第12章　Talagrand の集中不等式

12.1　導入

Talagrand　は有界な関数族で添え字づけれた一般的な経験過程に対する集中不等式を導出した.

Giné and Nickl (2016, pp.149-) も参考にすること.

12.2　証明のための技術的な補題

P を $\mathbb{X}$ 上の確率分布とする. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とし

$$\begin{aligned}\boldsymbol{X} &:= (X_1, X_2, \ldots, X_n),\\ \boldsymbol{X}_{-j} &:= (X_1, X_2, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n) \qquad (j = 1, 2, \ldots, n)\end{aligned}$$

とする. $\{X_j'\}_{j=1}^n$ を $\{X_j\}_{j=1}^n$ の独立複製としたとき

$$Z_j := h(X_1, X_2, \ldots, X_{j-1}, X_j', X_{j+1}, \ldots, X_n) \tag{12.1}$$

とする. ただし, $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ は有界関数とする. さらに

$$V^+ = \mathsf{E}\left[\sum_{j=1}^n \{Z - Z_j\}^2\right] \tag{12.2}$$

とする.

次の定理は Efron-Morris の不等式 (定理 5.17) の指数関数版と解釈できる.

定理 12.1. Z を $\boldsymbol{X}$ の有界統計量とする. このとき, 任意の $\theta > 0$ と $\lambda \in \left(0, \dfrac{1}{\theta}\right)$ に対して

$$\log \mathsf{E}\big[\lambda\big(Z - \mathsf{E}[Z]\big)\big] \leq \frac{\lambda\theta}{1-\lambda\theta} \log \mathsf{E}\left[\exp\left(\frac{\lambda V^+}{\theta}\right)\right] \tag{12.3}$$

が成立する. ただし, V^+ は (12.2) で与えられる.

定理を証明するために次の 2 つ補題を用意する.

補題 12.2. Z を W を $\boldsymbol{X}$ の有界な統計量とする. このとき, 任意の $\lambda \in \mathbb{R}$ に対して

$$\frac{\mathsf{E}\left[\lambda W e^{\lambda Z}\right]}{\mathsf{E}\left[e^{\lambda Z}\right]} \leq \frac{\mathsf{E}\left[\lambda Z e^{\lambda Z}\right]}{\mathsf{E}\left[e^{\lambda Z}\right]} - \log \mathsf{E}\left[e^{\lambda Z}\right] + \log \mathsf{E}\left[e^{\lambda W}\right]$$

が成立する.

Proof. Q を $\mathbb{X}^n$ 上の確率測度で

$$\mathrm{d}\mathsf{Q} = \frac{e^{\lambda Z}}{\mathsf{E}\left[E^{\lambda Z}\right]} \mathrm{d}\mathsf{P}^n$$

で与えれるるとする. 分布 Q に関する期待値を $\mathsf{E}_{\mathsf{Q}}[\cdot]$ と記すと

$$\mathsf{E}_{\mathsf{Q}}[\lambda W] = \frac{\mathsf{E}\left[\lambda W e^{\lambda Z}\right]}{\mathsf{E}\left[e^{\lambda Z}\right]}, \tag{12.4}$$

$$\mathsf{E}_{\mathsf{Q}}\left[e^{\lambda Z}\right] = \frac{\mathsf{E}\left[\lambda Z e^{\lambda Z}\right]}{\mathsf{E}\left[e^{\lambda Z}\right]}, \tag{12.5}$$

$$\begin{aligned} \log \mathsf{E}_{\mathsf{Q}}\left[e^{\lambda(W-Z)}\right] &= \log \mathsf{E}\left[e^{\lambda(W-Z)} \frac{e^{\lambda Z}}{\mathsf{E}\left[e^{\lambda Z}\right]}\right] = \log \mathsf{E}\left[\frac{e^{\lambda W}}{\mathsf{E}\left[e^{\lambda Z}\right]}\right] \\ &= \log \mathsf{E}\left[e^{\lambda W}\right] - \log \mathsf{E}\left[e^{\lambda Z}\right] \end{aligned} \tag{12.6}$$

となる. 関数 $x \mapsto e^x$ は convex なので, Jensen の不等式から

$$\begin{aligned} \log \mathsf{E}_{\mathsf{Q}}\left[e^{\lambda(W-Z)}\right] &\geq \log \exp\left(\mathsf{E}_{\mathsf{Q}}\left[\lambda(W-Z)\right]\right) \\ &= \mathsf{E}_{\mathsf{Q}}\left[\lambda(W-Z)\right] \\ &= \mathsf{E}_{\mathsf{Q}}\left[\lambda W\right] - \mathsf{E}_{\mathsf{Q}}\left[\lambda Z\right] \end{aligned}$$

となる. この不等式に (12.4) – (12.6) と代入すると

$$\log \mathsf{E}\left[e^{\lambda W}\right] - \log \mathsf{E}\left[e^{\lambda Z}\right] \geq \frac{\mathsf{E}\left[\lambda W e^{\lambda Z}\right]}{\mathsf{E}\left[e^{\lambda Z}\right]} - \frac{\mathsf{E}\left[\lambda Z e^{\lambda Z}\right]}{\mathsf{E}\left[e^{\lambda Z}\right]}$$

から補題は証明された. □

補題 12.3. $\theta > 0$ と $\lambda > 0$ に対して

$$\widetilde{H}(\lambda) = \log \mathsf{E}\left[\exp\left(\lambda V^{+}\right)\right]$$

とする. このとき, 写像

$$\left(0, \frac{1}{\theta}\right) \ni s \mapsto \frac{\widetilde{H}\left(s/\theta\right)}{s(1-s\theta)}$$

は非減少である.

Proof. $0 < s < t < 1/\theta$ を取り, $s = \alpha t\,(0 < \alpha < 1)$ と書く. H́older の不等式から, 関数 $[0, \infty) \ni \lambda \mapsto \log \mathsf{E}\big[e^{\lambda V^*}\big]$ は convex となる. また, $\widetilde{H}(0) = 0$ である. 関数 $\widetilde{H}$ の凸性から

$$\widetilde{H}\left(\frac{s}{\theta}\right) = \widetilde{H}\left(\alpha\frac{t}{\theta} + (1-\alpha)0\right) \le \alpha\widetilde{H}\left(\frac{t}{\theta}\right) + (1-\alpha)\underbrace{\widetilde{H}(0)}_{=0} = \alpha\widetilde{H}\left(\frac{t}{\theta}\right)$$

となる. $\alpha = s/t$ だったので

$$\widetilde{H}\left(\frac{s}{\theta}\right) \le \alpha\widetilde{H}\left(\frac{t}{\theta}\right) \Leftrightarrow \frac{\widetilde{H}\left(\frac{s}{\theta}\right)}{s} \le \frac{\widetilde{H}\left(\frac{t}{\theta}\right)}{t}$$

である. さらに

$$s < t \Leftrightarrow \theta s < \theta t \Leftrightarrow 1 - \theta s > 1 - \theta t \Leftrightarrow \frac{1}{1-\theta s} < \frac{1}{1-\theta t}$$

となる. これらの 2 つの不等式の辺々をかけ合わせると

$$\frac{\widetilde{H}\left(\frac{s(1-\theta s)}{\theta}\right)}{s} \le \frac{\widetilde{H}\left(\frac{t}{\theta}\right)}{t(1-\theta t)}$$

を得る. よって, G の非減少性が証明された. □

<u>定理 12.1 の証明</u>. $\psi(x) = x(e^x - 1)\,(x \ge 0)$ とする. すると

$$\psi(-x) = x(1 - e^{-x}) \le x^2 \tag{12.7}$$

となる. $\forall\lambda > 0$ に対して

$$H(\lambda) = \mathsf{E}\big[\exp(\lambda Z)\big]$$

とおいたとき

$$\begin{aligned}
\lambda \overset{\bullet}{H}(\lambda) - \widetilde{H}(\lambda)\log H(\lambda) &= \lambda\mathsf{E}\big[Ze^{\lambda Z}\big] - \mathsf{E}\big[e^{\lambda Z}\big]\log\mathsf{E}\big[e^{\lambda Z}\big] \\
&\le \sum_{j=1}^{n}\mathsf{E}\left[e^{\lambda Z}\psi\big(-\lambda(Z - Z^{(j)})_+\big)\right] \\
&\le \lambda^2\mathsf{E}\left[e^{\lambda Z}\sum_{j=1}^{n}(Z - Z_j)_+^2\right] \qquad \big(\because (12.7)\big) \\
&= \lambda^2\mathsf{E}\big[V^+e^{\lambda Z}\big] \qquad (12.8)
\end{aligned}$$

となる. $G(\lambda) = \log \mathsf{E}\bigl[e^{\lambda V^+}\bigr]$ $(\lambda \geq 0)$ おくと $G(0) = 0$ かつ G は凸が関数となる. $W = \dfrac{V^+}{\theta}$ として, 補題 12.3 を用いると

$$\begin{aligned}\mathsf{E}\left[\lambda\left(\frac{V^+}{\theta}\right)e^{\lambda Z}\right] &\leq \mathsf{E}\bigl[\lambda Z e^{\lambda Z}\bigr] - \mathsf{E}\bigl[e^{\lambda Z}\bigr]\log \mathsf{E}\bigl[e^{\lambda Z}\bigr] + \mathsf{E}\bigl[e^{\lambda Z}\bigr]\log \mathsf{E}\bigl[e^{\lambda V^+/\theta}\bigr] \\ &= \lambda \overset{\bullet}{H}(\lambda) - H(\lambda)\log H(\lambda) + H(\lambda) G\left(\frac{\lambda}{\theta}\right) \end{aligned} \tag{12.9}$$

となる. (12.8) と (12.9) を合わせると

$$\begin{aligned}&\lambda \overset{\bullet}{H}(\lambda) - H(\lambda)\log H(\lambda) \\ &\quad = \lambda^2 \mathsf{E}\bigl[V^+ e^{\lambda Z}\bigr] \\ &\quad = \lambda\theta \mathsf{E}\left[\lambda\left(\frac{V^+}{\theta}\right)e^{\lambda Z}\right] \\ &\quad \leq \lambda\theta\left\{\lambda \overset{\bullet}{H}(\lambda) - H(\lambda)\log H(\lambda) + H(\lambda) G\left(\frac{\lambda}{\theta}\right)\right\} \\ &\quad = \lambda^2\theta\left\{\overset{\bullet}{H}(\lambda) + \frac{1}{\lambda}H(\lambda)G\left(\frac{\lambda}{\theta}\right) - \frac{1}{\lambda}H(\lambda)\log H(\lambda)\right\}\end{aligned}$$

を得る. この不等式の辺々を $\lambda^2 H(\lambda)$ で割って, 整理をすると

$$\begin{aligned}&\frac{\overset{\bullet}{H}(\lambda}{H(\lambda)} - \frac{1}{\lambda^2}\log H(\lambda) \leq \theta\left\{\frac{\overset{\bullet}{H}(\lambda)}{H(\lambda)} + \frac{1}{\lambda}G\left(\frac{\lambda}{\theta}\right) - \frac{1}{\lambda}\log H(\lambda)\right\} \\ &\qquad \Leftrightarrow\ \frac{1-\lambda\theta}{\lambda}\frac{\overset{\bullet}{H}(\lambda)}{H(\lambda)} - \frac{1-\lambda\theta}{\lambda^2}\log H(\lambda) \leq \frac{\theta}{\lambda}G\left(\frac{\lambda}{\theta}\right) \\ &\qquad \Leftrightarrow\ \frac{1}{\lambda}\frac{\overset{\bullet}{H}(\lambda)}{H(\lambda)} - \frac{1}{\lambda^2}\log H(\lambda) \leq \frac{\theta}{1-\lambda\theta}G\left(\frac{\lambda}{\theta}\right)\end{aligned}$$

を得る. ここで, $\widetilde{H}(\lambda) = \lambda^{-1}\log H(\lambda)$ とおくと

$$\overset{\bullet}{\widetilde{H}}(\lambda) = \frac{1}{\lambda}\frac{\overset{\bullet}{H}(\lambda)}{H(\lambda)} - \frac{1}{\lambda^2}\log H(\lambda) \leq \frac{\theta}{1-\lambda\theta}G\left(\frac{\lambda}{\theta}\right)$$

と

$$\lim_{\lambda\to 0;\,\lambda>0} H(\lambda) = \left.\frac{\overset{\bullet}{H}(\lambda)}{H(\lambda)}\right|_{\lambda=0} = \left.\frac{\mathsf{E}\bigl[Ze^{\lambda Z}\bigr]}{\mathsf{E}\bigl[e^{\lambda Z}\bigr]}\right|_{\lambda=0} = \mathsf{E}[Z]$$

となる．これらのことと補題 12.3 に注意すると

$$
\begin{aligned}
\widetilde{H}(\lambda) - \underbrace{\widetilde{H}(0)}_{=\mathsf{E}[Z]} &= \int_0^\lambda \overset{\bullet}{\widetilde{H}}\,(s)\,\mathrm{d}s \\
&\leq \int_0^\lambda \frac{\theta G(s/\theta)}{s(1-s\theta)}\,\mathrm{d}s \\
&\leq \int_0^\lambda \frac{\theta G(\lambda/\theta)}{\lambda(1-\lambda\theta)}\,\mathrm{d}s \qquad (\because \text{補題 } 12.3) \\
&= \frac{\theta G(\lambda/\theta)}{s(1-\lambda\theta)}
\end{aligned}
$$

となる．この不等式を整理すると

$$
\begin{aligned}
&\frac{\log H(\lambda)}{\lambda} \leq \mathsf{E}[Z] + \frac{\theta G(\lambda/\theta)}{s(1-\lambda\theta)} \\
\Leftrightarrow \quad &\log H(\lambda) \leq \lambda\mathsf{E}[Z] + \frac{\lambda\theta G(\lambda/\theta)}{s(1-\lambda\theta)}
\end{aligned}
$$

がわかる．よって

$$
\begin{aligned}
\log\mathsf{E}\big[\exp\big\{\lambda(Z-\mathsf{E}[Z])\big\}\big] &= \log\mathsf{E}\left[\frac{e^{\lambda Z}}{e^{\lambda\mathsf{E}[Z]}}\right] \\
&= \log\underbrace{\mathsf{E}\big[e^{\lambda Z}\big]}_{=H(\lambda)} - \lambda\mathsf{E}[Z] \\
&\leq \frac{\lambda\theta G(\lambda/\theta)}{s(1-\lambda\theta)} \\
&= \frac{\lambda\theta}{1-\lambda\theta}\log\mathsf{E}\left[\exp\left(\frac{\theta V^+}{\lambda}\right)\right]
\end{aligned}
$$

となり, 定理は証明された． □

定理 12.4. P を標本空間 $\mathbb{X}$ 上の確率分布とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とする．Z を $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の有界な統計量とする．さらに, $\boldsymbol{X}_{-j} = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n)\,(j = 1, 2, \ldots, n)$ とし, $Z_j\,(j = 1, 2, \ldots, n)$ は $\boldsymbol{X}_{-j}$ に依存する有界な統計量とし

$$
0 \leq Z - Z_j \leq 1\,(j = 1, 2, \ldots, n) \qquad \sum_{j=1}^n (Z - Z_j) \leq Z
$$

をみたすとする．このとき, 任意の $\lambda \in \mathbb{R}$ と $j = 1, 2, \ldots, n$ に対して

$$
\log\mathsf{E}\big[e^{\lambda Z}\big] \leq \big(e^\lambda - 1\big)\mathsf{E}[Z] \tag{12.10}
$$

が成り立つ. さらに, 任意の $x > 0$ に対して

$$\mathsf{Pr}\left(Z - \mathsf{E}[Z] \geq \sqrt{2\mathsf{E}[Z]x} + \frac{2}{3}x\right) \leq e^{-x} \tag{12.11}$$

と

$$\mathsf{Pr}\left(Z - \mathsf{E}[Z] \leq -\sqrt{2\mathsf{E}[Z]x}\right) \leq e^{-x} \tag{12.12}$$

が成り立つ.

Proof. <u>(12.10) の証明</u>: $\varphi(x) = e^x - x - 1\,(x \in \mathbb{R})$ と定める. すると $\varphi(0) = 0$, $\varphi(x) \geq 0$, $\varphi(-\lambda u) \leq u\varphi(-\lambda)\,(\lambda \in \mathbb{R},\, 0 \leq u \leq 1)$ となること [1] がわかる. 命題 5.20 の不等式 (5.23) と $\sum_{j=1}^{n}(Z - Z_j) \leq Z$ から

$$\begin{aligned}
\lambda\mathsf{E}\big[Ze^{\lambda Z}\big] - \mathsf{E}\big[e^{\lambda Z}\big]\log\mathsf{E}\big[e^{\lambda Z}\big] &\leq \sum_{j=1}^{n}\mathsf{E}\Big[e^{\lambda Z}\varphi\big(-\lambda(Z - Z_j)\big)\Big] \\
&\leq \sum_{j=1}^{n}\mathsf{E}\Big[e^{\lambda Z}(Z - Z_j)\varphi(-\lambda)\Big] \\
&\qquad \big(\because e^{\lambda Z} > \text{と } \varphi(-\lambda u) \leq u\varphi(-\lambda)\ \big) \\
&= \varphi(-\lambda)\mathsf{E}\Big[e^{\lambda Z}\sum_{j=1}^{n}(Z - Z_j)\Big] \\
&\leq \varphi(-\lambda)\mathsf{E}\big[Ze^{\lambda Z}\big]
\end{aligned} \tag{12.13}$$

がわかる. $\widetilde{Z} = Z - \mathsf{E}[Z]$ とし

$$\begin{aligned}
H(\lambda) = \mathsf{E}\big[e^{\lambda\widetilde{Z}}\big] = \mathsf{E}\big[e^{\lambda Z - \lambda\mathsf{E}[Z]}\big] = \frac{\mathsf{E}\big[e^{\lambda Z}\big]}{e^{-\lambda\mathsf{E}[Z]}} \\
\Leftrightarrow \log H(\lambda) = \log\mathsf{E}\big[e^{\lambda Z}\big] - \lambda\mathsf{E}[Z]
\end{aligned}$$

とおく. すると

$$\frac{\overset{\bullet}{H}(\lambda)}{H(\lambda)} = \frac{\mathsf{E}\big[Ze^{\lambda Z}\big]}{\mathsf{E}\big[e^{\lambda Z}\big]} - \mathsf{E}[Z]$$

[1] 不等式は下記からわかる. 関数 $x \mapsto e^x$ は凸であるので

$$e^{-\lambda u} = \exp\big\{-(u\lambda + (1-u)\times 0)\big\} \leq ue^{-\lambda} + (1-u)e^0 = ue^{-\lambda} + 1 - u$$

となる. よって

$$\varphi(-\lambda u) = e^{-\lambda u} + \lambda u - 1 \leq ue^{-\lambda} + 1 - u + \lambda u - 1 = u\big\{e^{-\lambda} + \lambda - 1\big\} = u\varphi(-\lambda)$$

からわかる.

となるので

$$
\begin{aligned}
\lambda \mathsf{E}\big[Ze^{\lambda Z}\big] - \mathsf{E}\big[e^{\lambda Z}\big]\log \mathsf{E}\big[e^{\lambda Z}\big] &= \mathsf{E}\big[e^{\lambda Z}\big]\left\{\frac{\mathsf{E}\big[Ze^{\lambda Z}\big]}{\mathsf{E}\big[e^{\lambda Z}\big]} - \log \mathsf{E}\big[e^{\lambda Z}\big]\right\} \\
&= \mathsf{E}\big[e^{\lambda Z}\big]\left\{\lambda\frac{\dot{H}(\lambda)}{H(\lambda)} - \log H(\lambda)\right\}, \quad (12.14)
\end{aligned}
$$

$$
\varphi(-\lambda)\mathsf{E}\big[Ze^{\lambda Z}\big] = \varphi(-\lambda)\mathsf{E}\big[e^{\lambda Z}\big]\left\{\frac{\dot{H}(\lambda)}{H(\lambda)} + \mathsf{E}[Z]\right\} \tag{12.15}
$$

となる. (12.14) と (12.15) を (12.13) に代入して整理をすると

$$
\Big\{\underbrace{\lambda - \varphi(-\lambda)}_{\lambda - e^{-\lambda}+\lambda-1}\Big\}\frac{\dot{H}(\lambda)}{H(\lambda)} - \log H(\lambda) \le \varphi(-\lambda)\mathsf{E}[Z] \tag{12.16}
$$

を得る. $G(\lambda) = \log H(\lambda)$ とおくと

$$
\big(1 - e^{-\lambda}\big)\,\dot{G}(\lambda) - G(\lambda) \le \varphi(-\lambda)\mathsf{E}[Z]
$$

となる. いま, $G_0(\lambda) = \varphi(-\lambda)\mathsf{E}[Z]$ とおくと, G_0 は g の常微分方程式

$$
\big(1 - e^{-\lambda}\big)\,\dot{g}(\lambda) - g(\lambda) = \varphi(-\lambda)\mathsf{E}[Z]
$$

の解[2] である. もし

$$
G(\lambda) \le G_0(\lambda) \tag{12.17}
$$

が証明できる (この定理の証明の最後に (12.17) の証明を記しておく) と

$$
\begin{aligned}
&G(\lambda) = \log \mathsf{E}\big[e^{\lambda \widetilde{Z}}\big] = \log \mathsf{E}\big[e^{\lambda Z - \lambda \mathsf{E}[Z]}\big] = \log \mathsf{E}\big[e^{\lambda Z}\big] - \lambda \mathsf{E}[Z], \\
&G_0(\lambda) = \varphi(\lambda)\mathsf{E}[Z]
\end{aligned}
$$

から

$$
G(\lambda) \le G_0(\lambda) \ \Leftrightarrow\ \log \mathsf{E}\big[e^{\lambda Z}\big] \le (\varphi(\lambda) + \lambda)\mathsf{E}[Z] = (e^{\lambda} - 1)\mathsf{E}[Z]
$$

がわかる.

[2] $\dot{G}_0(\lambda) = \big\{e^{\lambda} - 1\big\}\mathsf{E}[Z]$ となるので

$$
\begin{aligned}
(1 - e^{-\lambda})\,\dot{G}_0(\lambda) - G_0(\lambda) &= (1 - e^{-\lambda})(e^{\lambda} - 1)\mathsf{E}[Z] - (e^{\lambda} - \lambda - 1)\mathsf{E}[Z] \\
&= \big\{e^{\lambda} - 1 - 1 + e^{-\lambda} - e^{\lambda} + \lambda + 1\big\}\mathsf{E}[Z] \\
&= e^{-\lambda} + \lambda - 1\big\}\mathsf{E}[Z] = \varphi(-\lambda)\mathsf{E}[Z]
\end{aligned}
$$

からわかる.

(12.11) の証明: $x > 0$ とする. Markov の不等式と (12.10) から, 任意の $\lambda > 0$ に対して

$$
\begin{aligned}
\mathsf{Pr}\big(Z - \mathsf{E}[Z] \geq x\big) &= \mathsf{Pr}\Big(e^{\lambda\left(Z - \mathsf{E}[Z]\right)} \geq e^{\lambda x}\Big) \\
&\leq e^{-\lambda x}\mathsf{E}\Big[e^{\lambda\left(Z-\mathsf{E}[Z]\right)}\Big] \\
&= \exp\left\{-\lambda x - \lambda \mathsf{E}[Z] + \log \mathsf{E}\big[e^{\lambda Z}\big]\right\} \\
&\leq \exp\left\{-\lambda x - \lambda \mathsf{E}[Z] + (e^{\lambda} - 1)\mathsf{E}[Z]\right\} \\
&= \exp\left\{-\lambda x + \varphi(\lambda)\mathsf{E}[Z]\right\}
\end{aligned}
$$

がわかる. あとは上の不等式の最右辺の指数の中を最小化すればよい. そのために

$$
h(\lambda) := -\lambda x + a\varphi(\lambda) = ae^{\lambda} - (x+a)\lambda - a, \qquad a = \mathsf{E}[Z]
$$

とおく. すると

$$
\dot{h}(\lambda) = ae^{\lambda} - (x+a)
$$

となるので

$$
\dot{h}(\lambda) = 0
$$

の解は $\lambda = \log\left(1+\dfrac{x}{a}\right) =: \lambda^*$ となる. 明らかに, 関数 $(0, \infty) \ni \lambda \mapsto h(\lambda)$ は $\lambda = \lambda^*$ で最小となり, その最小値は

$$
\begin{aligned}
h(\lambda^*) &= a\left(1 + \frac{x}{a}\right) - (x+a)\log\left(1+\frac{x}{a}\right) - a \\
&= -a\left\{\left(1+\frac{x}{a}\right)\log\left(1+\frac{x}{a}\right) - \frac{x}{a}\right\} \\
&=: -aq\left(\frac{x}{a}\right)
\end{aligned}
$$

となる. ただし, $q(y) = (1+y)\log(1+y) - y$ とした. すると $y > 0$ に対して

$$
q(y) \geq \frac{y^2}{2\left(1+\dfrac{y}{3}\right)} \tag{12.18}
$$

となること (この定理の最後の方で証明を与える) がわかるので

$$\begin{aligned}\mathsf{Pr}\big(Z - \mathsf{E}[Z] \geq x\big) &\leq e^{h(\lambda^*)} \\ &= \exp\left\{ -aq\left(\frac{x}{a}\right)\right\} \\ &\leq \exp\left\{-a\frac{\left(\frac{x}{a}\right)^2}{2\left(1+\frac{x}{3a}\right)}\right\} \\ &= \exp\left\{-\frac{x^2}{2\big(a+x/3\big)}\right\}\end{aligned}$$

を得る. ここで, $\sqrt{c+d} \leq \sqrt{c} + \sqrt{d}\,(c \geq 0,\, d \geq 0)$ から

$$y = \frac{x^2}{2\big(a+x/3\big)} \Leftrightarrow x = \frac{y}{3} + \sqrt{\frac{y^2}{9} + 2ay} \leq \frac{2y}{3} + \sqrt{2ay}$$

となることに注意する. このことから

$$\begin{aligned}\mathsf{Pr}\left(Z - \mathsf{E}[Z] \geq \frac{2y}{3} + \sqrt{2ay}\right) &\leq \mathsf{Pr}\left(Z - \mathsf{E}[Z] \geq \frac{y}{3} + \sqrt{\frac{y^2}{9} + 2ay}\right) \\ &\leq e^{-y}\end{aligned}$$

がわかる. 以上から (12.11) は証明された.

(12.12) の証明:

(12.17) の証明:

(12.18) の証明: □

12.3 Talagrand の集中不等式

定理 12.5. $\mathbb{X} \subset \mathbb{R}^d$ とする. P を $\mathbb{X}$ 上の確率測度とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim}$ P とする. $\mathcal{G}$ を $\mathbb{X}$ 上の各点で可測関数の族で, ある定数 $B > 0$ が存在して

$$\|g\|_\infty := \sup_{x \in \mathbb{X}} |f(x)| \leq B$$

とする. さらに, $\mathsf{P}g = 0\,(\forall \in \mathcal{G})$ とする. $\sigma^2\,(\sigma > 0)$ は定数で $\sigma^2 \geq \sup_{g \in \mathcal{G}} Pg^2$ とし

$$\begin{aligned}Z &:= \sup_{g \in \mathcal{G}} \Big|\sum_{j=1}^n g(X_j)\Big|, \\ V &:= n\sigma^2 + 2B\mathsf{E}[Z]\end{aligned}$$

とおく. このとき, 任意の $x > 0$ に対して

$$\mathsf{Pr}\left(Z - \mathsf{E}[Z] \geq \sqrt{2Vx} + \frac{Bx}{3}\right) \leq e^{-x} \tag{12.19}$$

が成り立つ. その結果, 任意の $x > 0$ に対して

$$\mathsf{Pr}\left(Z \geq \sup_{\alpha>0} \inf\left\{(1+\alpha)\mathsf{E}[Z] + \sigma\sqrt{2nx} + \left(\frac{1}{3} + \frac{1}{\alpha}\right)Bx\right\}\right) \leq e^{-x} \tag{12.20}$$

が成立する.

Proof. (12.19) のかわりによりシャープではない不等式

$$\mathsf{Pr}\left(Z - \mathsf{E}[Z] \geq \sqrt{\left(2n\sigma^2 + 16B\mathsf{E}[Z]\right)x} + 2Bx\right) \leq e^{-x} \tag{12.21}$$

を証明する.

各点の可測性により, 一般性を失うことなく, $\mathcal{G}$ を高々可算集合としてよい. さらに, 近似をすることにより, $\mathcal{G}$ を有限集合としてよい. また, $g/\|g\|_\infty$ としても不等式は変わらないので, 一般性を失うことなく, $B = 1$ としてよい. すなわち, $\|g\|_\infty \leq 1\,(\forall g \in \mathcal{G})$ が成立していると仮定してよい.

$\{X_j'\}_{j=1}^n$ を $\{X_j\}_{j=1}^n$ の独立複製とし

$$Z_j := \sup_{g\in\mathcal{G}} \left|\sum_{k=1}^{j-1} g(X_k) + g(X_j') + \sum_{k=j+1}^{n} g(X_k)\right|$$

とおく. さらに, $(X_1, X_2, \ldots, X_n) =: \boldsymbol{X}$ を与えたとき, $\widetilde{g} \in \mathcal{G}$ を Z で最大に達するものとする. すなわち, $\boldsymbol{X}$ を与えたとき

$$Z = \max_{g\in\mathcal{G}} \left|\sum_{j=1}^{n} g(X_j)\right| = \left|\sum_{j=1}^{n} \widetilde{g}(X_j)\right|$$

が成り立つ. $\widetilde{g}$ は $\boldsymbol{X}$ の実現値ごとに決まるので, ランダムな関数となる

ことに注意する. すると

$$
\begin{aligned}
Z - Z_j &= \left|\sum_{k=1}^{n} \widetilde{g}(X_k)\right| - \max_{g\in\mathcal{G}} \left|\sum_{k=1}^{j-1} g(X_k) + g(X'_j) + \sum_{k=j+1}^{n} g(X_k)\right| \\
&= \left|\sum_{k=1}^{n} \widetilde{g}(X_k)\right| - \max_{g\in\mathcal{G}} \left|\sum_{k=1}^{n} g(X_k) + g(X'_j) - g(X_j)\right| \\
&\le \left|\sum_{k=1}^{n} \widetilde{g}(X_k)\right| - \left|\sum_{k=1}^{n} \widetilde{g}(X_k) + \widetilde{g}(X'_j) - \widetilde{g}(X_j)\right| \\
&\qquad \Biggl(\because \max_{g\in\mathcal{G}} \left|\sum_{k=1}^{n} g(X_k) + g(X'_j) - g(X_j)\right| \\
&\qquad\qquad \ge \left|\sum_{k=1}^{n} \widetilde{g}(X_k) + \widetilde{g}(X'_j) - \widetilde{g}(X_j)\right|\Biggr) \\
&\le \left|\sum_{k=1}^{n} \widetilde{g}(X_k)\right| - \left|\sum_{k=1}^{n} \widetilde{g}(X_k)\right| + \left|\widetilde{g}(X'_j) - \widetilde{g}(X_j)\right| \\
&= \left|\widetilde{g}(X'_j) - \widetilde{g}(X_j)\right|
\end{aligned}
$$

と上から評価できる. したがって

$$
\begin{aligned}
V^+ :&= \mathsf{E}\left[\sum_{j=1}^{n} \bigl(Z - Z_j\bigr)^2 \,\middle|\, \boldsymbol{X}\right] \\
&\le \mathsf{E}\left[\sum_{j=1}^{n} \bigl\{\widetilde{g}(X_j) - \widetilde{g}(X'_j)\bigr\}^2 \,\middle|\, \boldsymbol{X}\right] \\
&= \mathsf{E}\left[\sum_{j=1}^{n} \widetilde{g}^2(X_j) \,\middle|\, \boldsymbol{X}\right] - 2\sum_{j=1}^{n} \bigl\{\widetilde{g}(X_j) \underbrace{\mathsf{E}\left[\widetilde{g}(X'_j) \,\middle|\, \boldsymbol{X}\right]}_{=0} + \sum_{j=1}^{n} \underbrace{\mathsf{E}\left[\widetilde{g}^2(X'_j) \,\middle|\, \boldsymbol{X}\right]}_{\le\sigma^2} \\
&\le \sum_{j=1}^{n} \widetilde{g}^2(X_j) + n\sigma^2 \\
&\le \max_{g\in\mathcal{G}} \sum_{j=1}^{n} g^2(X_j) + n\sigma^2 \\
&=: W + v
\end{aligned}
\tag{12.22}
$$

となる.

ここで, 定理 12.4 の Z を W として定理 **??** を用いるために, 以下の記号を導入する. $\boldsymbol{X}$ を与えたとき, $\check{g}$ を $\mathcal{G}$ のなかで W を最大にするものとする. すなわち

$$
\max_{g\in\mathcal{G}} \sum_{j=1}^{n} g^2(X_j) = \sum_{j=1}^{n} \check{g}^2(X_j)
$$

である. さらに, $j = 1, 2, \ldots, n$ に対して

$$W_j := \max_{g \in \mathcal{G}} \left\{ \sum_{k \neq j}^{j-1} g^2(X_k) \right\}$$

とおく. すると

$$\begin{aligned}
\sum_{k=1}^{n} g^2(X_j) \geq \sum_{k \neq j} g^(X_k) \;\Rightarrow\; \max_{g \in \mathcal{G}} \sum_{k=1}^{n} g^2(X_j) \geq \max_{g \in \mathcal{G}} \sum_{k \neq j} g^(X_k) \\
\Leftrightarrow\; W - W_j \geq 0
\end{aligned}$$

となる. さらに, $\boldsymbol{X}$ を条件付きにすると

$$\begin{aligned}
W - W_j &= \max_{g \in \mathcal{G}} \sum_{k=1}^{n} g^2(X_j) - \max_{g \in \mathcal{G}} \sum_{k \neq j} g^2(X_j) \\
&= \max_{g \in \mathcal{G}} \sum_{k=1}^{n} g^2(X_j) - \max_{g \in \mathcal{G}} \left\{ \sum_{k=1}^{n} g^2(X_j) - g^2(X_j) \right\} \\
&\qquad \left(\because \max_{g \in \mathcal{G}} \left\{ \sum_{k=1}^{n} g^2(X_j) - g^2(X_j) \right\} \geq \sum_{k=1}^{n} \check{g}^2(X_j) + \check{g}^2(X_j) \right) \\
&\leq \max_{g \in \mathcal{G}} \sum_{k=1}^{n} \check{g}^2(X_j) - \sum_{k=1}^{n} \check{g}^2(X_j) + \check{g}^2(X_j)
\end{aligned}$$

となる. これより

$$\sum_{j=1}^{n} (W - W_j) \leq \sum_{j=1}^{n} \check{g}^2(X_j) = W$$

となる. 以上から, W に対して定理 12.4 を適用できるので

$$\log \mathsf{E}\big[e^{\lambda W}\big] \leq (e^{\lambda} - 1)\mathsf{E}[W] \qquad (\lambda \in \mathbb{R}) \tag{12.23}$$

を得る. つぎに, (12.22) と (12.23) を用いてから, 定理 12.1 を $\theta = 1$ として適用すると, $\lambda \in (0, 1)$ に対して

$$\begin{aligned}
\log \mathsf{E}\Big[\exp\big(\lambda(Z - \mathsf{E}[Z])\big)\Big] &\leq \frac{\lambda}{1 - \lambda} \log \mathsf{E}\big[\exp(\lambda V^{+})\big] \\
&\leq \frac{\lambda}{1 - \lambda} \log \mathsf{E}\big[\exp\{\lambda(W + v)\}\big] \\
&\leq \frac{\lambda}{1 - \lambda} \Big\{ \log \mathsf{E}\big[e^{\lambda W}\big] + \lambda v \Big\} \\
&\leq \frac{\lambda}{1 - \lambda} \Big\{ (e^{\lambda} - 1)\mathsf{E}[W] + \lambda v \Big\}
\end{aligned}$$

を得る．ここで不等式 $(e^{\lambda}-1)(1-\lambda) \le \lambda\,(\lambda \ge 0)$ が成り立つこと[3]に注意すると, $\lambda \in (0,\,1/2)$ に対して

$$
\begin{aligned}
\log \mathsf{E}\Big[\exp\big(\lambda(Z-\mathsf{E}[Z])\big)\Big] &= \frac{\lambda}{(1-\lambda)^2}\Big\{\underbrace{\lambda(1-\lambda)}_{\le\lambda}v + \underbrace{(e^{\lambda}-1)(1-\lambda)}_{\le\lambda}\mathsf{E}[W]\Big\} \\
&\le \frac{\lambda^2}{\underbrace{(1-\lambda)^2}_{\ge 1-2\lambda}}\big\{v+\mathsf{E}[W]\big\} \\
&\le \frac{\lambda^2}{1-2\lambda}\big\{v+\mathsf{E}[W]\big\}
\end{aligned}
$$

を得る．これらのことを踏まえて, Markov の不等式を用いると, 任意の $x>0$ と $\lambda \in (0,\,1/2)$ に対して

$$
\begin{aligned}
\mathsf{Pr}\Big(Z-\mathsf{E}[Z] \ge x\Big) &= \mathsf{Pr}\Big(e^{\lambda(Z-\mathsf{E}[Z])} \ge e^{\lambda x}\Big) \\
&= e^{-\lambda x}\mathsf{E}\Big[e^{\lambda(Z-\mathsf{E}[Z])} \ge e^{\lambda x}\Big] \\
&\le \exp\Big\{-\lambda x + \frac{\lambda^2}{1-2\lambda}\big\{v+\mathsf{E}[W]\big\}\Big\} \\
&=: \exp\big\{h(\lambda)\big\}
\end{aligned}
$$

となる．h を λ に関して微分すると

$$
\overset{\bullet}{h}(\lambda) = -x + \frac{2\lambda}{1-2\lambda}\big\{v+\mathsf{E}[W]\big\} + \frac{2\lambda^2}{(1-2\lambda)^2}\big\{v+\mathsf{E}[W]\big\}
$$

から

$$
\begin{aligned}
(1-2\lambda)^2\,\overset{\bullet}{h}(\lambda) &= -(1-2\lambda)^2 x + \big\{2\lambda(1-2\lambda)+2\lambda^2\big\}\big\{v+\mathsf{E}[W]\big\} \\
&= -2\{v+\mathsf{E}[W]+2x\}\Big(\lambda-\frac{1}{2}\Big)^2 + \frac{v+\mathsf{E}[W]}{2} \\
&= 0 \;\Leftrightarrow\; \lambda = \frac{1}{2} \pm \sqrt{\frac{v+\mathsf{E}[W]}{4\{v+\mathsf{E}[W]+2x\}}}
\end{aligned}
$$

となる．よって

$$
\lambda = \frac{1}{2}\Bigg(1-\sqrt{\frac{v+\mathsf{E}[W]}{v+\mathsf{E}[W]+2x}}\Bigg) =: \lambda^*
$$

[3] $h(\lambda) := \lambda - (1-\lambda)(e^{\lambda}-1)\,(\lambda \ge 0)$ とおくと, $\overset{\bullet}{h}(\lambda) = 1 + (e^{\lambda}-1) - (1-\lambda)e^{\lambda} = \lambda e^{\lambda}\,(\lambda \ge 0)$ となる．$h(0)=0$ なので, $h(\lambda) \ge 0\,(\lambda \ge 0)$ がわかる．よって, 不等式は証明された.

は $\overset{\bullet}{h}(\lambda) = 0$ の解で $0 < \lambda^* < 1/2$ となる．よって, 関数 $(0, 1/2) \ni \lambda \mapsto \mathbb{R}$ は $\lambda = \lambda^*$ で最小となる．$a = v + \mathsf{E}[W]$ とおいて, 単純な計算をすると

$$h(\lambda^*) = \frac{-x - a + \sqrt{a^2 + 2ax}}{2}$$

を得る．したがって

$$\mathsf{Pr}\Big(Z - \mathsf{E}[Z] \geq x\Big) \leq \exp\bigg\{-\frac{x + a - \sqrt{a^2 + 2ax}}{2}\bigg\}$$

を得る．ここで

$$2y = x + a - \sqrt{a^2 + 2ax}\,(x > 0) \;\Leftrightarrow\; x = 2y + 2\sqrt{(v + \mathsf{E}[W])y}$$

とおくと

$$\mathsf{Pr}\Big(Z - \mathsf{E}[Z] \geq 2\sqrt{(v + \mathsf{E}[W])y} + 2y\Big) \leq e^{-y} \tag{12.24}$$

を得る．さらに, 系 6.27 を用いてから, 補題 6.22(6.26) を用いると

$$\begin{aligned}
\mathsf{E}[W] &= \mathsf{E}\bigg[\max_{g\in\mathcal{G}} \sum_{j=1}^{n} g^2(X_j)\bigg] \\
&\leq \underbrace{n\sigma^2}_{=v\,\because(12.22)} + 8\mathsf{E}\bigg[\max_{g\in\mathcal{G}}\Big|\epsilon_j g(X_j)\Big|\bigg] \\
&\leq v + 16\mathsf{E}\bigg[\max_{g\in\mathcal{G}}\Big|g(X_j)\Big|\bigg]
\end{aligned}$$

を得る．これを (12.24) に代入すると

$$\mathsf{Pr}\Big(Z - \mathsf{E}[Z] \geq 2\sqrt{(2v + 16\mathsf{E}[Z])y} + 2y\Big) \leq e^{-y}$$

を得る．よって, (12.21) は証明された． □

12.4 Talagrand の不等式の統計学版

12.5 Fuk-Nagaev 形の不等式