{k=0}^\infty (|s|\lambda)^k \\
&\leq 1 + s^2\lambda^2 \sum_{k=0}^\infty \left(\frac{1}{2}\right)^k = 1 + 2s^2\lambda^2 \leq e^{2s^2\lambda^2}.
\end{aligned}$$

□

定義 2.21 X は母数 $\lambda\,(\lambda > 0)$ の劣指数型分布に従うであるとは, 次をみたすときをいう.
(1) $\mathbb{E}[X] = 0$.
(2) $\forall |s| < 1/\lambda$ に対して,

$$\mathbb{E}\big[e^{sX}\big] \le \exp\left(\frac{1}{2}s^2\lambda^2\right).$$

これを $\mathsf{subE}(\lambda)$ と記す.

補題 2.22 $X \sim \mathsf{subG}(\sigma^2)$ とする. $Z := X^2 - \mathbb{E}[X^2]$ としたとき, $Z \sim \mathsf{subE}(16\sigma^2)$ となる.

証明　優収束定理から

$$
\begin{aligned}
\mathbb{E}[e^{sZ}] &= \mathbb{E}\left[1+\sum_{k=2}^{\infty}\frac{(sZ)^k}{k!}\right] \\
&= \mathbb{E}\left[1+\sum_{k=2}^{\infty}\frac{s^k\left\{X^2-\mathbb{E}[X^2]\right\}^k}{k!}\right] \\
&= \mathbb{E}\left[1+\sum_{k=2}^{\infty}\frac{s^k 2^{k-1}\left\{X^{2k}+\left\{\mathbb{E}[X^2]\right\}^k\right.}{k!}\right] \\
&\left(\because (a+b)^k \le 2^{k-1}(a^k+b^k)\ \mathrm{for}\ k\ge 2\ (k\in\mathbb{N}),\ a,\ b\ge 0\right) \\
&= \mathbb{E}\left[1+\sum_{k=2}^{\infty}\frac{s^k 2^{k-1}\left\{\mathbb{E}\left[X^{2k}\right]+\left\{\mathbb{E}[X^2]\right\}^k\right\}}{k!}\right] \\
&= \mathbb{E}\left[1+\sum_{k=2}^{\infty}\frac{s^k 2^{k-1}\left\{\mathbb{E}\left[X^{2k}\right]+\mathbb{E}\left[X^{2k}\right]\right\}}{k!}\right] \qquad (\because \text{Jensen}) \\
&= \mathbb{E}\left[1+\sum_{k=2}^{\infty}\frac{s^k 2^k\mathbb{E}\left[X^{2k}\right]}{k!}\right] \\
&= \mathbb{E}\left[1+\sum_{k=2}^{\infty}\frac{s^k 2^k\mathbb{E}\left[X^{2k}\right]}{k!}\right] \\
&\le 1+\sum_{k=2}^{\infty}\frac{s^k 2^k(2\sigma^2)^k(2k)\Gamma(k)}{k!} \qquad (\text{補題 } 2.11) \\
&\le 1+2\sum_{k=2}^{\infty}s^k 4^k\sigma^{2k} \\
&\le 1+(4s\sigma^2)^2\sum_{k=0}^{\infty}(8s\sigma^2)^k \\
&\quad \left(\because \textstyle\sum \text{ の前の「2」を取り込むために } 4s\sigma^2 \text{ を } 8s\sigma^2\right) \\
&= 1+(8s\sigma^2)^2\frac{8s\sigma^2}{1-8s\sigma^2} \\
&= 1+128s^2\sigma^4 \qquad \left(|s|\le\frac{1}{16\sigma^2}\text{ に対して}\right) \\
&\le e^{128s^2\sigma^4}
\end{aligned}
$$

よりわかる. □ ここの
ックの

3.1 Bernstein の不等式

命題 2.23 (Bernstein の不等式) $X_1, X_2, \ldots, X_n$ を独立な確率変数列とし, $\mathbb{E}[X_j] = 0$ と $X_j \sim \mathsf{subE}(\lambda)\,(\lambda > 0;\, j = 1, 2, \ldots, n)$ とする.

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

とする. このとき, $\forall t > 0$ に対して,

$$\mathbb{P}\big(\overline{X} > t\big) \vee \mathbb{P}\big(\overline{X} < -t\big) \le \exp\left[-\frac{n}{2}\left(\frac{t^2}{\lambda^2} \wedge \frac{t}{\lambda}\right)\right]$$

となる. ただし, $a, b \in \mathbb{R}$ に対して, $a \wedge b$ は a と b の大きくない方で, $a \vee b$ は a と b の小さくない方である.

証明 一般性を失わず, $\lambda = 1$ としてよい. 次に, Chernoff 限界を用いれば, $\forall s > 0$ に対して,

$$\begin{aligned}
\mathbb{P}\big(\overline{X} > t\big) &= \mathbb{P}\left(\exp\left(s\sum_{j=1}^{n} X_j\right) > e^{snt}\right) \\
&\le e^{-nst}\,\mathbb{E}\left[\exp\big(s\sum_{j=1}^{n} X_j\big)\right] \\
&= e^{-snt}\prod_{j=1}^{n}\mathbb{E}[e^{sX_j}] \\
&\le e^{-snt}\prod_{j=1}^{n} e^{s^2/2} \\
&\qquad \big(\because \text{劣指数分布の定義より}, |s| \le 1 \text{ のとき}, \mathbb{E}\big[e^{sX_j}\big] \le e^{s^2/2}\big) \\
&= \exp\left(\frac{ns^2}{2} - snt\right)
\end{aligned}$$

を得る．よって

$$\begin{aligned}\mathbb{P}\big(\overline{X} > t\big) &= \inf_{0<s\le 1} \exp\left(\frac{ns^2}{2} - snt\right)\\ &= \inf_{0<s\le 1} \left\{\frac{n}{2}\big(s-t\big) - \frac{n}{2}t^2\right\}\end{aligned}$$

を得る．上の式の右辺は $t \le 1$ のとき, $s = t$ で最小値, $-\dfrac{n}{2}t^2$ を取り, $t > 1$ のとき, $s = 1$ で最小値 $-\dfrac{n}{2}t$ を取る．さらに

$$\left(-\frac{n}{2}t^2\right) \vee \left(-\frac{n}{2}t\right) = -\frac{n}{2}\big(t^2 \wedge t\big)$$

に注意すると

$$\mathbb{P}(\overline{X} > t) \le \exp\left(-\frac{n}{2}(t^2 \wedge t)\right)$$

を得る．$\mathbb{P}(\overline{X} < -t)$ に対しても同じようの求めれば

$$\mathbb{P}(\overline{X} < -t) \le \exp\left(-\frac{n}{2}(t^2 \wedge t)\right)$$

となる．$\mathbb{P}(\overline{X} > t)$ と $\mathbb{P}(\overline{X} < -t)$ は同じ上限をもつので, 命題の主張は証明された． □

4　最大値の確率不等式

前節の指数不等式は独立な確率変数列の線形結合に対して成立した．多くの場合, 確率変数列の線形結合のパラメータ上で最大を取ったものの評価に興味がある．この節では確率変数列の 独立性を仮定せず に議論をすることに注意すること．

4.1　有限集合上の最大値

命題 2.24 $n \in \mathbb{N}$ とし, $X_1, X_2, \ldots, X_n$ を確率変数列[3]とし,

$$X_j \sim \mathsf{subG}(\sigma^2)\,(\sigma > 0;\, j = 1, 2, \ldots, n)$$

[3]独立性を仮定する必要はないことに注意せよ．

とする．このとき，

$$
\begin{aligned}
\mathbb{E}\left[\max_{1\le j\le n} X_j\right] &\le \sigma\sqrt{2\log n}, \\
\mathbb{E}\left[\max_{1\le j\le n} |X_j|\right] &\le \sigma\sqrt{2\log(2n)}
\end{aligned}
$$

が成り立つ．さらに，$\forall t>0$ に対して，

$$
\begin{aligned}
\mathbb{P}\left(\max_{1\le j\le n} X_j > t\right) &\le n\exp\left(-\frac{t^2}{2\sigma^2}\right), \\
\mathbb{P}\left(\max_{1\le j\le n} |X_j| > t\right) &\le 2n\exp\left(-\frac{t^2}{2\sigma^2}\right)
\end{aligned}
$$

が成り立つ．

証明　$\forall s>0$ に対して，

$$
\begin{aligned}
\mathbb{E}[\max_{1\le j\le n} X_j] &= \frac{1}{s}\mathbb{E}\left[\log\exp\left(s\max_{1\le j\le n} X_j\right)\right] \\
&\le \frac{1}{s}\log\mathbb{E}\left[\max_{1\le j\le n} e^{sX_j}\right] \\
&\qquad (\because \text{ Jensen の不等式より}) \\
&\le \frac{1}{s}\log\left(\sum_{j=1}^{n}\mathbb{E}[e^{sX_j}]\right) \\
&\le \frac{1}{s}\log\left\{\sum_{j=1}^{n}\exp\left(\frac{s^2\sigma^2}{2}\right)\right\} \\
&= \frac{\log n}{s} + \frac{\sigma^2 s}{2}.
\end{aligned}
$$

ここで

$$
s = \sqrt{\frac{\sigma^2}{2\log n}}
$$

とおけば，

$$
\begin{aligned}
\frac{\log n}{s} + \frac{s^2\sigma^2}{2} &= \sqrt{\frac{\sigma^2\log n}{2}} + \frac{\sigma^2}{2}\sqrt{\frac{2\log n}{\sigma^2}} \\
&= 2\sqrt{\frac{\sigma^2\log n}{2}} = \sigma\sqrt{2\log n}
\end{aligned}
$$

から 1 番目の主張はわかる.

確率に対する最初の不等式は和集合の上限 (union bound) より得られる.

$$\begin{aligned}\mathbb{P}\left(\max_{1\le j\le n} X_j > t\right) &= \mathbb{P}\left(\bigcup_{j=1}^{n}\{X_j > t\}\right)\\ &\le \sum_{j=1}^{n}\mathbb{P}(X_j > t)\\ &\le n\exp\left(-\frac{t^2}{2\sigma^2}\right). \qquad (\because \text{補題 } 2.10)\end{aligned}$$

残りの 2 つの不等式は, $X_{n+i} = -X_i$ $(i = 1, 2\ \ldots, n)$ とおけば,

$$\max_{1\le j\le n} |X_i| = \max_{1\le j\le 2n} X_i$$

と書けることに注意すればよい. □

4.2 凸多面体上の最大値

凸多面体 P とは, 有限個の頂点 $\mathcal{V}(\mathsf{P})$ を持つコンパクトな集合である. P の頂点の凸多面体の頂点の凸包を $\mathrm{conv}(\mathcal{V}(\mathsf{P}))$ と書く. このとき,

$$\mathsf{P} = \mathrm{conv}(\mathcal{V}(\mathsf{P}))$$

が成立する.

$d \in \mathbb{N}$ とし, $X \in \mathbb{R}^d$ を確率ベクトルとして, 確率変数の族

$$\mathsf{F} := \{\boldsymbol{\theta}^\top \boldsymbol{X};\, \boldsymbol{\theta} \in \mathsf{P}\}$$

を考える. ただし, $\mathsf{P} \subset \mathbb{R}^d$ は頂点を n 個持つ凸多面体である. F は無限集合だが, F 上の最大値は有限集合上の最大値の議論に帰着できる. そのために, 次の補題を用いる.

補題 2.25 $\boldsymbol{c} \in \mathbb{R}^d$ とする. 線形写像

$$\boldsymbol{x} \mapsto \boldsymbol{c}^\top \boldsymbol{x} \qquad (\boldsymbol{x} \in \mathbb{R}^d)$$

を考える. このとき, 任意の凸多面体 $\mathsf{P} \subset \mathbb{R}^d$ に対して,

$$\max_{\boldsymbol{x}\in\mathsf{P}} \boldsymbol{c}^\top \boldsymbol{x} = \max_{\boldsymbol{x}\in\mathcal{V}(\mathsf{P})} \boldsymbol{c}^\top \boldsymbol{x}$$

となる．ただし, $\mathcal{V}(\mathsf{P})$ は P の頂点集合である．

証明　$\mathcal{V}(\mathsf{P}) = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ とする．$\forall \boldsymbol{x} \in \mathsf{P} = \mathrm{conv}(\mathcal{V}(\mathsf{P}))$ に対して, ある非負の定数 $\lambda_1, \lambda_2, \ldots, \lambda_n$ があって,

$$\boldsymbol{x} = \sum_{j=1}^{n} \lambda_j \boldsymbol{x}_j; \qquad \sum_{j=1}^{n} \lambda_j = 1$$

と書ける．したがって,

$$\begin{aligned} \boldsymbol{c}^\top \boldsymbol{x} &= \boldsymbol{c}^\top \left(\sum_{i=1}^{n} \lambda_i \boldsymbol{c}^\top \boldsymbol{x}_i \right) \\ &\le \sum_{i=1}^{n} \lambda_i \max_{\boldsymbol{x} \in \mathcal{V}(\mathsf{P})} \boldsymbol{c}^\top \boldsymbol{x} \\ &= \max_{\boldsymbol{x} \in \mathcal{V}(\mathsf{P})} \boldsymbol{c}^\top \boldsymbol{x}. \end{aligned}$$

よって,

$$\max_{\boldsymbol{x} \in \mathsf{P}} \boldsymbol{c}^\top \boldsymbol{x} \le \max_{\boldsymbol{x} \in \mathcal{V}(\mathsf{P})} \boldsymbol{c}^\top \boldsymbol{x} \le \max_{\boldsymbol{x} \in \mathsf{P}} \boldsymbol{c}^\top \boldsymbol{x}.$$

□

命題 2.26 P を頂点を n 個持つ凸多面体とし, その頂点を $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots, \boldsymbol{v}^{(n)} \in \mathbb{R}^d$ とする．$\boldsymbol{X} \in \mathbb{R}^d$ は確率ベクトルで, 各 $[\boldsymbol{v}^{(i)}]^\top X\,(i = 1, 2, \ldots, n)$ は代行分散 $\sigma^2\,(\sigma > 0)$ を持つ劣ガウス型分布に従う．このとき,

$$\mathbb{E}\left[\max_{\boldsymbol{\theta} \in \mathsf{P}} \boldsymbol{\theta}^\top \boldsymbol{X}\right] \le \sigma\sqrt{2\log n}; \qquad \mathbb{E}\left[\max_{\boldsymbol{\theta} \in \mathsf{P}} |\boldsymbol{\theta}^\top \boldsymbol{X}|\right] \le \sigma\sqrt{2\log(2n)}.$$

さらに, $\forall t > 0$ に対して,

$$\begin{aligned} \mathbb{P}\left(\max_{\boldsymbol{\theta} \in \mathsf{P}} \boldsymbol{\theta}^\top \boldsymbol{X} > t\right) &\le n \exp\left(-\frac{t^2}{2\sigma^2}\right), \\ \mathbb{P}\left(\max_{\boldsymbol{\theta} \in \mathsf{P}} |\boldsymbol{\theta}^\top \boldsymbol{X}| > t\right) &\le 2n \exp\left(-\frac{t^2}{2\sigma^2}\right). \end{aligned}$$

証明　補題 2.25 と命題 2.23 からわかる．□

4.3 ℓ_2 球上の最大値

$\mathbb{B}_2$ の ℓ_2 球は

$$\mathbb{B}_2 := \left\{ \boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top \in \mathbb{R}^d ; \sum_{j=1}^d x_j^2 \le 1 \right\}$$

で定義される．これは明らかに凸多面体ではない．しかし, ℓ_2 球上の最大値を制御できる．これは, $\mathbb{B}_2$ の有限部分集合 A が存在して, A 上の最大値と $\mathbb{B}_2$ 上の最大値が同一のオーダーであるという事実からわかる．

定義 2.27 $K \subset \mathbb{R}^d$ と $\epsilon > 0$ を固定する．集合 N は $\mathbb{R}^d$ 上の距離 $d(\cdot, \cdot)$ に関する ϵ 網であるとは, $N \subset K$ であって, $\forall \boldsymbol{x}$ に対して, ある $\boldsymbol{x} \in N$ が存在して,

$$d(\boldsymbol{x}, \boldsymbol{z}) \le \epsilon$$

となっているときをいう．

注意 2.28 N がノルム $\|\cdot\|$ に関する ϵ 網のとき, K のすべての点は N の点からの距離が ϵ 以下になる．すなわち, $\forall \boldsymbol{x} \in K$ に対して,

$$\inf_{\boldsymbol{x} \in N} \|\boldsymbol{x} - \boldsymbol{z}\| \le \epsilon$$

となる． □

次の補題は $\mathbb{B}_2$ の最小の ϵ 網のサイズの上限を与える．

補題 2.29 $0 < \epsilon < 1$ を固定する．このとき, 単位球 $\mathbb{B}_2$ は Euclid 距離に関する ϵ 網 N で,

$$\#(N) \le \left(\frac{3}{\epsilon} \right)^d$$

をみたすものが存在する．ただし, $\#(N)$ は集合 N の要素の個数である．

証明 再帰的に ϵ 網を作る．$\boldsymbol{x}_1 = \boldsymbol{0}$ を選ぶ．$\forall i \in \mathbb{N}\,(i \ge 2)$ に対して, $\boldsymbol{x}_i$ は $\forall \boldsymbol{x} \in \mathbb{B}_2$ で, $|\boldsymbol{x} - \boldsymbol{x}_j|_2 > \epsilon\,(j = 1, 2, \ldots, i-1)$ ととる．このような $\boldsymbol{x}$ が存在しないとき, この操作を止めることにする．明らか, これは ϵ 網を作る．次に, こ

の操作で作られた ϵ 網のサイズを評価する．$\forall \boldsymbol{x}, \boldsymbol{y} \in N$ に対して, $|\boldsymbol{x}-\boldsymbol{y}|_2 > \epsilon$ なので, $\boldsymbol{x} \in N$ を中心とした半径 $\epsilon/2$ の Euclid 球は互いに排反である．さらに

$$\bigcup_{\boldsymbol{z}\in N}\left\{\boldsymbol{z}+\frac{\epsilon}{2}\mathbb{B}_2\right\} \subset \left(1+\frac{\epsilon}{2}\right)\mathbb{B}_2.$$

ただし

$$\{\boldsymbol{z}+\epsilon\mathbb{B}_2\} := \{\boldsymbol{z}+\epsilon\boldsymbol{x};\, \boldsymbol{x}\in\mathbb{B}_2\}$$

である．よって, 体積を測れば

$$\begin{aligned}\mathrm{vol}\left(\left(1+\frac{\epsilon}{2}\mathbb{B}_2\right)\right) &\geq \mathrm{vol}\left(\bigcup_{\boldsymbol{z}\in N}\left\{\boldsymbol{z}+\frac{\epsilon}{2}\mathbb{B}_2\right\}\right) \\ &= \sum_{\boldsymbol{z}\in N}\mathrm{vol}\left(\left\{\boldsymbol{z}+\frac{\epsilon}{2}\mathbb{B}_2\right\}\right).\end{aligned}$$

これは

$$\left(1+\frac{\epsilon}{2}\right)^d \geq \#(N)\left(\frac{\epsilon}{2}\right)^d$$

と同値である．したがって

$$\#(N) \leq \left(1+\frac{2}{\epsilon}\right)^d \leq \left(\frac{3}{\epsilon}\right)^d.$$

□

命題 2.30 $\boldsymbol{X} \in \mathbb{R}^d$ を代行分散 $\sigma^2\,(\sigma > 0)$ の劣ガウス確率ベクトルとする．このとき

$$\mathbb{E}\left[\max_{\boldsymbol{\theta}\in\mathbb{B}_2}\boldsymbol{\theta}^\top\boldsymbol{X}\right] = \mathbb{E}\left[\max_{\boldsymbol{\theta}\in\mathbb{B}_2}|\boldsymbol{\theta}^\top\boldsymbol{X}|\right] \leq 4\sigma\sqrt{d}.$$

さらに $\forall\delta > 0$ に対して,

$$\mathbb{P}\left(\max_{\boldsymbol{\theta}\in\mathbb{B}_2}\boldsymbol{\theta}^\top\boldsymbol{X} = \max_{\boldsymbol{\theta}\in\mathbb{B}_2}|\boldsymbol{\theta}^\top\boldsymbol{X}| \leq 4\sigma\sqrt{d}+2\sigma\sqrt{2\log(1/\delta)}\right) \geq 1-\delta$$

が成立する．

証明　N を Euclid 距離に関する $\mathbb{B}_2$ の $(1/2)$ 網とする．すると, 補題 2.16 より

$$\#(N) \leq 6^d$$

となる．$\boldsymbol{\theta} \in \mathbb{B}_2$ に対して, $\exists \boldsymbol{z} \in N$ と $\boldsymbol{x} \in \mathbb{R}^d$ で $|\boldsymbol{x}|_2 \leq 1/2$ なるものがあって,

$$\boldsymbol{\theta} = \boldsymbol{z} + \boldsymbol{x}$$

と書ける．したがって

$$\max_{\boldsymbol{\theta} \in \mathbb{B}_2} \boldsymbol{\theta}^\top \boldsymbol{X} = \max_{\boldsymbol{z} \in N} \boldsymbol{z}^\top \boldsymbol{X} + \max_{\boldsymbol{x} \in (1/2)\mathbb{B}_2} \boldsymbol{x}^\top \boldsymbol{X}.$$

しかし

$$\max_{\boldsymbol{x} \in (1/2)\mathbb{B}_2} \boldsymbol{x}^\top \boldsymbol{X} = \frac{1}{2} \max_{\boldsymbol{\theta} \in \mathbb{B}_2} \boldsymbol{\theta}^\top \boldsymbol{X}$$

なので

$$\max_{\boldsymbol{\theta} \in \mathbb{B}_2} \boldsymbol{\theta}^\top \boldsymbol{X} \leq 2 \max_{\boldsymbol{z} \in N} \boldsymbol{z}^\top \boldsymbol{X}.$$

よって, 命題 2.24 より

$$\begin{aligned}
\mathbb{E}\left[\max_{\boldsymbol{\theta} \in \mathbb{B}_2} \boldsymbol{\theta}^\top \boldsymbol{X}\right] &\leq \mathbb{E}\left[2 \max_{\boldsymbol{z} \in N} \boldsymbol{z}^\top \boldsymbol{X}\right] \leq 2\sigma\sqrt{2\log(\#(N))} \leq 2\sigma\sqrt{2d\log 6} \\
&\leq 4\sigma\sqrt{d}.
\end{aligned}$$

確率不等式は以下からわかる．命題 2.24 より

$$\begin{aligned}
\mathbb{P}\left(\max_{\boldsymbol{\theta} \in \mathbb{B}_2} \boldsymbol{\theta}^\top \boldsymbol{X} > t\right) &\leq \mathbb{P}\left(2 \max_{\boldsymbol{z} \in N} \boldsymbol{z}^\top \boldsymbol{X} > t\right) \leq \#(N) \exp\left(-\frac{t^2}{8\sigma^2}\right) \\
&\leq 6^d \exp\left(-\frac{t^2}{8\sigma^2}\right).
\end{aligned}$$

最後に

$$5^d \exp\left(-\frac{t^2}{8\sigma^2}\right) \leq \delta \Longleftrightarrow t^2 \geq 8\sigma^2 d \log 6 + 8\sigma^2 \log(1/\delta)$$

と $\sqrt{8\log 6} < 4$ に注意[4]して,

$$t := 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}$$

[4]$\log(6) \doteq 1.791759$ である．

とおけば,

$$t^2 \geq 8\sigma^2 d \log 6 + 8\sigma^2 \log(1/\delta)$$

となるので,

$$\begin{aligned}\mathbb{P}\left(\max_{\boldsymbol{\theta}\in\mathbb{B}_2} \boldsymbol{\theta}^\top \boldsymbol{X} > 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}\right) &\leq \mathbb{P}\left(\max_{\boldsymbol{\theta}\in\mathbb{B}_2} \boldsymbol{\theta}^\top \boldsymbol{X} > t\right) \\ &\leq \exp\left(-\frac{t^2}{8\sigma^2} + d\log 6\right) \leq \delta\end{aligned}$$

となる. 上式の補事象をとれば,

$$\mathbb{P}\left(\max_{\boldsymbol{\theta}\in\mathbb{B}_2} \boldsymbol{\theta}^\top \boldsymbol{X} \leq 4\sigma\sqrt{d} + 2\sigma\sqrt{2\log(1/\delta)}\right) \geq 1-\delta$$

がわかる. □

5　独立なランダム行列の和に対する最大値の確率不等式

この節では, 確率集中不等式の主張がどのようにランダム行列の和に対して拡張できるかを議論する. 独立なランダム行列の和に対する Bernstein の不等式の拡張を与える. このようなタイプの不等式はランク回復問題に対するある種の保障を与えることになる.

特に, n 個の独立な対称行列の和の最大固有値に対する確率 $\mathbb{P}\big(\lambda_{\max} \sum_{j=1}^n \boldsymbol{X}_j > t\big)$ の制御は, 様々な場面で必要となる. このようなことをスカラー値の確率変数に対して考えている. しかし, $d \times d$ の対称行列 $\boldsymbol{A}$, $\boldsymbol{B}$ に対して, $\boldsymbol{AB} \neq \boldsymbol{BA}$ のとき,

$$\exp\{\boldsymbol{A} + \boldsymbol{B}\} = \exp \boldsymbol{A} \exp \boldsymbol{B}$$

となるので, 解析上の困難が発生する. ただし,

$$\exp \boldsymbol{A} = \sum_{j=1}^{\infty} \frac{\boldsymbol{A}^j}{j!}; \quad \boldsymbol{A}^0 = \boldsymbol{I}_d; \quad 0! = 1$$

である.

準備

$\mathrm{Mat}(d;\,\mathbb{R})$: $d \times d$ の実行列全体の成す集合.
$\mathrm{Sym}(d;\,\mathbb{R}) := \{\boldsymbol{A} \in \mathrm{Mat}(d;\,\mathbb{R});\ \boldsymbol{A} = \boldsymbol{A}^\top\}$.
$\mathrm{Sym}^+(d;\,\mathbb{R}) := \{\boldsymbol{A} \in \mathrm{Sym}(d;\,\mathbb{R});\ \boldsymbol{A}$ は半正定値 $\Longleftrightarrow \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0\,(\forall \boldsymbol{x} \in \mathbb{R}^d)\ \}$.
$\mathrm{Sym}^{++}(d;\,\mathbb{R}) := \{\boldsymbol{A} \in \mathrm{Sym}(d;\,\mathbb{R});\ \boldsymbol{A}$ は正定値 $\Longleftrightarrow \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} > 0\,(\forall \boldsymbol{x} \in \mathbb{R}^d,\ \boldsymbol{x} \neq \boldsymbol{0})\ \}$.

次のことに注意する. 対称行列 $\boldsymbol{A}$ に対して, $\lambda_j(\boldsymbol{A})\,(j = 1, 2, \ldots, d)$ を j 番目の大きさの固有値とする. このとき,

$$\lambda_d(\boldsymbol{A}) \geq 0 \quad \Longleftrightarrow \quad \boldsymbol{A} \in \mathrm{Sym}^+(d;\,\mathbb{R}),$$
$$\lambda_d(\boldsymbol{A}) > 0 \quad \Longleftrightarrow \quad \boldsymbol{A} \in \mathrm{Sym}^{++}(d;\,\mathbb{R}),$$

定義 2.31 関数 $f : \mathbb{R} \longrightarrow \mathbb{R}$ を考える. $\boldsymbol{A} \in \mathrm{Sym}(d;\,\mathbb{R})$, $\lambda_1, \lambda_2, \ldots, \lambda_d$ をその固有値, $\lambda_j(\boldsymbol{A})\,(j = 1, 2, \ldots, d)$ に対応する固有ベクトルを $\boldsymbol{u}_j \in \mathbb{R}^d$ とする. すなわち,

$$\boldsymbol{A} = \sum_{j=1}^{d} \lambda_j \boldsymbol{u}_j \boldsymbol{u}_j^\top$$

である. このとき,

$$f(\boldsymbol{A}) := \sum_{j=1}^{d} f(\lambda_j) \boldsymbol{u}\boldsymbol{u}^\top$$

で定義する.

すると,

$$\exp \boldsymbol{A} = \sum_{j=1}^{\infty} \frac{\boldsymbol{A}^j}{j!}; \quad \boldsymbol{A}^0 = \boldsymbol{I}_d; \quad 0! = 1$$

と書けることに注意せよ. また, $\boldsymbol{A} \in \mathrm{Sym}^{++}(d;\,\mathbb{R})$ に対して, 写像 $\log : \mathrm{Sym}^{++}(d;\,\mathbb{R}) \longrightarrow \mathrm{Sym}(d;\,\mathbb{R})$ は定義される. $\|\boldsymbol{A} - \boldsymbol{I}_d\| < 1$ のとき,

$$\log \boldsymbol{A} = \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} (\boldsymbol{A} - \boldsymbol{I}_d)^j$$

と書けることがわかる. ただし, $\boldsymbol{A} = (a_{jk}) \in \mathrm{Mat}(d;\,\mathbb{R})$ に対して,

$$\|\boldsymbol{A}\| := \sqrt{\sum_{j=1}^{d} \sum_{k=1}^{d} a_{jk}^2}$$

である．この点については, 村上 (1973, p.88) を参照せよ．

さらに, $\boldsymbol{A}$, $\boldsymbol{B} \in \mathrm{Sym}(d; \mathbb{R})$ に対して, $\boldsymbol{A} \precsim \boldsymbol{B}$ ならば,

$$\mathrm{Tr}\,\exp \boldsymbol{A} \precsim \mathrm{Tr}\,\exp \boldsymbol{B}$$

である．また, $\boldsymbol{A}$, $\boldsymbol{B} \in \mathrm{Sym}^{++}(d; \mathbb{R})$ に対して, $\boldsymbol{A} \precsim \boldsymbol{B}$ ならば,

$$\log \boldsymbol{A} \precsim \log \boldsymbol{B} \tag{2.5}$$

である．

注意 2.32 $\boldsymbol{A}$, $\boldsymbol{B} \in \mathrm{Mat}(d; \mathbb{R})$ に対して, (2.5) に対応するものは成立しない．
□

5.1 Laplace 変換法

命題 2.33 $\boldsymbol{Y} \in \mathrm{Sym}(d; \mathbb{R})$ をランダム行列とする．このとき, $\forall t \in \mathbb{R}$ に対して,

$$\mathbb{P}\Big(\lambda_{\max}(\boldsymbol{Y}) \geq t\Big) \leq \inf_{\theta>0}\Big\{e^{-\theta t}\mathbb{E}\Big[\mathrm{Tr}\left\{\exp(\theta \boldsymbol{Y})\right\}\Big]\Big\}$$

である．ただし, $\lambda_{\max}(\boldsymbol{Y})$ は $\boldsymbol{Y}$ の最大固有値である．

証明

$$\begin{aligned}
\mathbb{P}\Big(\lambda_{\max}(\boldsymbol{Y}) \geq t\Big) &= \mathbb{P}\Big(\lambda_{\max}(\theta\boldsymbol{Y}) \geq \theta t\Big) \leq e^{-\theta t}\mathbb{E}\Big[\exp\Big(\lambda_{\max}(\theta\boldsymbol{Y})\Big)\Big] \\
&\leq e^{-\theta t}\mathbb{E}\Big[\mathrm{Tr}\Big(\exp(\theta\boldsymbol{Y})\Big)\Big]
\end{aligned}$$

□

スカラー値の独立な n 個の確率変数列 $X_1, X_2, \ldots, X_n$ の和に対して, Bernstein の不等式や Hoeffding の不等式を証明するときに,

$$\mathbb{E}\Big[e^{\sum_{j=1}^{n} X_j}\Big] = \mathbb{E}\Big[\prod_{j=1}^{n} e^{X_j}\Big] = \prod_{j=1}^{n}\mathbb{E}\big[e^{X_j}\big]$$

という事実は重要であった．しかし，$\boldsymbol{A}$, $\boldsymbol{B} \in \mathrm{Sym}(d;\,\mathbb{R})$ に対して，一般には，

$$\exp(\boldsymbol{A}+\boldsymbol{B}) \neq \exp(\boldsymbol{A})\exp(\boldsymbol{B})$$

である．その代りに，Golden-Thompson の不等式

$$\mathrm{Tr}\left(\exp\{\theta(\boldsymbol{A}+\boldsymbol{B})\}\right) \leq \mathrm{Tr}\left(\exp(\theta\boldsymbol{A})\exp(\theta\boldsymbol{B})\right)$$

でおきかえることはできる．だが，不幸なことに，この不等式は 3 個以上の行列に対しては成立しない．

定義 2.34 写像 $f : \mathrm{Sym}(d;\,\mathbb{R}) \longrightarrow \mathrm{Sym}(d;\,\mathbb{R})$ は凹であるとは，$\boldsymbol{A}$, $\boldsymbol{B} \in \mathrm{Sym}(d;\,\mathbb{R})$ と $0 \leq \lambda \leq 1$ に対して，

$$f(\lambda\boldsymbol{A}+(1-\lambda)\boldsymbol{B}) \geq \lambda f(\boldsymbol{A})+(1-\lambda)f(\boldsymbol{B})$$

が成立することである．

命題 2.35 (Lieb の不等式) $\boldsymbol{H} \in \mathrm{Sym}(d;\,\mathbb{R})$ とする．$\mathrm{Sym}^{++}(d;\,\mathbb{R})$ 上の写像を

$$f(\boldsymbol{X}) := \mathrm{Tr}\,\exp\big(\boldsymbol{H}+\log\boldsymbol{X}\big)$$

で定義する．このとき，f は $\mathrm{Sym}^{++}(d;\,\mathbb{R})$ 上の凹写像である．

証明 Ruskai (2002, J.Math.Phy., pp.4358–4375) を参照. □

系 2.36 $\boldsymbol{X}$, $\boldsymbol{H} \in \mathrm{Sym}(d;\,\mathbb{R})$ とし，$\boldsymbol{X}$ をランダム行列で，$\boldsymbol{H}$ を定数行列とする．このとき，

$$\mathbb{E}\big[\mathrm{Tr}\,\{\exp(\boldsymbol{H}+\boldsymbol{X})\}\big] \leq \mathrm{Tr}\,\exp\big\{\boldsymbol{H}+\log\mathbb{E}[e^{\boldsymbol{X}}]\big\}$$

が成立する．

証明 $\boldsymbol{Y} = e^{\boldsymbol{X}}$ と書く．Jensen の不等式を用いると，

$$\begin{aligned}
\mathbb{E}\big[\mathrm{Tr}\,\{\exp(\boldsymbol{X}+\boldsymbol{H})\}\big] &= \mathbb{E}\big[\mathrm{Tr}\,\{\exp(\boldsymbol{H}+\log\boldsymbol{Y})\}\big] \\
&= \mathrm{Tr}\,\big\{\mathbb{E}\big[\exp(\boldsymbol{H}+\log\boldsymbol{Y})\big]\big\} \\
&\leq \mathrm{Tr}\,\big\{\exp\big(\boldsymbol{H}+\mathbb{E}[\log\boldsymbol{Y}]\big)\big\} \\
&= \mathrm{Tr}\,\big\{\exp\big(\boldsymbol{H}+\log\mathbb{E}[e^{\boldsymbol{X}}]\big)\big\}.
\end{aligned}$$

□

補題 2.37 $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n \in \mathrm{sym}(d; \mathbb{R})$ を独立なランダム行列とする．このとき, $\forall\theta \in \mathbb{R}$ に対して,

$$\mathbb{E}\big[\mathrm{Tr}\ \exp\big(\sum_{j=1}^{n} \theta \boldsymbol{X}_j\big)\big] \leq \mathrm{Tr}\ \exp\big(\sum_{j=1}^{n} \log \mathbb{E}[\theta \boldsymbol{X}_j]\big).$$

証明　一般性を失わず, $\theta = 1$ としてよい．$j = 1, 2, \ldots, n-1$ に対して, $\mathbb{E}_j$ を $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_j$ を与えたときの条件付き期待値とする．また, $\mathbb{E}_0 = \mathbb{E}$ とする．条件付き期待値の性質と Lieb の不等式より

$$\begin{aligned}
\mathbb{E}\big[\mathrm{Tr}\ \exp\big(\sum_{j=1}^{n} \boldsymbol{X}_j\big)\big] &= \mathbb{E}_0\mathbb{E}_1 \cdots \mathbb{E}_{n-1}\Big[\mathrm{Tr}\ \exp\big(\sum_{j=1}^{n} \boldsymbol{X}_j\big)\Big] \\
&\leq \mathbb{E}_0\mathbb{E}_1 \cdots \mathbb{E}_{n-2}\Big[\mathrm{Tr}\ \exp\big(\sum_{j=1}^{n-1} \boldsymbol{X}_j + \log \mathbb{E}_{n-1}[e^{\boldsymbol{X}_n}]\big)\Big] \\
&= \mathbb{E}_0\mathbb{E}_1 \cdots \mathbb{E}_{n-2}\Big[\mathrm{Tr}\ \exp\Big(\sum_{j=1}^{n-1} \boldsymbol{X}_j + \log \mathbb{E}[e^{\boldsymbol{X}_n}]\Big)\Big] \\
&= \mathbb{E}_0\mathbb{E}_1 \cdots \mathbb{E}_{n-3}\big[\mathrm{Tr}\ \exp\Big(\sum_{j=1}^{n-2} \boldsymbol{X}_j \\
&\qquad + \log \mathbb{E}_{n-1}[e^{\boldsymbol{X}_{n-1}} + \log \mathbb{E}[e^{\boldsymbol{X}_n}]\Big)\Big] \\
&\ \ \vdots \\
&\leq \mathrm{Tr}\ \exp\left(\sum_{j=1}^{n} \log \mathbb{E}[e^{\boldsymbol{X}_j}]\right)
\end{aligned}$$

□

定理 2.38 (裾確率のマスター定理) $\forall t \in \mathbb{R}$ に対して,

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_{j=1}^{n} \boldsymbol{X}_j\big) \geq t\Big) \leq \inf_{\theta > 0}\Big[e^{-\theta}\mathrm{Tr}\left\{\exp\Big(\sum_{j=1}^{n} \log \mathbb{E}[e^{\theta \boldsymbol{X}_j}]\Big)\right\}\Big]$$

が成り立つ.

証明　補題 2.37 と命題 $pro:3-8$ を用いる.

$$\begin{aligned}\mathbb{P}\left(\lambda_{\max}\big(\sum_{j=1}^{n}\boldsymbol{X}_j\big)\geq t\right) &\leq \inf_{\theta>0}\left\{e^{-\theta}\mathbb{E}\left[\mathrm{Tr}\left\{\exp\big(\sum_{j=1}^{n}\theta\boldsymbol{X}_j\big)\right\}\right]\right\} \\ &\leq \inf_{\theta>0}\left[e^{-\theta t}\mathrm{Tr}\left\{\exp\left(\sum_{j=1}^{n}\log\mathbb{E}[e^{\theta\boldsymbol{X}_j}]\right)\right\}\right].\end{aligned}$$

□

系 2.39 ある関数 $g:\ (0,\,+\infty)\longrightarrow[0,\,+\infty)\cup\{+\infty\}$ と固定した行列 $\boldsymbol{A}_i\in\mathrm{Sym}(d;\,\mathbb{R})\,(j=1,\,2,\,\ldots,\,n)$ があって, $\forall j\in\{1,\,2,\,\ldots,\,n\}$ に対して,

$$\mathbb{E}[e^{\theta\boldsymbol{X}_j}]\precsim\exp\big\{g(\theta)\boldsymbol{A}_j\big\}\qquad(\theta\in(0,\,+\infty))$$

が成り立つと仮定する.

$$\rho=\lambda_{\max}\left(\sum_{j=1}^{n}\boldsymbol{A}_j\right)$$

とおく. このとき, $\forall t\in\mathbb{R}$ に対して,

$$\mathbb{P}\left(\lambda_{\max}\big(\sum_{j=1}^{n}\boldsymbol{X}_j\big)\geq t\right)\leq d\inf_{\theta>0}\left\{\exp\big(-\theta t+g(\theta)\rho\big)\right\}$$

が成り立つ.

証明　定理 2.38 より

$$\begin{aligned}\mathbb{P}\left(\lambda_{\max}\left(\sum_{j=1}^{n}\boldsymbol{X}_j\right)\geq t\right) &\leq \inf_{\theta>0}e^{-\theta t}\mathrm{Tr}\left\{\exp\left(g(\theta)\sum_{j=1}^{n}\boldsymbol{A}_j\right)\right\} \\ &\leq \inf_{\theta>0}e^{-\theta t}\mathrm{Tr}\left\{\exp\left(g(\theta)\sum_{j=1}^{n}\boldsymbol{A}_j\right)\right\} \\ &\leq \inf_{\theta>0}e^{-\theta t}\mathrm{Tr}\left\{\exp\Big(g(\theta)\rho\boldsymbol{I}_d\Big)\right\} \\ &\qquad(\because\ \boldsymbol{A}\precsim\boldsymbol{B}\Longrightarrow\exp(\boldsymbol{A})\precsim\exp(\boldsymbol{B})) \\ &= \inf_{\theta>0}e^{-\theta t}d\left\{\exp\big(g(\theta)\rho\big)\right\} \\ &= \inf_{\theta>0}\exp\left\{-\theta t+g(\theta)\rho\right\}.\end{aligned}$$

$\square$

補題 2.40 $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{0}$ と $\lambda_{\max}(\boldsymbol{X}) \le 1$ ならば, $\forall \theta \in \mathbb{R}$ に対して,

$$\mathbb{E}[e^{\theta \boldsymbol{X}}] \precsim \exp\bigl\{(e^\theta - \theta - 1)\mathbb{E}[\boldsymbol{X}^2]\bigr\}$$

が成り立つ.

証明 関数 $f; \mathbb{R} \longrightarrow \mathbb{R}$ を

$$f(x) = \begin{cases} \dfrac{e^{\theta x} - \theta x - 1}{x^2} & (x \neq 0), \\ \dfrac{\theta^2}{2} & (x = 0) \end{cases}$$

と定める. この関数は $\mathbb{R}$ 上でなめらかで単調増加である. したがって,

$$f(x) \le f(1) \qquad (x \le 1).$$

$\lambda_{\max}(\boldsymbol{X}) \le 1$ ならば, $\boldsymbol{X} \precsim \boldsymbol{I}_d$ なので,

$$f(\boldsymbol{X}) \precsim f(\boldsymbol{I}_d) \precsim f(1)\boldsymbol{I}_d.$$

このことより

$$\begin{aligned} e^{\theta \boldsymbol{X}} &= \boldsymbol{I}_d + \theta \boldsymbol{X} + \boldsymbol{X} f(\boldsymbol{X}) \boldsymbol{X} \\ &\precsim \boldsymbol{I}_d + \theta \boldsymbol{X} + f(1) \boldsymbol{X}^2. \end{aligned}$$

この式の両辺の期待値を取れば,

$$\begin{aligned} \mathbb{E}[e^{\theta \boldsymbol{X}}] &\precsim \boldsymbol{I}_d + f(1)\mathbb{E}[\boldsymbol{X}^2] \precsim \exp\bigl(\exp f(1)\mathbb{E}[\boldsymbol{X}^2]\bigr) \\ &= \exp\bigl\{(e^\theta - \theta - 1)\mathbb{E}[\boldsymbol{X}^2]\bigr\}. \end{aligned}$$

$\square$

定理 2.41 (Bernstein の不等式) $j = 1, 2, \ldots, n$ に対して, $\mathbb{E}[\boldsymbol{X}_j] = \boldsymbol{0}$ で, ほとんど確実に $\lambda_{\max}(\boldsymbol{X}_j) \le R$ とする. ただし, R は正の定数とする.

$$\sigma^2 = \| \sum_{j=1}^{\infty} \mathbb{E}[\boldsymbol{X}_j^2 \|_{\mathrm{op}}; \qquad \|\boldsymbol{A}\|_{\mathrm{op}} := \sup_{\boldsymbol{u} \in \mathbb{S}^{d-1}} \sqrt{\boldsymbol{u}^\top \boldsymbol{A} \boldsymbol{u}}\, (\boldsymbol{A} \in \mathrm{Sym}(d; \mathbb{R}))$$

とする．このとき, $\forall t > 0$ に対して,

$$\begin{aligned}\mathbb{P}\left(\lambda_{\max}\left(\sum_{j=1}^{n}\boldsymbol{X}_j\right) \geq t\right) &\leq d\exp\left(-\frac{\sigma^2}{R}h\left(\frac{Rt}{\sigma^2}\right)\right) \\ &\leq d\exp\left(-\frac{t^2/2}{\sigma^2+(Rt)/3}\right) \\ &\leq \begin{cases} d\exp\left(-\dfrac{3t^2}{8\sigma^2}\right) & \left(t \leq \frac{\sigma^2}{R}\right) \\ d\exp\left(-\dfrac{3t}{8R}\right) & \left(t > \frac{\sigma^2}{R}\right)\end{cases}\end{aligned}$$

が成り立つ．ただし,

$$h(u) = (1+u)\log(1+u) - u \qquad (u > 0).$$

証明　一般性を失わず, $R = 1$ としてよい．補題 2.40 より

$$\mathbb{E}[e^{\theta \boldsymbol{X}_j}] \precsim \exp\left(g(\theta)\mathbb{E}[\boldsymbol{X}_j^2]\right); \qquad g(\theta) = e^{\theta} - \theta - 1 \qquad (\theta > 0).$$

系 2.39 より

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{j=1}^{n}\boldsymbol{X}_j\right) \geq t\right) \leq e\exp(-\theta t + g(\theta)\sigma^2).$$

ここで,

$$f(\theta) := -\theta t - \sigma^2 e^{\theta} - \sigma^2\theta - \sigma^2$$

とおけば,

$$f'(\theta) = -t - \sigma^2 e^{\theta} - \sigma^2 = 0 \Longleftrightarrow e^{\theta} = 1 + \frac{t}{\sigma^2} \Longleftrightarrow \theta = \log\left(1 + \frac{t}{\sigma^2}\right)$$

なので,

$$
\begin{aligned}
\mathbb{P}\left(\lambda_{\max}\left(\sum_{j=1}^{n}\boldsymbol{X}_j\right) \geq t\right) &\leq d\exp\left\{-t\log\left(1+\frac{t}{\sigma^2}\right)+g\left(1+\frac{t}{\sigma^2}\right)\sigma^2\right\} \\
&= d\exp\left\{-t\log\left(1+\frac{t}{\sigma^2}\right)+\sigma^2\left(1+\frac{t}{\sigma^2}\right)\right. \\
&\qquad\qquad \left.-\sigma^2\log\left(1+\frac{t}{\sigma^2}\right)-\sigma^2\right\} \\
&= d\exp\left\{-\sigma^2\left(1+\frac{t}{\sigma^2}\right)\log\left(1+\frac{t}{\sigma^2}\right)+\frac{t}{\sigma^2}\right\} \\
&= d\exp\left(-\sigma^2 h\left(\frac{t}{\sigma^2}\right)\right).
\end{aligned}
$$

次に, 不等式

$$
h(u) \geq \frac{u^2/2}{1+u/3} \qquad (u>0)
$$

に注意すれば、

$$
\begin{aligned}
d\exp\left(-\sigma^2 h\left(\frac{t}{\sigma^2}\right)\right) &\leq d\exp\left(-\sigma^2\frac{\frac{t^2}{2\sigma^4}}{1+\frac{t}{3\sigma^2}}\right) = d\exp\left(-\frac{t^2}{\sigma^2+t/3}\right) \\
&\leq \begin{cases} d\exp\left(-\dfrac{3t^2}{8\sigma^2}\right) & \left(t \leq \sigma^2\right) \\ d\exp\left(-\dfrac{3t}{8}\right) & \left(t > \sigma^2\right) \end{cases}
\end{aligned}
$$

を得る. □

6 補遺

命題 2.42 確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の非負値確率変数 X と $p>0$ に対して, X^p は $\mathbb{P}$ 可積分とする. このとき,

$$
\mathbb{E}[X^p] = p\int_0^\infty x^{p-1}\mathbb{P}(X>x)\,dx
$$

が成り立つ. 非負値確率変数

証明　Fubini の定理より

$$\begin{aligned}
\mathbb{E}[X^p] &= \int_0^\infty \left\{ \int_0^\infty \mathbb{1}_{[0,\, x^p]}(y)\, dy \right\} dP_X(x) = \int_0^\infty \left\{ \int_{y^{1/p}}^\infty dP_X(x) \right\} dy \\
&= \int_0^\infty P_X((y^{1/p},\, \infty))\, dy = p \int_0^\infty x^{p-1} P_X((x,\, +\infty))\, dx \\
&= p \int_0^\infty x^{p-1} \mathbb{P}(X > x)\, dx.
\end{aligned}$$

□