Rademacher 複雑度と正則化 ver.1

楠岡成雄

1　初めに

$(E,\mathcal{E})$ は可測空間、Θ は距離空間とする。Θ をパラメータの集合と考える。$F: E\times\Theta\to\mathbf{R}$ は可測関数であり、$F(x,\cdot):\Theta\to\mathbf{R}$, $x\in E$, は連続とする。E-値確率変数 X に対して

$$F_0(\theta)=E[F(X,\theta)],\qquad \theta\in\Theta$$

とおき、$F_0(\theta)$ を最小とする $\theta_0\in\Theta$ を見つけるという問題を考える。このような問題は機械学習（教師付学習、強化学習）などにしばしば現れる。しかし、多くの場合、X の確率法則が不明である、あるいはわかっていても $F_0(\theta)$ の計算が容易でないといった理由により θ_0 を見つけることは困難であることが多い。

この問題を解決する方法として X の確率法則と同じ確率法則を持つ独立な E-値確率変数 $X_1,X_2,\ldots,X_n$ が与えられており得て、

$$\hat{F}_n(\theta)=\frac{1}{n}\sum_{i=1}^n F(X_i,\theta)$$

は計算可能として考えることがある。この時、

$$\hat{\theta}_0=\mathrm{argmin}_{\theta\in\Theta}\hat{F}_n(\theta)$$

を θ_0 の「推定値」として良いであろうか？

今、$F(x,\theta)$ の例として以下のような線形モデルを考える

$g_i: E\to\mathbf{R}$, $i=1,\ldots,m$, は可測関数、$\Theta=\mathbf{R}^m$ とし、$L:\mathbf{R}\times E\to[0,\infty)$ が与えられており、

$$F(x,\theta)=L(\sum_{i=1}^m\theta_i g_i(x),x),\qquad x\in E,\ \theta\in\mathbf{R}$$

とする。もし $m>n$ であれば

$$\hat{F}_n(\hat{\theta})=\frac{1}{n}\sum_{i=1}^n F(X_i,\hat{\theta})=\min_{\theta\in\mathbf{R}^m}\hat{F}_n(\theta)$$

となる $\hat{\theta}$ が無限に存在する。これが、Overfittig や過学習と呼ばれる問題である。このような問題は、まず最尤推定量に関する問題として現れた。この問題に対する赤池さんの答え (AIC) は「パラメータの次元は観測数 n に対してあまり大きくなってはならない」ということである。しかし近年の「機械学習」においてはパラメータの次元 $>>n$ となることが普通である。そのために正則化という概念が生まれた。

正規化とは、正則化関数 $\phi:\Theta\to[0,\infty)$ 及び $\lambda>0$ を選び

$$\hat{F}_n(\theta)+\lambda\phi(\theta)$$

を最小にする θ を選ぶという考えである。

正則化関数の例としては以下のようなものがある。

(L^1 正則化)　$\phi(\theta)=||\theta||_1=\sum_{i=1}^m|\theta_i|$

(L^2 正則化) $\phi(\theta)=||\theta||_2=(\sum_{i=1}^m|\theta_i|^2)^{1/2}$ 　または $\phi(\theta)=||\theta||_2^2=\sum_{i=1}^m|\theta_i|^2$

ここで以下のような疑問が生ずる。

(1) 正則化の役割は overfittig を回避するだけか?

(2) どのような正則化関数を用いるべきか?

(3) λ をどこまで小さくとれるのか？

これらに対する一つの解答を与えるのが本ノートの目的である。

なお参考文献として　金森敬文「統計的学習理論」講談社　MLP　2015 がある。

2　準備

命題 2.1 $p\in[0,1]$ とする。この時

$$p\exp((1-p)u)+(1-p)\exp(-pu)\leqq\exp(u^2/8),\qquad u\in\mathbf{R}$$

が成立する。

証明. $f(u)=p\exp((1-p)u)+(1-p)\exp(-pu)$ とおくと

$$f(u)=\exp(-pu)(1-p+p\exp(u)),$$

$$f'(u)=p(1-p)(\exp((1-p)u)-\exp(-pu))=\exp(-pu)p(1-p)(-1+\exp(u)),$$

$$f''(u)=p(1-p)((1-p)\exp((1-p)u)+p\exp(-pu))=\exp(-pu)p(1-p)(p+(1-p)\exp(u)).$$

よって $f(0)=1,\ f'(0)=0,\ f(u)>0$ であるので、$g(u)=\log f(u)$ とおくと

$$g(0)=0,\qquad g'(0)=f(0)^{-1}f'(0)=0,$$

$$g''(u)=f(u)^{-1}f''(u)-f(u)^{-2}(f'(u))^2.$$

$$\begin{aligned}&f''(u)f(u)-(f'(u))^2\\&=\exp(-2pu)\{p(1-p)(p+(1-p)\exp(u))(1-p+p\exp(u))-p^2(1-p)^2(-1+\exp(u))^2\}\\&=\exp(-2pu)p(1-p)(p^2+(1-p)^2+2p(1-p))\exp(u)=\exp(-2pu)p(1-p)\exp(u)\end{aligned}$$

であるので

$$\begin{aligned}g''(u)&=\frac{p(1-p)\exp(u)}{(1-p+p\exp(u))^2}=\frac{(1-p)p\exp(u)}{(1-p+p\exp(u))^2}\\&=\frac{1-p}{1-p+p\exp(u)}\frac{p\exp(u)}{1-p+p\exp(u)}=\frac{1-p}{1-p+p\exp(u)}(1-\frac{1-p}{1-p+p\exp(u)})\leqq\frac{1}{4}.\end{aligned}$$

よって、

$$g(u)=\int_0^1(1-t)\frac{d^2}{dt^2}g(tu)dt\leqq\frac{u^2}{8}.$$

∎

命題 2.2 (Hoeffding's lemma) $a,b\in\mathbf{R}$ とする。X が確率変数で $a\leqq X\leqq b\ a.s.$ かつ $E[X]=0$ であれば

$$E[\exp(tX)]\leqq\exp(t^2(b-a)^2/8),\qquad t\in\mathbf{R}.$$

証明. $\exp(tx)$ は x の凸関数であるので

$$\exp(tx) \leqq \frac{x-a}{b-a}e^{tb} + \frac{b-x}{b-a}e^{ta} \qquad x \in [a,b]$$

が成立する。よって

$$E[\exp(tX)] \leqq \frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta}$$

$p = (-a)/(b-a) \in [0,1]$ とおくと、命題 2.1 より

$$\frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta} = p\exp(t(b-a)(1-p)) + (1-p)\exp(-t(b-a)p) \leqq \exp(t^2(b-a)^2/8).$$

∎

命題 2.3 (1) X は確率変数で $E[\exp(X^2)] < \infty,\ E[X] = 0$ とする。この時、

$$\sup_{t\neq 0} \frac{1}{t^2}\log E[\exp(tX)] \leqq \log E[\exp(X^2)] + \frac{1}{2}$$

が成り立つ。

(2) X は確率変数、$\gamma_0 > 0$ で $E[\exp(\gamma_0^{-1}X^2)] < \infty,\ E[X] = 0$ が成り立つとする。この時、

$$\sup_{t\neq 0} \frac{1}{t^2}\log E[\exp(tX)] \leqq \gamma_0(\log E[\exp(\gamma_0^{-1}X^2)] + \frac{1}{2})$$

が成り立つ。

証明. (1) を示す。$\varphi(t) = \log E[\exp(tX)],\ t \in \mathbf{R}$ とおくと、

$$\varphi(0) = 0,$$

$$\varphi'(t) = \frac{E[X\exp(tX)]}{E[\exp(tX)]}, \qquad t \in \mathbf{R},$$

$$\varphi''(t) = \frac{E[X^2\exp(tX)]}{E[\exp(tX)]} - (\frac{E[X\exp(tX)]}{E[\exp(tX)]})^2 \quad \geqq 0, \qquad t \in \mathbf{R},$$

が成り立つ。

$\varphi'(0) = E[X] = 0$ であるので

$$\varphi(t) = \int_0^t ds_1(\int_0^{s_1}\varphi''(s_2)ds_2)$$

となる。また、

$$0 \leqq \varphi''(t) \leqq \frac{E[X^2\exp(tX)]}{E[\exp(tX)]}$$

となる。

これより、

$$\varphi''(t) \leqq \frac{E[X^2\exp(tX)]}{E[\exp(tX)]} \leqq 2\log(\frac{E[\exp(\frac{1}{2}X^2)\exp(tX)]}{E[\exp(tX)]})$$

となる。

$$E[\exp(\frac{1}{2}X^2)\exp(tX)] = E[\exp(X^2 - \frac{1}{2}(X-t)^2 + \frac{t^2}{2})] \leqq E[\exp(X^2)]\exp(\frac{t^2}{2}).$$

よって、

$$\varphi``(t) \leqq 2\log E[\exp(X^2)] + t^2, \qquad t \in (0,1],$$

であり

$$\frac{1}{t^2}\log E[\exp(tX)] = \frac{1}{t^2}\varphi(t) \leqq \log E[\exp(X^2)] + \frac{1}{2}, \qquad t \in (0,1],$$

となる。また、

$$\frac{1}{t^2}\log E[\exp(tX)] = \frac{1}{t^2}\log E[\exp(X^2-(X-\frac{t}{2})^2+\frac{t^2}{2})] \leqq \log E[\exp(X^2)]+\frac{1}{2}, \qquad t \geqq 1,$$

である。よって、

$$\sup_{t>0}\frac{1}{t^2}\log E[\exp(tX)] \leqq \log E[\exp(X^2)]+\frac{1}{2}$$

を得る。これより主張 (1) を得る。

(2) は

$$\sup_{t\neq 0}\frac{1}{t^2}\log E[\exp(tX)] \leqq \sup_{t\neq 0}\frac{\gamma_0}{t^2}\log E[\exp(t\gamma_0^{-1/2}X)]$$

よりわかる。

∎

3 Rademacher 複雑度

$(E,\mathcal{E})$ を可測空間、$\mathcal{M}(E)$ を E 上の可測関数全体とする。

定義 3.1 (1) $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とする。$\hat{\mathcal{R}}_n(\cdot;\mathcal{K}) : E^n \to [0,\infty)$, $n \geqq 1$, を

$$\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}) = \frac{1}{n}E[\sup_{g\in\mathcal{K}}\sum_{k=1}^n g(x_k)\sigma_k], \qquad (x_1,\ldots,x_n)\in E^n$$

により定義する。ここで、$\sigma_1,\sigma_2,\ldots$ は独立同分布な確率変数列で $P(\sigma_k=\pm 1)=1/2$ となるものである。
(2) $\mathcal{G}$ を $\mathcal{M}(E)$ の空でない部分集合とし、ν を $(E,\mathcal{E})$ 上の確率測度とする。$\mathcal{R}_n(\mathcal{G},\nu)\in[0,\infty]$, $n\geqq 1$, を

$$\mathcal{R}_n(\mathcal{G},\nu) = \int_{E^n}\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{G})\nu^{\otimes n}(dx_1,\ldots dx_n)$$

により定義する。

命題 3.2 (1) $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とする。この時、$a>0$ に対して

$$\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\{ag;\ g\in\mathcal{K}\}) = a\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}).$$

また、$-\mathcal{K}=\{-g;\ g\in\mathcal{K}\}$ とおくと、

$$\hat{\mathcal{R}}_n(x_1,\ldots,x_n;-\mathcal{K}) = \hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K})$$

である。さらに、$g_0\in\mathcal{M}(E)$ に対して $\mathcal{K}+g_0=\{g+g_0;\ g\in\mathcal{K}\}$ とおくと、

$$\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}+g_0) = \hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K})$$

である。
(2) $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とする。この時、

$$\hat{\mathcal{R}}_n(x_1,\ldots,x_n;conv(\mathcal{K})) = \hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}), \qquad n\geqq 1,\ (x_1,\ldots,x_n)\in E^n$$

である。ここで $conv(\mathcal{K})$ は $m\geqq 1$, $\lambda_i\in(0,1], i=1,\ldots,m$, $\sum_{i=1}^m\lambda_i=1$, $g_i\in\mathcal{K}\ i=1,\ldots,m$, が存在して、$g=\sum_{i=1}^m\lambda_i g_i$ となる g の集合である。
(3) $\mathcal{K}_i$, $i=1,2$, を $\mathcal{M}(E)$ の空でない部分集合とする。この時、

$$\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}_1\cup\mathcal{K}_2\cup\{0\}) \leqq \hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}_1\cup\{0\})+\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}_2\cup\{0\}),$$

$n \geqq 1,\ (x_1, \ldots, x_n) \in E^n$ である。

(4) $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とする。今、$\mathcal{G}$ を $m \geqq 1,\ a_i \in \mathbf{R}, i = 1, \ldots, m,\ \sum_{i=1}^m |a_i| \leqq 1,$ $g_i \in \mathcal{K},\ i = 1, \ldots, m,$ が存在して、$g = \sum_{i=1}^m a_i g_i$ となる g の集合とする。

この時、

$$\hat{\mathcal{R}}_n(x_1, \ldots, x_n; \mathcal{G}) \leqq 2\hat{\mathcal{R}}_n(x_1, \ldots, x_n; \mathcal{K} \cup \{0\}) \qquad n \geqq 1,\ (x_1, \ldots, x_n) \in E^n$$

である。

(5) $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とし、$g_0 \in \mathcal{K}$ とする。この時、

$$\hat{\mathcal{R}}_n(x_1, \ldots, x_n; \mathcal{K} \cup \{0\}) \leqq \hat{\mathcal{R}}_n(x_1, \ldots, x_n; \mathcal{K}) + \frac{1}{n}(\sum_{k=1}^n g_0(x_k)^2)^{1/2}.$$

証明. (1) は

$$\sup_{g \in \mathcal{K}} \sum_{k=1}^n a g(x_k)\sigma_k = a \sup_{g \in \mathcal{K}} \sum_{k=1}^n g(x_k)\sigma_k,$$

$$\sup_{g \in \mathcal{K}} \sum_{k=1}^n (-g(x_k))\sigma_k = \sup_{g \in \mathcal{K}} \sum_{k=1}^n g(x_k)(-\sigma_k),$$

$$\sup_{g \in \mathcal{K}} \sum_{k=1}^n (g(x_k) + g_0(x_k))\sigma_k = \sup_{g \in \mathcal{K}} \sum_{k=1}^n g(x_k)\sigma_k + \sum_{k=1}^n g_0(x_k)\sigma_k$$

よりわかる。

(2) は $m \geqq 1,\ \lambda_i \in (0, 1], i = 1, \ldots, m,\ \sum_{i=1}^m \lambda_i = 1,\ g_i \in \mathcal{K},\ i = 1, \ldots, m,\ g = \sum_{i=1}^m \lambda_i g_i$ ならば

$$\sum_{k=1}^n g(x_k)\sigma_k = \sum_{i=1}^m \lambda_i (\sum_{k=1}^n g_i(x_k)\sigma_k) \leqq \sup_{h \in \mathcal{K}} \sum_{k=1}^n h(x_k)\sigma_k$$

よりわかる。

(3) は

$$\sup_{g \in \mathcal{K}_i \cup \{0\}} \sum_{k=1}^n g(x_k)\sigma_k = \sup_{g \in \mathcal{K}_i} (\sum_{k=1}^n g(x_k)\sigma_k) \vee 0, \qquad i = 1, 2,$$

であるので、

$$\sup_{g \in \mathcal{K}_1 \cup \mathcal{K}_2 \cup \{0\}} \sum_{k=1}^n g(x_k)\sigma_k \leqq \sup_{g \in \mathcal{K}_1} (\sum_{k=1}^n g(x_k)\sigma_k) \vee 0 + \sup_{g \in \mathcal{K}_2} (\sum_{k=1}^n g(x_k)\sigma_k) \vee 0$$

よりわかる。

(4) を示す。$g \in \mathcal{G}$ ならば $m \geqq 1,\ g_i \in \mathcal{K}_i,\ i = 1, \ldots, m,\ a_i \in \mathbf{R},\ \sum_{i=1}^m |a_i| \leqq 1$ が存在して $g = \sum_{i=1}^m a_i g_i$ となる。$h_i,\ i = 1, \ldots, m+1,$ を、$a_i \geqq 0,\ i \leqq m,$ ならば $h_i = g_i,$ また $a_i < 0,\ i \leqq m,$ ならば $h_i = -g_i,$ $h_{m+1} = 0,$ で定める。この時、

$$g = \sum_{i=1}^m |a_i| h_i + (1 - \sum_{i=1}^m |a_i|) h_{m+1}$$

となるので、$g \in conv(\mathcal{K} \cup (-\mathcal{K}) \cup \{0\})$ がわかる。よって、$\mathcal{G} \subset conv(\mathcal{K} \cup (-\mathcal{K}) \cup \{0\})$. これより (4) がわかる。

(5) を示す。

$$\sup_{g \in \mathcal{K} \cup \{0\}} \sum_{k=1}^n g(x_k)\sigma_k = \sup_{g \in \mathcal{K}} ((\sum_{k=1}^n g(x_k)\sigma_k) \vee 0)$$

$$= \sup_{g \in \mathcal{K}} (\sum_{k=1}^n (g(x_k) - g_0(x_k))\sigma_k + \sum_{k=1}^n g_0(x_k)\sigma_k) \vee 0$$

$$\leqq \sup_{g\in\mathcal{K}}((\sum_{k=1}^n (g(x_k)-g_0(x_k))\sigma_k)\vee 0+(\sum_{k=1}^n g_0(x_k)\sigma_k)\vee 0$$

$$=\sup_{g\in\mathcal{K}}((\sum_{k=1}^n (g(x_k)-g_0(x_k))\sigma_k)+(\sum_{k=1}^n g_0(x_k)\sigma_k)-(\sum_{k=1}^n g_0(x_k)\sigma_k))\wedge 0$$

$$=\sup_{g\in\mathcal{K}}(\sum_{k=1}^n g(x_k)\sigma_k)-(\sum_{k=1}^n g_0(x_k)\sigma_k)\wedge 0$$

であるので、

$$\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}\cup\{0\})\leqq \hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K})+\frac{1}{n}E[(\sum_{k=1}^n g_0(x_k)(-\sigma_k))\vee 0]$$

$$\leqq \hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K})+\frac{1}{n}E[|\sum_{k=1}^n g_0(x_k)\sigma_k|]$$

を得る。

$$E[|\sum_{k=1}^n g_0(x_k)\sigma_k|]\leqq E[|\sum_{k=1}^n g_0(x_k)\sigma_k|^2]^{1/2}=(\sum_{k=1}^n g_0(x_k)^2)^{1/2}$$

であるので (5) を得る。

∎

命題 3.3 (Talagrand) $\mathcal{K}$ は $\mathcal{M}(E)$ の空でない部分集合とし、$\rho:\mathbf{R}\times E\to\mathbf{R}$ は $|\rho(s,x)-\rho(t,x)|\leqq|t-s|$, $s,t\in\mathbf{R},x\in E,$ を満たす可測関数とする。いま、

$$\mathcal{K}\text{'}=\{\rho(g(\cdot),\cdot);\ g\in\mathcal{K}\}$$

で定める。この時、

$$\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}')\leqq\hat{\mathcal{R}}_n(x_1,\ldots,x_n;\mathcal{K}),\qquad n\geqq 1,\ (x_1,\ldots,x_n)\in E^n$$

が成立する。

証明. $\sigma_1,\sigma_2,\ldots$ は独立同分布な確率変数列で $P(\sigma_k=\pm 1)=1/2$ となるものとする。いま、$n\geqq 1$ を固定し、$g\in\mathcal{K}$ に対して確率変数 $u_m(g)$, $m=1,\ldots,n,$ を

$$u_m(g)=\sum_{k=1}^{m-1} g(x_k)\sigma_k+\sum_{k=m+1}^n \rho(g(x_k),x_k)\sigma_k$$

で定める。$u_m(g)$ は $\sigma\{\sigma_k;\ 1\leqq k\leqq n,\ k\neq m\}$-可測である。この時、

$$E[\sup_{g\in\mathcal{K}}(\sum_{k=1}^{m-1} g(x_k)\sigma_k+\sum_{k=m}^n \rho(g_k(x_k),x_k)\sigma_k)]$$

$$=E[E[\sup_{g\in\mathcal{K}}(u_m(g)+\rho(g(x_m),x_k)\sigma_m)|\sigma_1,\ldots,\sigma_{m-1},\sigma_{m+1},\ldots,\sigma_n]]$$

$$=\frac{1}{2}E[\sup_{g\in\mathcal{K}}(u_m(g)+\rho(g(x_m),x_m))]+\frac{1}{2}E[\sup_{g\in\mathcal{K}}(u_m(g)-\rho(g(x_m),x_m)]$$

である。任意の $\varepsilon>0$ 及び $\sigma_1,\ldots,\sigma_{m-1},\sigma_{m+1},\ldots,\sigma_n$ に対して $f_0,f_1\in\mathcal{K}$ で

$$\sup_{g\in\mathcal{K}}(u_m(g)+\rho(g(x_m),x_m))\leqq u_m(f_0)+\rho(f_0(x_m),x_m)+\varepsilon$$

$$\sup_{g\in\mathcal{K}}(u_m(g)-\rho(g(x_m),x_m)) \leqq u_m(f_1)-\rho(f_1(x_m),x_m)+\varepsilon$$

となるものが存在する。

いま、c を

$$c=\mathrm{sign}(f_0(x_m)-f_1(x_m))\in\{-1,1\}$$

とおくと、

$$\rho(f_0(x_m),x_m)-\rho(f_1(x_m),x_m)\leqq c(f_0(x_m)-f_1(x_m))$$

となる。この時

$$\begin{aligned}
&\sup_{g\in\mathcal{K}}(u_m(g)+\rho(g(x_m),x_m))+\sup_{g\in\mathcal{K}}(u_m(g)-\rho(g(x_m),x_m))\\
&\leqq u_m(f_0)+\rho(f_0(x_m),x_m)+u_m(f_1)-\rho(f_1(x_m),x_m)+2\varepsilon\\
&\leqq u_m(f_0)+u_m(f_1)+cf_0(x_m)-cf_1(x_m)+2\varepsilon\\
&\leqq \sup_{g\in\mathcal{K}}(u_m(g)+g(x_m))+\sup_{g\in\mathcal{K}}(u_m(g)-g(x_m))+2\varepsilon.
\end{aligned}$$

ε は任意なので

$$\begin{aligned}
&\sup_{g\in\mathcal{K}}(u_m(g)+\rho(g(x_m),x_m))+\sup_{g\in\mathcal{K}}(u_m(g)-\rho(g(x_m),x_m))\\
&\leqq \sup_{g\in\mathcal{K}}(u_m(g)+g(x_m))+\sup_{g\in\mathcal{K}}(u_m(g)-g(x_m))
\end{aligned}$$

を得る。よって、$m=1,\ldots,n$ に対して

$$\begin{aligned}
&E[\sup_{g\in\mathcal{K}}(\sum_{k=1}^{m-1}g(x_k)\sigma_k+\sum_{k=m}^{n}\rho(g(x_k),x_k)\sigma_k)]\\
&\leqq E[\sup_{g\in\mathcal{K}}(\sum_{k=1}^{m}g(x_k)\sigma_k+\sum_{k=m+1}^{n}\rho(g(x_k),x_k)\sigma_k)]
\end{aligned}$$

となる。これより主張を得る。

∎

命題 3.4 ν を $(E,\mathcal{E})$ 上の確率測度とし、$\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とする。

(1) $a>0$ に対して

$$\mathcal{R}_n(\{ag;\ g\in\mathcal{K}\};\nu)=a\mathcal{R}_n(\mathcal{K};\nu).$$

また、$-\mathcal{K}=\{-g;\ g\in\mathcal{K}\}$ とおくと、

$$\mathcal{R}_n(-\mathcal{K};\nu)=\mathcal{R}_n(\mathcal{K};\nu)$$

である。

さらに、$g_0\in L^1(E,d\nu)$ に対して $\mathcal{K}+g_0=\{g+g_0;\ g\in\mathcal{K}\}$ とおくと、

$$\hat{\mathcal{R}}_n(\mathcal{K}+g_0;\nu)=\hat{\mathcal{R}}_n(\mathcal{K};\nu)$$

である。

(2) $\mathcal{R}_n(conv(\mathcal{K});\nu)=\mathcal{R}_n(\mathcal{K};\nu),\quad n\geqq 1.$

(3) $\mathcal{K}_i,\ i=1,2,$ を $\mathcal{M}(E)$ の空でない部分集合とする。この時、

$$\mathcal{R}_n(\mathcal{K}_1\cup\mathcal{K}_2\cup\{0\};\nu)\leqq\mathcal{R}_n(\mathcal{K}_1\cup\{0\};\nu)+\mathcal{R}_n(\mathcal{K}_2\cup\{0\};\nu),\qquad n\geqq 1$$

である。
(4) 今、$\mathcal{G}$ を $m \geqq 1$, $a_i \in \mathbf{R}, i = 1, \ldots, m$, $\sum_{i=1}^m |a_i| \leqq 1$, $g_i \in \mathcal{K}$ $i = 1, \ldots, m$, が存在して、$g = \sum_{i=1}^m a_i g_i$ となる g の集合とする。
　この時、

$$\mathcal{R}_n(\mathcal{G};\nu) \leqq 2\mathcal{R}_n(\mathcal{K} \cup \{0\}) \qquad n \geqq 1,$$

である。
(5) $\mathcal{K} \cap L^2(E; d\nu) \neq \emptyset$ とする。この時、

$$\mathcal{R}_n(\mathcal{K} \cup \{0\};\nu) \leqq \mathcal{R}_n(\mathcal{K};\nu) + \frac{1}{n^{1/2}} \inf\{||g||_{L^2(E;d\nu)};\ g \in \mathcal{K} \cap L^2(E; d\nu)\}.$$

(6) $\rho : \mathbf{R} \times E \to \mathbf{R}$ は $|\rho(s,x) - \rho(t,x)| \leqq |t - s|$, $s, t \in \mathbf{R}, x \in E$, を満たす可測関数とする。いま、

$$\mathcal{G} = \{\rho(g(\cdot), \cdot);\ g \in \mathcal{K}\}$$

で定める。この時、

$$\mathcal{R}_n(\mathcal{G};\nu) \leqq \mathcal{R}_n(\mathcal{K};\nu), \qquad n \geqq 1$$

が成立する。
(7) $\mathcal{K} \subset L^1(E; d\nu)$ であり、$\bar{\mathcal{K}}$ を $\mathcal{K}$ の $L^1(E; d\nu)$ での閉包とする。この時、

$$\mathcal{R}_n(\bar{\mathcal{K}};\nu) = \mathcal{R}_n(\mathcal{K};\nu)$$

である。

証明. 主張 (1),(2),(3),(4) は命題 3.2 よりわかる。また、命題 3.2 (5) より $g_0 \in \mathcal{K} \cap L^2(E; d\nu)$ ならば

$$\mathcal{R}_n(\mathcal{K} \cup \{0\},\nu) \leqq \mathcal{R}_n(\mathcal{K},\nu) + \frac{1}{n}(\int_{E^n} (\sum_{k=1}^n g_0(x_k)^2)^{1/2} \nu^{\otimes n}(dx_1 \ldots dx_n)$$

$$\leqq \mathcal{R}_n(\mathcal{K},\nu) + \frac{1}{n}(\int_{E^n} (\sum_{k=1}^n g_0(x_k)^2) \nu^{\otimes n}(dx_1 \ldots dx_n))^{1/2}$$

$$= \mathcal{R}_n(\mathcal{K},\nu) + \frac{1}{n^{1/2}}(\int_E g_0(x)^2 \nu(dx))^{1/2}.$$

これより主張 (5) を得る。
　主張 (6) は命題 3.3 よりわかる。主張 (7) を示す。$g_0 \in \bar{\mathcal{K}}$ ならば、$g_n \in \mathcal{K}$, $n = 1, 2, \ldots,$ が存在して、$||g_0 - g_m||_{L^1(E,d\nu)} \to 0$, $m \to \infty$, となる。
　$X_1, \ldots, X_n$ は E-値確率変数列、$\sigma_1, \ldots, \sigma_N$ は $\{-1, 1\}$-値確率変数であり、$X_1, \ldots, X_n, \sigma_1, \ldots, \sigma_n$ は独立で X_i, $i = 1, \ldots, n$, の確率法則は ν, $P(\sigma_i = \pm 1) = 1/2$, $i = 1, \ldots, n$, とすると

$$E[|\frac{1}{n}\sum_{k=1}^n g_0(X_k)\sigma_k - \frac{1}{n}\sum_{k=1}^n g_m(X_k)\sigma_k|]$$

$$\leqq \frac{1}{n} E[\sum_{k=1}^n |g_0(X_k) - g_m(X_k)|] \leqq E[|g_0(X_1) - g_m(X_1)|] \to 0, \qquad m \to \infty$$

となる。よって部分列 $\{m_\ell\}_{\ell=1}^\infty$ が存在して

$$\frac{1}{n}\sum_{k=1}^n g_{m_\ell}(X_k)\sigma_k \to \frac{1}{n}\sum_{k=1}^n g_0(X_k)\sigma_k \quad a.s. \qquad \ell \to \infty$$

となる。これより

$$\frac{1}{n}\sum_{k=1}^{n} g_0(X_k)\sigma_k \leqq \sup_{m\geqq 1}\frac{1}{n}\sum_{k=1}^{n} g_m(X_k)\sigma_k \leqq \sup_{g\in\mathcal{K}}\frac{1}{n}\sum_{k=1}^{n} g(X_k)\sigma_k \quad a.s.$$

となり、

$$\sup_{g\in\bar{\mathcal{K}}}\frac{1}{n}\sum_{k=1}^{n} g(X_k)\sigma_k \leqq \sup_{g\in\mathcal{K}}\frac{1}{n}\sum_{k=1}^{n} g(X_k)\sigma_k \quad a.s.$$

となる。よって

$$\mathcal{R}_n(\bar{\mathcal{K}};\nu) = E[\sup_{g\in\bar{\mathcal{K}}}\frac{1}{n}\sum_{k=1}^{n} g(X_k)\sigma_k] \leqq E[\sup_{g\in\mathcal{K}}\frac{1}{n}\sum_{k=1}^{n} g(X_k)\sigma_k] = \mathcal{R}_n(\mathcal{K};\nu)$$

となり、主張 (7) を得る。

∎

定義 3.5 (1) $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とする。$\hat{\mathcal{Q}}_n(\cdot;\mathcal{K}) : E^{2n} \to [0,\infty)$, $n \geqq 1$, を

$$\hat{\mathcal{Q}}_n(x_1,\ldots,x_{2n};\mathcal{K}) = \frac{1}{n}\sup_{g\in\mathcal{K}}(\sum_{k=1}^{n} g(x_{2k-1}) - \sum_{k=1}^{n} g(x_{2k})) \qquad (x_1,\ldots,x_{2n}) \in E^{2n}$$

により定義する。
(2) $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とし、ν を $(E,\mathcal{E})$ 上の確率測度とする。$\mathcal{Q}_n(\mathcal{K};\nu) \in [0,\infty)$, $n \geqq 1$, を

$$\mathcal{Q}_n(\mathcal{K};\nu) = \int_{E^{2n}} \hat{\mathcal{Q}}_n(x_1,\ldots,x_{2n};\mathcal{K})\nu^{\otimes 2m}(dx_1,\ldots,dx_{2n})$$

により定義する。

命題 3.6 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合、ν を E 上の確率測度とする。この時、

$$\int_{E^n}\sup_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^{n} g(x_k) - \int_E g d\nu)\nu^{\otimes n}(dx_1,\ldots,x_n) \leqq \mathcal{Q}_n(\mathcal{K};\nu),$$

かつ

$$\int_{E^n}\sup_{g\in\mathcal{K}}(\int_E g d\nu - \frac{1}{n}\sum_{k=1}^{n} g(x_k))\nu^{\otimes n}(dx_1,\ldots,x_n) \leqq \mathcal{Q}_n(\mathcal{K};\nu),$$

が成り立つ。特に、

$$\int_{E^n}\sup_{g\in\mathcal{K}\cup\{0\}}|\frac{1}{n}\sum_{k=1}^{n} g(x_k) - \int_E g d\nu|\nu^{\otimes n}(dx_1,\ldots,x_n) \leqq 2\mathcal{Q}_n(\mathcal{K}\cup\{0\};\nu),$$

が成り立つ。

証明. $X_1,\ldots,X_{2n}$ を独立で同分布を持つ E-値確率変数でその分布を ν とする。この時、

$$\mathcal{Q}_n(\mathcal{K};\nu), = E[\sup_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^{n} g(X_{2k-1}) - \frac{1}{n}\sum_{k=1}^{n} g(X_{2k}))]$$

となる。$g_0 \in \mathcal{K}$ に対して

$$\frac{1}{n}\sum_{k=1}^{n} g_0(X_{2k-1}) - \int_E g_0(x)\nu(dx) = E[\frac{1}{n}\sum_{k=1}^{n} g_0(X_{2k-1}) - \frac{1}{n}\sum_{k=1}^{n} g(X_{2k}))|X_1,X_3,\ldots,X_{2n-1}]$$

$$\leqq E[\sup_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^n g(X_{2k-1})-\frac{1}{n}\sum_{k=1}^n g(X_{2k}))|X_1,X_3,\dots,X_{2n-1}].$$

よって、

$$E[\sup_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^n g(X_{2k-1})-\int_E g(x)\nu(dx))]\leqq E[\sup_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^n g(X_{2k-1})-\frac{1}{n}\sum_{k=1}^n g(X_{2k}))].$$

これより最初の主張がわかる。2 番目の主張も

$$E[\sup_{g\in\mathcal{K}}(\int_E g(x)\nu(dx)-\frac{1}{n}\sum_{k=1}^n g(X_{2k}))]$$

$$=E[\sup_{g\in\mathcal{K}}E[\frac{1}{n}\sum_{k=1}^n g(X_{2k-1})-\frac{1}{n}\sum_{k=1}^n g(X_{2k})|X_2,X_4,\dots,X_{2n}]]$$

$$\leqq E[E[\sup_{g\in\mathcal{K}}\frac{1}{n}\sum_{k=1}^n g(X_{2k-1})-\frac{1}{n}\sum_{k=1}^n g(X_{2k})|X_2,X_4,\dots,X_{2n}]]$$

よりわかる。

最後の主張はこれら 2 つの主張よりわかる。 ∎

$I\subset\mathbf{Z}_{\geqq 1}$ に対して $\tau_I:\mathbf{Z}_{\geqq 1}\to\mathbf{Z}_{\geqq 1}$ を $i\in\mathbf{Z}_{\geqq 1}$ に対して

$$\tau_I(2i-1)=\begin{cases} 2i-1, & i\in I \text{ の時} \\ 2i, & i\notin I \text{ の時}\end{cases}$$

$$\tau_I(2i)=\begin{cases} 2i, & i\in I \text{ の時} \\ 2i-1, & i\notin I \text{ の時}\end{cases}$$

により定める。

この時、次が成立する。

命題 3.7 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とする。
任意の $n\geqq 1$ 及び $x_1,\dots,x_{2n}\in E^{2n}$ に対して

$$\sum_{I\subset\{1,\dots,n\}}\frac{1}{2^n}\hat{\mathcal{Q}}_n(x_{\tau_I(1)},\dots,x_{\tau_I(2n)};\mathcal{K})$$

$$\leqq\hat{\mathcal{R}}_n(x_1,x_3,\dots,x_{2n-1};\mathcal{K})+\hat{\mathcal{R}}_n(x_2,x_4,\dots,x_{2n};\mathcal{K})$$

が成立する。特に、

$$\mathcal{Q}_n(\mathcal{K};\nu)\leqq 2\mathcal{R}_n(\mathcal{K};\nu)$$

となる。

証明. $\sigma_1,\sigma_2,\dots$ は独立同分布な確率変数列で $P(\sigma_k=\pm 1)=1/2$ となるものとする。この時、

$$E[\sup_{g\in\mathcal{K}}(\sum_{i=1}^n g(x_{2i-1})\sigma_i-\sum_{i=1}^n g(x_{2i})\sigma_i)]$$

$$=\sum_{I\subset\{1,\dots,n\}}E[\sup_{g\in\mathcal{K}}(\sum_{i=1}^n g(x_{2i-1})\sigma_i-\sum_{i=1}^n g(x_{2i})\sigma_i),\{i=1,\dots,n;\ \sigma_i=1\}=I]$$

$$= \sum_{I \subset \{1,\dots,n\}} \frac{1}{2^n} \hat{\mathcal{Q}}_n(x_{\tau_I(1)}, \dots, x_{\tau_I(2n)}; \mathcal{K})$$

を得る。

また、

$$E[\sup_{g \in \mathcal{K}}(\sum_{k=1}^n g(x_{2i-1})\sigma_i - \sum_{k=1}^n g(x_{2i})\sigma_i)]$$
$$\leqq E[\sup_{g \in \mathcal{K}}(\sum_{k=1}^n g(x_{2i-1})\sigma_i) + \sup_{g \in \mathcal{K}}(\sum_{k=1}^n g(x_{2i})(-\sigma_i))]$$
$$= \hat{\mathcal{R}}_n(x_1, x_3, \dots, x_{2n-1}; \mathcal{K}) + \hat{\mathcal{R}}_n(x_2, x_4, \dots, x_{2n}; \mathcal{K})$$

となるので、最初の主張を得る。

後半は

$$\int_{E^{2n}} \sum_{I \subset \{1,\dots,n\}} \frac{1}{2^n} \hat{\mathcal{Q}}_n(x_{\tau_I(1)}, \dots, x_{\tau_I(2n)}; \mathcal{K}) \nu^{\otimes 2n}(dx_1, \dots, x_{2n}) = \mathcal{Q}_n(\mathcal{K}; \nu)$$

よりわかる。

∎

命題 3.6, 3.7 より以下を得る。

命題 3.8 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合、ν を E 上の確率測度とする。この時以下が成立する。

$$\int_{E^n} \sup_{g \in \mathcal{K}}(\frac{1}{n}\sum_{k=1}^n g(x_k) - \int_E g d\nu) \nu^{\otimes}(dx_1, \dots, x_n) \leqq 2\mathcal{R}_n(\mathcal{K}; \nu).$$

$$\int_{E^n} \sup_{g \in \mathcal{K}}(\int_E g d\nu - \frac{1}{n}\sum_{k=1}^n g(x_k)) \nu^{\otimes}(dx_1, \dots, x_n) \leqq 2\mathcal{R}_n(\mathcal{K}; \nu).$$

命題 3.9 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合、ν を $(E, \mathcal{B})$ 上の確率測度 とする。また、すべての $g \in \mathcal{K}$ は ν-可積分であり、

$$\int_E g(x)\nu(dx) = 0, \qquad \mathrm{g} \in \mathcal{K}$$

と仮定する。この時、

$$n\mathcal{R}_n(\mathcal{K}; \nu) \leqq 2 \max_{m=1,\dots,n} m\mathcal{Q}_m(\mathcal{K}; \nu)$$

が成り立つ。

証明. $X_1, \dots, X_{2n}$ は E-値確率変数、$\sigma_1, \dots, \sigma_n$ は確率変数、$X_1, \dots, X_{2n}, \sigma_1, \dots, \sigma_n$ は独立であり、X_k, $k = 1, \dots, 2n$, の分布は ν, $P(\sigma_k = \pm 1) = 1/2$, $k = 1, \dots, n$, とする。この時

$$n\mathcal{R}_n(\mathcal{K}; \nu) = E[\sup_{g \in \mathcal{K}}(\sum_{k=1}^n \sigma_k g(X_k))]$$

である。$I_+ = \{k = 1, \dots, n;\ \sigma_k = 1\}$, $I_- = \{k = 1, \dots, n;\ \sigma_k = -1\}$ とおくと

$$\sup_{g \in \mathcal{K}}(\sum_{k=1}^n \sigma_k g(X_k)) = \sup_{g \in \mathcal{K}}(\sum_{k \in I_+} g(X_k) - \sum_{k \in I_-} g(X_k))$$

$$\leqq \sup_{g \in \mathcal{K}}(\sum_{k \in I_+} g(X_k)) + \sup_{g \in \mathcal{K}}(-\sum_{k \in I_-} g(X_k))$$

となる。よって、$n_+ = \#(I_+)$, $n_- = \#(I_-)$, とおくと

$$E[\sup_{g\in\mathcal{K}}(\sum_{k=1}^{n}\sigma_k g(X_k))|\sigma_1,\dots,\sigma_n]$$

$$\leqq \int_{E^{n_+}}(\sum_{i=1}^{n_+} g(x_i))|\nu^{\otimes n_+}(dx_1,\dots,dx_{n_+})) + \int_{E^{n_-}}(-\sum_{i=1}^{n_-} g(x_i))\nu^{\otimes n_-}(dx_1,\dots,dx_{n_-}))$$

$$\leqq \max_{m=1,\dots,n}\sup_{g\in\mathcal{K}} E[(\sum_{i=1}^{m} g(X_i))] + \max_{m=1,\dots,n}\sup_{g\in\mathcal{K}} E[(-\sum_{i=1}^{m} g(X_i))]$$

が成り立つ。よって、命題 3.6 より

$$\mathcal{R}_n(\mathcal{K};\nu)$$

$$\leqq \max_{m=1,\dots,n} E[\sup_{g\in\mathcal{K}}(\sum_{i=1}^{m} g(X_i))] + \max_{m=1,\dots,n} E[\sup_{g\in\mathcal{K}}(-\sum_{i=1}^{m} g(X_i))]$$

$$\leqq 2\max_{m=1,\dots,n} m\mathcal{Q}_m(\mathcal{K};\nu).$$

これより主張を得る。

∎

命題 3.10 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合、ν を $(E,\mathcal{B}_E)$ 上の確率測度 とする。さらに $\int_E |g|d\nu \leqq 1$, $g\in\mathcal{K}$, であると仮定する。この時、

$$n\mathcal{R}_n(\mathcal{K};\nu) \leqq 2\max_{m=1,\dots,n} m\mathcal{Q}_m(\mathcal{K};\nu) + n^{1/2}$$

が成り立つ。

証明. $\mathcal{K}_\nu \subset \mathcal{M}(E)$ を

$$\mathcal{K}_\nu = \{g - \int_E g d\nu;\ g\in\mathcal{K}\}$$

で定める。この時、定義より

$$\mathcal{Q}_n(\mathcal{K};\nu) = \mathcal{Q}_n(\mathcal{K}_\nu;\nu)$$

である。$\sigma_1,\dots,\sigma_n$ は独立な確率変数で $P(\sigma_k = \pm 1) = 1/2$, $k = 1,\dots,n$, とする。この時、$g\in\mathcal{K}$ に対して

$$\sum_{k=1}^{n}\sigma_k g(x_k) = \sum_{k=1}^{n}\sigma_k(g(x_k) - \int_E g d\nu) + (\sum_{k=1}^{n}\sigma_k\int_E g d\nu)$$

$$\leqq \sum_{k=1}^{n}\sigma_k(g(x_k) - \int_E g d\nu) + |\sum_{k=1}^{n}\sigma_k|$$

よって、

$$\sup_{g\in\mathcal{K}}\sum_{k=1}^{n}\sigma_k g(x_k) \leqq \sup_{g\in\mathcal{K}_\nu}\sum_{k=1}^{n}\sigma_k g(x_k) + |\sum_{k=1}^{n}\sigma_k|.$$

$$E[|\sum_{k=1}^{n}\sigma_k|] \leqq E[(\sum_{k=1}^{n}\sigma_k)^2]^{1/2} = n^{1/2}$$

であるので、

$$n\mathcal{R}_n(\mathcal{K};\nu) \leqq n\mathcal{R}_n(\mathcal{K}_\nu;\nu) + n^{1/2} \leqq \max_{m=1,\dots,n} m\mathcal{Q}_m(\mathcal{K}_\nu;\nu) + n^{1/2}$$

がわかり、主張を得る。

∎

注意. $\mathcal{K}=\{-1,0,1\}$ とすると、

$$n^{1/2}\mathcal{R}_n(\mathcal{K},\nu)=n^{-1/2}E[|\sum_{k=1}^n\sigma_k|]$$

$$\to(\frac{1}{2\pi})^{1/2}\int_{-\infty}^{\infty}|t|\exp(-\frac{t^2}{2})dt=(\frac{2}{\pi})^{1/2},\qquad n\to\infty$$

となる。

命題 3.11 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない有限部分集合、ν を $(E,\mathcal{B}_E)$ 上の確率測度 とする。
(1) $\gamma_0>0$ が存在して

$$\sup_{g\in\mathcal{K}}\int_E\exp(\gamma_0^{-1}g(x)^2)\nu(dx)=C_0<\infty$$

と仮定する。この時、

$$\mathcal{R}_n(\mathcal{K};\nu)\leqq 2n^{-1/2}\gamma_0^{1/2}(\log(\#(\mathcal{K}))^{1/2}(\log C_0+\frac{1}{2})^{1/2}$$

が成り立つ。
(2) $c_0>0$ が存在して

$$|g(x)|\leqq c_0,\qquad g\in\mathcal{K},\ x\in E$$

と仮定する。この時、

$$\mathcal{R}_n(\mathcal{K};\nu)\leqq n^{-1/2}c_0(\log(2\#(\mathcal{K}))^{1/2}$$

が成り立つ。
(3) $g_i\in L^2(E,d\nu)$, $i=1,2,\ldots,$ とし $\mathcal{K}\subset L^2(E,d\nu)$ を

$$\mathcal{K}=\{\sum_{i=1}^{\infty}a_ig_i;\ a_i\in\mathbf{R},\ i=1,2,\ldots,\sum_{i=1}^{\infty}a_i^2\leqq 1\}$$

とおく。この時、

$$\mathcal{R}_n(\mathcal{K};\nu)\leqq n^{-1/2}(\sum_{i=1}^{\infty}\int_E g_i^2d\nu)^{1/2}$$

が成り立つ。

証明. $X_1,\ldots,X_n$ は E-値確率変数、$\sigma_1,\ldots,\sigma_n$ は確率変数、$X_1,\ldots,X_n,\sigma_1,\ldots,\sigma_n$ は独立であり、X_k, $k=1,\ldots,n,$ の分布は ν, $P(\sigma_k=\pm1)=1/2$, $k=1,\ldots,n,$ とする。

まず (1) を示す。$t>0$ に対して

$$tE[\sup_{g\in\mathcal{K}}\sum_{k=1}^n g(X_k)\sigma_k]\leqq\log E[\exp(t\sup_{g\in\mathcal{K}}\sum_{k=1}^n g(X_k)\sigma_k)]$$

$$\leqq\log E[\sum_{g\in\mathcal{K}}\exp(t\sum_{k=1}^n g(X_k)\sigma_k)]=\log(\sum_{g\in\mathcal{K}}E[\exp(tg(X_1)\sigma_1)]^n).$$

$E[g(X_1)\sigma_1]=0$, $E[\exp(\gamma_0^{-1}(g(X_1)\sigma_1)^2]\leqq C_0$ であるので、命題 2.3 (2) より

$$\log E[\exp(tg(X_1)\sigma_1)]\leqq\gamma_0t^2(\log C_0+\frac{1}{2}).$$

よって、

$$tE[\sup_{g\in\mathcal{K}}\sum_{k=1}^n g(X_k)\sigma_k]\leqq\log(\sum_{g\in\mathcal{K}}\exp(\gamma_0nt^2(\log C_0+\frac{1}{2})))$$

$$= \log(\#(\mathcal{K})\exp(\gamma_0 n t^2(\log C_0 + \frac{1}{2}))) = \log(\#(\mathcal{K})) + \gamma_0 n t^2(\log C_0 + \frac{1}{2}).$$

これより

$$n\mathcal{R}_n(\mathcal{K},\nu) = E[\sup_{g\in\mathcal{K}}\sum_{k=1}^n g(X_k)\sigma_k] \leqq \frac{1}{t}\log(\#(\mathcal{K})) + \gamma_0 n t(\log C_0 + \frac{1}{2})$$

となる。$t = n^{-1/2}\gamma_0^{-1/2}(\log(\#(\mathcal{K}))^{1/2}(\log C_0 + \frac{1}{2})^{-1/2}$ とおくと

$$n\mathcal{R}_n(\mathcal{K},\nu) \leqq 2n^{1/2}\gamma_0^{-1/2}(\log(\#(\mathcal{K}))^{1/2}(\log C_0 + \frac{1}{2})^{1/2}$$

となり主張を得る。

主張 (2) は命題 2.2 を用いて同様に示される。

主張 (3) は

$$\sup_{g\in\mathcal{K}}\sum_{k=1}^n g(X_k)\sigma_k = \sup\{\sum_{i=1}^\infty a_i(\sum_{k=1}^n g_i(X_k)\sigma_k);\ \sum_{i=1}^\infty a_i^2 \leqq 1\} = (\sum_{i=1}^\infty(\sum_{k=1}^n g_i(X_k)\sigma_k)^2)^{1/2}$$

であるので、

$$n\mathcal{R}_n(\mathcal{K},\nu) \leqq E[\sum_{i=1}^\infty(\sum_{k=1}^n g_i(X_k)\sigma_k)^2]^{1/2} = (\sum_{i=1}^\infty E[(\sum_{k=1}^n g_i(X_k)\sigma_k)^2])^{1/2} = (\sum_{i=1}^\infty n\int_E g_i^2 d\nu)^{1/2}$$

よりわかる。

∎

4 一様大数の法則

$(E,\mathcal{E})$ を可測空間、$\mathcal{M}(E)$ を E 上の可測関数全体とする。

命題 4.1 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とし、ν を $(E,\mathcal{E})$ 上の確率測度とする。$X_k, k = 1,2,\dots,$ は独立同分布な E-値確率変数でその分布は ν とする。さらに、$\gamma_0 > 0$ が存在して

$$C_0 = \int_E \exp(\gamma_0^{-1}\sup_{g\in\mathcal{K}} g(x)^2)\nu(dx) < \infty$$

が成立すると仮定する。

今、確率変数 Y_n, $n \geqq 1$, を

$$Y_n = \sup_{g\in\mathcal{K}}(\sum_{k=1}^n g(X_k) - n\int_E g d\nu)$$

$$Y_n' = \sup_{g\in\mathcal{K}}(n\int_E g d\nu - \sum_{k=1}^n g(X_k))$$

で定める。この時、以下が成立する。

(1) $t \in \mathbf{R}$ に対して

$$E[\exp(t(Y_n - E[Y_n]))] \leqq \exp(c_1 n t^2),$$

$$E[\exp(t(Y_n' - E[Y_n']))] \leqq \exp(c_1 n t^2)$$

が成立する。ここで

$$c_1 = \gamma_0(2\log C_0 + \frac{1}{2})$$

である。
(2) 任意の $a>0$ に対して

$$P(\frac{1}{n}(Y_n - E[Y_n]) > an^{-1/2}) \leqq \exp(-\frac{1}{4c_1}a^2),$$

$$P(\frac{1}{n}(Y'_n - E[Y'_n]) > an^{-1/2}) \leqq \exp(-\frac{1}{4c_1}a^2),$$

が成立する。

証明. $n \geqq 1$ を固定して考える。確率変数 M_m, $m=0,1,\ldots,n.$ を

$$M_m = E[Y_n|\sigma\{X_1,\cdots,X_m\}], \qquad m=0,\ldots,n,$$

で定める。$f: E^n \to [0,\infty]$ を

$$f(x_1,\ldots,x_n) = \sup_{g\in\mathcal{K}}(\sum_{k=1}^n g(x_k) - n\int_E g d\nu) \qquad x_1,\ldots x_n \in E,$$

で定め、$f_m : E^n \to [0,\infty]$, $m=0,1,\ldots,n,$ を

$$f_m(x_1,\ldots,x_m) = E[f(x_1,\ldots,x_m,X_{m+1},\ldots,X_n)] \qquad x_1,\ldots x_m \in E,$$

で定めると

$$M_m = f_m(X_1,\ldots X_m), \qquad m=0,1,\ldots,n,$$

$$f_{m-1}(x_1,\ldots,x_{m-1}) = E[f_m(x_1,\ldots,x_{m-1},X_m)] = E[f_m(x_1,\ldots,x_{m-1},X_{n+1})],$$

$m=1,\ldots,n$, $x_1,\ldots x_{m-1}\in E$, となることがわかる。
　$m=1,\ldots,n$, $x_1,\ldots,x_{m-1},y,y'\in E$ に対して

$$f_m(x_1,\ldots,x_{m-1},y) - f_m(x_1,\ldots,x_{m-1},y')$$

$$= E[f(x_1,\ldots,x_{m-1},y,X_{m+1},\ldots,X_n)] - E[f(x_1,\ldots,x_{m-1},y',X_{m+1},\ldots,X_n)]$$

であり

$$f(x_1,\ldots,x_{m-1},y,X_{m+1},\ldots,X_n) - f(x_1,\ldots,x_{m-1},y',X_{m+1},\ldots,X_n)$$

$$= \sup_{g\in\mathcal{K}}(\sum_{k=1}^{m-1} g(x_k) + g(y) + \sum_{k=m+1}^{n} g(X_k) - n\int_E g d\nu) - \sup_{g\in\mathcal{K}}(\sum_{k=1}^{m-1} g(x_k) + g(y') + \sum_{k=m+1}^{n} g(X_k) - n\int_E g d\nu)$$

$$\leqq \sup_{g\in\mathcal{K}} |g(y) - g(y')|$$

であるので、

$$|f_m(x_1,\ldots,x_{m-1},y) - f_m(x_1,\ldots,x_{m-1},y')| \leqq \sup_{g\in\mathcal{K}}|g(y)-g(y')|,$$

$m=1,\ldots,n,\ x_1,\ldots,x_m,y,y'\in E$, がわかる。よって、

$$|f_m(x_1,\ldots,x_m) - f_{m-1}(x_1,\ldots,x_{m-1})|$$

$$\leqq E[|f_m(x_1,\ldots,x_m) - f_m(x_1,\ldots,x_{m-1},X_{n+1})|] \leqq E[\sup_{g\in\mathcal{K}}|g(x_m) - g(X_{n+1})|]$$

となる。よって

$$\exp(\frac{\gamma_0^{-1}}{2}|f_m(x_1,\ldots,x_m) - f_{m-1}(x_1,\ldots,x_{m-1})|^2) \leqq \exp(\frac{\gamma_0^{-1}}{2}E[\sup_{g\in\mathcal{K}}|g(x_m)-g(X_{n+1})|]^2)$$

$$\leqq E[\exp(\frac{\gamma_0^{-1}}{2}\sup_{g\in\mathcal{K}}|g(x_m)-g(X_{n+1})|^2)] \leqq E[\exp(\gamma_0^{-1}\sup_{g\in\mathcal{K}}(|g(x_m)|^2+|g(X_{n+1})|^2))]$$

であるので、

$$E[\exp(\frac{\gamma_0}{2}|M_m-M_{m-1}|^2)|\sigma\{X_1,\ldots,X_{m-1}\}]$$
$$\leqq E[\exp(\gamma_0^{-1}\sup_{g\in\mathcal{K}}(|g(X_m)|^2+|g(X_{n+1})|^2))] \leqq C_0^2$$

となる。また、

$$E[M_m-M_{m-1}|\sigma\{X_1,\ldots,X_{m-1}\}]=0$$

である。

よって、命題 2.3 (2) より

$$E[\exp(t(M_m-M_{m-1}))|\sigma\{X_1,\ldots,X_{m-1}\}] \leqq \exp(\gamma_0(2\log C_0+\frac{1}{2})t^2)\ a.s.$$

$t\in\mathbf{R}$, $m=1,\ldots,n$, を得る。よって帰納的に

$$E[\exp(t(M_n-M_m)|\sigma\{X_1,\ldots,X_m\}] \leqq \exp(c_1t^2(n-m))\ a.s. \qquad m=n-1,\ldots,0,$$

を得る。これより主張 (1) の前半がわかる。

$$Y_n'=\sup_{g\in-\mathcal{K}}(\sum_{k=1}^n g(X_k)-n\int_E gd\nu)$$

であるので、後半もわかる。

また、$a,t>0$ に対して

$$P(\frac{1}{n^{1/2}}(Y_n-E[Y_n])>a) \leqq \exp(-at)E[\exp(tn^{-1/2}(Y_n-E[Y_n]))]$$
$$\leqq \exp(-at+c_1t^2)$$

であるので、$t=\dfrac{a}{2c_1}$ とおくと、主張 (2) の前半を得る。後半の証明も同様である。

∎

命題 3.9, 4.1(2) より以下がわかる。

命題 4.2 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とし、ν を $(E,\mathcal{E})$ 上の確率測度とする。$X_k, k=1,2,\ldots,$ は独立同分布な E-値確率変数でその分布は ν とする。また、$\gamma_0>0$ が存在して

$$C_0=\int_E \exp(\gamma_0^{-1}\sup_{g\in\mathcal{K}}g(x)^2)\nu(dx)<\infty$$

が成立すると仮定する。この時、以下が成立する。任意の $a>0$ に対して

$$P(\sup_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^n g(X_k)-\int_E gd\nu)>\frac{a}{n^{1/2}}+2\mathcal{R}_n(\mathcal{K};\nu)) \leqq \exp(-\frac{1}{4c_1}a^2)$$

かつ

$$P(\inf_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^n g(X_k)-\int_E gd\nu)<-\frac{a}{n^{1/2}}-2\mathcal{R}_n(\mathcal{K};\nu)) \leqq \exp(-\frac{1}{4c_1}a^2)$$

が成立する。ただし

$$c_1=\gamma_0(2\log C_0+\frac{1}{2})$$

である。特に

$$P(\sup_{g\in\mathcal{K}}|\frac{1}{n}\sum_{k=1}^n g(X_k)-\int_E gd\nu|>\frac{a}{n^{1/2}}+2\mathcal{R}_n(\mathcal{K};\nu)) \leqq 2\exp(-\frac{1}{4c_1}a^2)$$

が成立する。

命題 4.3 $\mathcal{K}$ を $\mathcal{M}(E)$ の空でない部分集合とし、ν を $(E,\mathcal{E})$ 上の確率測度とする。$X_k, k=1,2,\dots,$ は独立同分布な E-値確率変数でその分布は ν とする。また、$c_0>0$ が存在して

$$\sup_{g\in\mathcal{K}}|g(x)| \leqq c_0, \qquad x\in E$$

が成立すると仮定する。この時、以下が成立する。任意の $a>0$ に対して

$$P(\sup_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^n g(X_k) - \int_E gd\nu) > \frac{a}{n^{1/2}} + 2\mathcal{R}_n(\mathcal{K};\nu)) \leqq \exp(-\frac{1}{2c_0^2}a^2)$$

かつ

$$P(\inf_{g\in\mathcal{K}}(\frac{1}{n}\sum_{k=1}^n g(X_k) - \int_E gd\nu) < -\frac{a}{n^{1/2}} + 2\mathcal{R}_n(\mathcal{K};\nu)) \leqq \exp(-\frac{1}{2c_0^2}a^2)$$

が成立する。

証明. 命題 4.1 と同様に確率変数 M_m, $m=0,1,\dots,n.$ を

$$M_m = E[Y_n|\sigma\{X_1,\cdots,X_m\}], \qquad m=0,\dots,n,$$

で定める。$f: E^n\to[0,\infty]$ を

$$f(x_1,\dots,x_n) = \sup_{g\in\mathcal{K}}(\sum_{k=1}^n g(x_k) - n\int_E gd\nu) \qquad x_1,\dots x_n\in E,$$

で定め、$f_m: E^n\to[0,\infty]$, $m=0,1,\dots,n,$ を

$$f_m(x_1,\dots,x_m) = E[f(x_1,\dots,x_m,X_{m+1},\dots,X_n)] \qquad x_1,\dots x_m\in E,$$

で定める。命題 4.1 の証明と同様に

$$\begin{aligned}
&|f_m(x_1,\dots,x_m) - f_{m-1}(x_1,\dots,x_{m-1})| \\
&= |E[(\sup_{g\in\mathcal{K}}(\sum_{k=1}^m g(x_k) + \sum_{k=m+1}^n g(X_k) - n\int_E gd\nu) - (\sup_{g\in\mathcal{K}}(\sum_{k=1}^{m-1} g(x_k) + \sum_{k=m}^n g(X_k) - n\int_E gd\nu)]| \\
&\leqq E[\sup_{g\in\mathcal{K}}|g(x_m) - g(X_{n+1})|] \leqq 2c_0
\end{aligned}$$

となる。よって　命題 2.2 より

$$E[\exp(t(M_m - M_{m-1}))|\sigma\{X_1,\dots,X_{m-1}\}] \leqq \exp(\frac{1}{2}c_0^2t^2)\ a.s.$$

$t\in\mathbf{R}$, $m=1,\dots,n,$ を得る。よって

$$E[\exp(t(M_n - M_0))|\sigma\{X_1,\dots,X_m\}] \leqq \exp(\frac{1}{2}c_0^2t^2n)\ a.s.$$

を得る。これより、命題 4.1, 4.2 の証明と同様にして主張を得る。

∎

5 正則化のため準備

$(E, \mathcal{E})$ は可測空間、ν を $(E, \mathcal{E})$ 上の確率測度とする。Θ は距離空間で $F: E \times \Theta \to \mathbf{R}$ は可測関数とする。また、$\phi_0 : \Theta \to [0, \infty)$ を連続関数とする。$r > 0$ に対して $\mathcal{M}(E)$ の空でない部分集合 $\mathcal{K}_r$ を

$$\mathcal{K}_r = \{F(\cdot, \theta) \in \mathcal{M}(E);\ \phi_0(\theta) \leqq r\}$$

で定める。

また、$(\Omega, P, \mathcal{F})$ は確率空間、$X_k, k = 1, 2, \ldots,$ は独立同分布な E-値確率変数でその分布は ν とする。

以下では $n \geqq 1$ は固定して考える。

この時、以下が成立する。

命題 5.1 単調増大関数 $\psi_i : [0, \infty) \to [\infty)$, $i = 1, 2$, 及び $r_0 \geqq 1$, が存在して、以下の *2* 条件が成立すると仮定する。

(A-1) $\sup_{g \in \mathcal{K}_r} |g(x)| \leqq \psi_1(r) \qquad r \geqq r_0.$

(A-2) $\mathcal{R}_n(\mathcal{K}_r; \nu) \leqq \psi_2(r), \qquad r \geqq r_0.$

この時、任意の $a > 0$ に対して $\Omega_a \in \mathcal{F}$ が存在して

$$P(\Omega_a) \geqq 1 - 2\frac{\exp(-2(r_0+1)a^2)}{1 - \exp(-2a^2)}$$

であり、$\omega \in \Omega_a$ に対して

$$|\frac{1}{n}\sum_{k=1}^{n} F(X_k, \theta) - \int_E F(x, \theta)\nu(dx)|$$

$$\leqq \frac{a}{n^{1/2}}\tilde{\psi}_1(\phi_0(\theta) \vee r_0 + 1) + 2\psi_2(\phi_0(\theta) \vee r_0 + 1), \qquad \theta \in \Theta$$

が成立する。ただし、$\tilde{\psi}_1(r) = \psi_1(r)r^{1/2}$, $r \geqq 1$, である。

証明. 命題 4.2 より、任意の $s > 0$ 及び $r \geqq r_0$ に対して

$$P(\sup_{g \in \mathcal{K}_r} |\frac{1}{n}\sum_{k=1}^{n} g(X_k) - \int_E g d\nu| > \frac{s}{n^{1/2}} + 2\mathcal{R}_n(\mathcal{K}_r, \nu))$$

$$\leqq 2\exp(-\frac{1}{\psi_1(r)^2}s^2)$$

が成り立つ。よって、$s = a\tilde{\psi}_1(r) = a\psi_1(r)r^{1/2}$ とおくと、

$$P(\sup_{g \in \mathcal{K}_r} |\frac{1}{n}\sum_{k=1}^{n} g(X_k) - \int_E g d\nu| > \frac{a\tilde{\psi}_1(r)}{n^{1/2}} + 2\psi_2(r)$$

$$\leqq 2\exp(-2ra^2) \qquad a > 0,\ r \geqq r_0$$

を得る。

さらに、$\ell \geqq 1$ に対して

$$\Omega_{a,\ell}\text{'} = \{\sup_{g \in \mathcal{K}_\ell} |\frac{1}{n}\sum_{k=1}^{n} g(X_k) - \int_E g d\nu| > \frac{a\tilde{\psi}_1(r_0+\ell)}{n^{1/2}} + 2\psi_2(r_0+\ell)\}$$

とおくと、

$$P(\bigcup_{\ell=\ell_0}^{\infty} \Omega'_{a,\ell}) \leqq \sum_{\ell=1}^{\infty} P(\Omega'_{a,\ell})$$

$$\leqq 2\sum_{\ell=1}^{\infty}\exp(-2(r_0+\ell)a^2) \leqq 2\frac{\exp(-2(r_0+1)a^2)}{1-\exp(-2a^2)}$$

を得る。

$$\Omega_a = \Omega \setminus (\bigcup_{\ell=\ell_0}^{\infty}\Omega'_{a,\ell})$$

とおくと $r_0+\ell-1 \leqq \phi_0(\theta)\vee r_0 < r_0+\ell,\ \ell \geqq 1,$ ならば $F(\cdot,\theta)\in\mathcal{K}_{r_0+\ell}$ であるので、$\omega\in\Omega_a$ に対して

$$|\frac{1}{n}\sum_{k=1}^{n}F(X_k,\theta) - \int_E F(x,\theta)\nu(dx)| \leqq \frac{a}{n^{1/2}}\tilde{\psi}_1(\phi_0(\theta)\vee r_0+1) + 2\psi_2(\phi_0(\theta)\vee r_0+1)$$

が成立する。

∎

命題 5.2 単調増大関数 $\psi_i : [0,\infty)\to[\infty),\ i=1,2,\ C_0\in(0,\infty)$ 及び $r_0\geqq 1,$ が存在して、以下の *2* 条件が成立すると仮定する。

(A-1) $\displaystyle\int_E \exp(\psi_1(r)^{-1}\sup_{g\in\mathcal{K}_r}|g(x)|^2)\nu(dx)) \leqq \psi_1(r) \qquad r\geqq r_0.$

(A-2) $\mathcal{R}_n(\mathcal{K}_r;\nu)\leqq\psi_2(r), \qquad r\geqq r_0.$

　この時、任意の $a>0$ に対して $\Omega_a\in\mathcal{F}$ が存在して

$$P(\Omega_a)\geqq 1-2\frac{\exp(-2(r_0+1)a^2)}{1-\exp(-2a^2)}$$

であり、$\omega\in\Omega_a$ に対して

$$|\frac{1}{n}\sum_{k=1}^{n}F(X_k,\theta) - \int_E F(x,\theta)\nu(dx)|$$

$$\leqq \frac{a}{n^{1/2}}\tilde{\psi}_1(\phi_0(\theta)\vee r_0+1) + 2\psi_2(\phi_0(\theta)\vee r_0+1) \qquad \theta\in\Theta$$

が成立する。ただし、$\tilde{\psi}_1(r)=(c_1\psi_1(r)r)^{1/2},\ r\geqq 1,$ であり、$c_1 = 8\log C_0+2$ である。

証明. 命題 4.2 より、任意の $s>0$ 及び $r\geqq r_0$ に対して

$$P(\sup_{g\in\mathcal{K}_r}|\frac{1}{n}\sum_{k=1}^{n}g(X_k) - \int_E gd\nu| > \frac{s}{n^{1/2}} + 2\mathcal{R}_n(\mathcal{K}_r,\nu))$$

$$\leqq 2\exp(-\frac{1}{\psi_1(r)c_1}s^2)$$

が成り立つ。よって、$s = a\tilde{\psi}_1(r) = a(c_1\psi_1(r)r)^{1/2}$ とおくと、

$$P(\sup_{g\in\mathcal{K}_r}|\frac{1}{n}\sum_{k=1}^{n}g(X_k) - \int_E gd\nu| > \frac{\tilde{\psi}_1(r)a}{n^{1/2}} + 2\psi_2(r))$$

$$\leqq 2\exp(-2ra^2) \qquad a>0,\ r\geqq r_0$$

を得る。

　よって、命題 5.1 の証明と同様にして主張を得る。

∎

6　多層ニューラルネットワークに対する評価

以下では $(E,\mathcal{E}))$ は可測空間、ν は $(E,\mathcal{E})$ 上の確率測度、$X_1,\ldots,X_n$ は E-値確率変数列、$X_1,\ldots,X_n$ は独立で X_i, $i=1,\ldots,n$, の確率法則は ν とする。

さらに $\rho:\mathbf{R}\to\mathbf{R}$ は連続関数で、以下の 2 条件を満たすとする。

(R-1) $|\rho(t)-\rho(s)|\leqq|t-s|$, $t,s\in\mathbf{R}$,

(R-2) $|\rho(0)|\leqq 1$

また、$\mathcal{K}\subset\mathcal{M}(E)$ は $\{0,1\}\subset\mathcal{K}$ 満たすとし、$r_k\geqq 0$, $k=0,1,2,\ldots,$ とする。

さらに、$\mathcal{K}_0\subset\mathcal{M}(E)$ を

$$\mathcal{K}_0=\{a_0+\sum_{i=1}^m a_ig_i;\ m\geqq 1,;g_i\in\mathcal{K},\ a_0,a_i\in\mathbf{R},\ i=1,\ldots,m,\ \sum_{i=0}^m|a_i|\leqq r_0\}$$

とおき、$\mathcal{K}_k\subset\mathcal{M}(E)$, $k=1,2,\ldots,$ を帰納的に

$$\mathcal{K}_k=\{a_0+\sum_{i=1}^m a_i\rho\circ g_i;\ m\geqq 1,;g_i\in\mathcal{K}_{k-1},\ a_0,a_i\in\mathbf{R},\ i=1,\ldots,m,\ \sum_{i=0}^m|a_i|\leqq r_k\},\qquad k=1,2,\ldots$$

で定める。

命題 6.1 (1) $\mathcal{R}_n(\mathcal{K}_0;\nu)\leqq 2r_0\mathcal{R}_n(\mathcal{K};\nu)$.

(2) $\mathcal{R}_n(\mathcal{K}_k;\nu)\leqq 2r_k(\mathcal{R}_n(\mathcal{K}_{k-1};\nu)+\dfrac{2}{n^{1/2}}),\qquad k=1,2,\ldots$

(3) $C>0$ が存在して

$$\mathcal{R}_n(\mathcal{K};\nu)\leqq\frac{C}{n^{1/2}}$$

が成り立つならば

$$\mathcal{R}_n(\mathcal{K}_k;\nu)\leqq\frac{2^{k+1}C}{n^{1/2}}r_k\prod_{j=0}^{k-1}(r_j+1),\qquad k=1,2,\ldots$$

(4) $|\rho(t)|\leqq 1$, $t\in\mathbf{R}$, が成立するならば

$$\sup\{|g(x)|;\ x\in E,\ g\in\mathcal{K}_n\}\leqq r_n,\qquad n\geqq 1.$$

(5) $|\rho(t)|\leqq|t|$, $t\in\mathbf{R}$, が成立すると仮定する。この時、

$$\sup\{|g(x)|;\ g\in\mathcal{K}_k\cup\{1\}\}\leqq(\prod_{i=0}^k r_i)\sup\{|g(x)|;\ g\in\mathcal{K}\cup\{1\}\}\qquad k=0,1,\ldots$$

が成立する。

証明. 主張 (1) は命題 3.4 (1),(4) よりわかる。

命題 3.4 (5) より

$$\mathcal{R}_n(\mathcal{K}_{k-1}\cup\{0\};\nu)\leqq\mathcal{R}_n(\mathcal{K}_{k-1};\nu)+\frac{1}{n^{1/2}}|\rho(0)|$$

また

$$\mathcal{R}_n(\{0,1\};\nu)=\frac{1}{n}E[(\sum_{k=n}^n\sigma_k)\vee 0]\leqq\frac{1}{n}E[(\sum_{k=n}^n\sigma_k)^2]^{1/2}\leqq\frac{1}{n^{1/2}}$$

よって、命題 3.4 (3) より

$$\mathcal{R}_n(\mathcal{K}_{k-1}\cup\{0,1\};\nu)\leqq\mathcal{R}_n(\mathcal{K}_{k-1};\nu)+\frac{2}{n^{1/2}}$$

これと命題 3.4 (1),(4) より主張 (2) を得る。

主張 (3) は主張 (1),(2) より帰納的に得る。

主張 (4) は $g_i \in \mathcal{M}(E)$. $i = 1, 2, \ldots, m$, $a_i \in \mathbf{R}$, $i = 0, 1, 2, \ldots, m$, ならば

$$|a_0 + \sum_{i=1}^{m} a_i \rho \circ g_i| \leqq \sum_{i=0}^{m} |a_i|$$

よりわかる。

主張 (5) を示す。$g_i \in \mathcal{M}(E)$. $i = 1, 2, \ldots, m$, $a_i \in \mathbf{R}$, $i = 0, 1, 2, \ldots, m$, ならば

$$|a_0 + \sum_{i=1}^{m} a_i \rho \circ g_i| \leqq (\sum_{i=0}^{m} |a_i|)(1 \vee \max_{i=1,\ldots,m} |g_i(x)|)$$

であるので、

$$\sup\{|g(x)|;\ g \in \mathcal{K}_k \cup \{1\}\} \leqq r_k \sup\{|g(x)|;\ g \in \mathcal{K}_{k-1} \cup \{1\}\}).$$

となる。これより主張 (5) は帰納的に示される。

∎

多層ニューラルネットワークを以下のように与える。、$m \geqq 1$ とし、$I_k \geqq 1$, $k = 0, 1, \ldots, m$, $I_{m+1} = 1$ とする。

$\vec{a} = (\vec{a}^{(0)}, \vec{a}^{(1)}, \ldots, \vec{a}^{(m)})$

$$\vec{a}^{(k)} = \{a_{i,j}^{(k)}\}_{i=1,\ldots,I_{k+1}\ j=1,\ldots,I_k+1} \in \mathbf{R}^{I_{k+1} \otimes (I_k+1)}, \qquad k = 0, \ldots, m,$$

と表すことにする。さらに、

$$\mathbf{A} = \oplus_{k=0}^{m} \mathbf{R}^{I_{k+1} \otimes (I_k+1)}$$

と表すことにする。

$h_i \in \mathcal{M}(E)$, $i = 1, \ldots, N_0$, とする。

$g_i^{(k)}(\cdot; \vec{a}) \in \mathcal{M}(E)$, $k = 0, 1, \ldots, m$, $i = 1, \ldots, I_{k+1}$ を以下のように帰納的に定める

$$g_i^{(0)}(x; \vec{a}) = \sum_{j=1,\ldots,I_0} a_{i,j}^{(0)} h_j + a_{i,I_0+1}^{(0)}$$

$$g_i^{(k+1)}(x; \vec{a}) = \sum_{j=1,\ldots,I_{k+1}} a_{i,j}^{(k+1)} \rho(g_j^{(k)}(x)) + a_{i,I_{k+1}+1}^{(k+1)}, \quad k = 0, 1, 2, \ldots, m-1.$$

$g_1^{(m)}(x; \vec{a})$ が出力である。

$$q_k(\vec{a}) = \max_{i=1,\ldots,I_k+1} \sum_{j=1}^{I_k+1} \sum_{j=1}^{I_k+1} |a_{i,j}^{(k)}|$$

$k = 0, 1, \ldots, m$, $\vec{a} \in \mathbf{A}$, とおく。

$\mathcal{K} = \{h_i;\ i = 1, \ldots, I_0\}$ とおき、$r_k \geqq 0$, $k = 0, 1, 2, \ldots,$ を定めると

$$\{g_i^{(k)}(\cdot; \vec{a});\ i = 1, \ldots, I_{k+1},\ q_\ell(\vec{a}) \leqq r_\ell,\ \ell = 0, 1, \ldots, k\} \subset \mathcal{K}_k, \qquad k = 0, 1, \ldots, m,$$

となることがわかる。よって，命題 より以下のことがわかる。

命題 6.2 $\gamma_0 > 0$, $C_0 \in (0, \infty)$ が存在して

$$\int_E \exp(\gamma_0 h_j(x)^2) \nu(dx) \leqq C_0, \qquad i = 1, \ldots, d,$$

かつ $\exp(\gamma_0) \leqq C_0$ と仮定する。この時、

$$\mathcal{R}_n(\{g_1^{(m)}(\cdot;\vec{a});\ q_\ell(\vec{a}) \leqq r_\ell,\ \ell=0,1,\ldots,k\};\ \nu) \leqq \frac{2^{k+1}C_1}{n^{1/2}} r_m \prod_{i=0}^{m-1}(r_i+1)$$

となる。ここで

$$C_1 = 2\gamma_0^{-1/2}(\log(\#(\mathcal{K}))^{1/2}(\log C_0 + \frac{1}{2})^{1/2}.$$

7 例

ここでは $\rho : \mathbf{R} \to \mathbf{R}$ は連続関数で、条件 (R-1), (R-2) を満たし、

$$|\rho(t)| \leqq 1, \qquad t \in \mathbf{R}$$

となると仮定する。

さらに、$\gamma_0 > 0$, $C_0 \in (0,\infty)$ が存在して

$$\int_E \exp(\gamma_0 h_j(x)^2)\nu(dx) \leqq C_0, \qquad i = 1,\ldots,d,$$

かつ $\exp(\gamma_0) \leqq C_0$ と仮定する。

今、$f : E \to \mathbf{R}$ は可測関数とし、$F : E \times \mathbf{A} \to \mathbf{R}$ を

$$F(x;\vec{a}) = (f(x) - g^{(m)}(x;\vec{a}))^2, \qquad x \in E$$

で定める。

例えば

$$\phi_0(\vec{a}) = \max_{k=0,\ldots,m}(q_k(\vec{a})+1)$$

とおき

$$\tilde{\mathcal{F}}_r = \{g_1^{(m)}(\cdot;\vec{a});\ \phi_0(\vec{a}) \leqq r\}$$

とおくと、

$$\mathcal{R}_n(\tilde{\mathcal{F}}_r;\nu) \leqq \frac{2^{m+1}C_1}{n^{1/2}}(r-1)r^m \qquad r \geqq 1.$$

となり、さらに

$$\sup\{|g(x)|;\ x \in E,\ g \in \tilde{\mathcal{F}}_r\} \leqq r-1, \qquad r \geqq 1.$$

となる。

場合１． $f : E \to \mathbf{R}$ は有界と仮定する。

今、$\mathcal{F}_r = \{F(x;\vec{a});\ \phi_0(\vec{a}) \leqq r\}$ とおく。$g \in \tilde{\mathcal{F}}_r$ ならば $(f(x)-g(x))^2 = (f(x) - ((-r) \vee g(x)) \wedge r)$ となる。

$\rho_{r,b}(t,x) = (f(x) - ((-r) \vee t) \wedge r)^2$ とおくと

$$|\frac{d}{dt}\rho_r(t,x)| \leqq 2|1_{A_b}(x)f(x) - ((-r)\vee t)\wedge r)| \leqq 2(||f||_\infty + r), \qquad t \in \mathbf{R},\ r \geqq 1$$

であるので、命題 3.4 より

$$\mathcal{R}_n(\mathcal{F}_r;\nu) \leqq 2(||f||_\infty + r)\mathcal{R}_n(\tilde{\mathcal{F}}_r;\nu) \leqq \frac{2^{m+1}C_1}{n^{1/2}}(||f||_\infty + r)r^{m+1}$$

であり、

$$\sup\{|g(x)|; x \in E,\ g \in \mathcal{F}_r\} \leqq (||f||_\infty + r)^2$$

となる。

よって、命題 5.1 より確率 $1-2\exp(-4a^2)/(1-\exp(-2a^2))$ で

$$
\begin{aligned}
&|\hat{F}_n(\vec{a})-F_0(\vec{a})| \\
&\leqq \frac{a}{n^{1/2}}(\phi_0(\vec{a})+||f||_\infty+1)^2(\phi_0(\vec{a})+1)^{1/2}+\frac{2^{m+2}C_1}{n^{1/2}}(\phi_0(\vec{a})+||f||_\infty+1)(\phi_0(\vec{a})+1)^{m+1}
\end{aligned}
$$

ただし、

$$
\hat{F}_n(\vec{a})=\frac{1}{n}\sum_{k=1}^{n}F(X_k,\vec{a}), \qquad F_0(\vec{a})=\int_E F(x,\vec{a})\nu(dx)
$$

である。今、$\phi:\mathbf{A}\to[0,\infty)$, $\lambda>0$ を定めて

$$
\theta_0=\mathrm{argmin}_{\vec{a}\in\mathbf{A}}F_0(\vec{a}),
$$

$$
\hat{\theta}_\lambda=\mathrm{argmin}_{\vec{a}\in\mathbf{A}}(\hat{F}_n(\vec{a})+\lambda\phi(\vec{a}))
$$

とおくと、

$$
\begin{aligned}
&F_0(\hat{\theta}_\lambda)-F_0(\theta_0) \\
&\leqq (\hat{F}_n(\hat{\theta}_\lambda)+\lambda\phi(\hat{\theta}_\lambda))-(\hat{F}_n(\theta_0)+\lambda\phi(\theta_0))-(\hat{F}_n(\hat{\theta}_\lambda)-F_0(\hat{\theta}_\lambda)-\lambda\phi(\hat{\theta}_\lambda)+(\hat{F}_n(\theta_0)-F_0(\theta_0))+\lambda\phi(\theta_0) \\
&\leqq \frac{a}{n^{1/2}}(\phi_0(\hat{\theta}_\lambda)+||f||_\infty+1)^2(\phi_0(\hat{\theta}_\lambda)+1)^{1/2}+\frac{2^{m+2}C_1}{n^{1/2}}(\phi_0(\hat{\theta}_\lambda)+||f||_\infty+1)(\phi_0(\hat{\theta}_\lambda)+1)^{m+1}-\lambda\phi(\hat{\theta}_\lambda) \\
&\quad +(\hat{F}_n(\theta_0)-F_0(\theta_0))+\lambda\phi(\theta_0)
\end{aligned}
$$

となる。

例えば、

$$
\phi(\theta)=(\phi_0(\theta)+1)^{m+2}, \qquad \theta\in\Theta
$$

とおくと、

$$
\begin{aligned}
&F_0(\hat{\theta}_\lambda)-F_0(\theta_0) \\
&\leqq \frac{a}{n^{1/2}}(||f||_\infty+1)^2(\phi_0(\hat{\theta}_\lambda)+1)^{5/2}+\frac{2^{m+2}C_1}{n^{1/2}}(||f||_\infty+1)(\phi_0(\hat{\theta}_\lambda)+1)^{m+2}-\lambda(\phi_0(\hat{\theta}_\lambda)+1)^{m+1} \\
&\quad +(\hat{F}_n(\theta_0)-F_0(\theta_0))+\lambda\phi(\theta_0)
\end{aligned}
$$

となり、

$$
\frac{a}{n^{1/2}}(||f||_\infty+1)^2+\frac{2^{m+2}C_1}{n^{1/2}}(||f||_\infty+1)\leqq\lambda
$$

であれば

$$
F_0(\hat{\theta}_\lambda)-F_0(\theta_0)\leqq(\hat{F}_n(\theta_0)-F_0(\theta_0))+\lambda\phi(\theta_0)
$$

となる。

場合 2. f が有界でない場合

$b>0$ に対して $A_b=\{x\in E;\ |f(x)|\leqq b\}$ とおく。$\tilde{\Omega}_{b,n}=\bigcap_{k=1}^{n}\{X_k\in A_b\}$ とおくと、$P(\tilde{\Omega}_{b,n})\geqq 1-n(1-\nu(A_b))$ となる。

$$
F_b(x;\vec{a})=(1_{A_b}(x)f(x)-g^{(m)}(x;\vec{a}))^2
$$

とおき、

$$
F_{b,0}(\vec{a})=\int_E F_b(x,\vec{a})\nu(dx)
$$

$$
\hat{F}_{b,n}(\vec{a})=\frac{1}{n}\sum_{k=1}^{n}F_b(X_k,\vec{a})
$$

とおく。

この時、$\omega \in \tilde{\Omega}_{b,n}$ ならば、

$$\hat{F}_n(\vec{a}) = \hat{F}_{b,n}(\vec{a})$$

であり、

$$|F_{b,0}(\vec{a}) - F_0(\vec{a})| \leqq |\int_E (F_b(x,\vec{a}) - F(x,\vec{a}))\nu(dx)|$$

$$\leqq 2\int_{E\setminus A_b} |f(x)g_1^{(m)}(x,\vec{a})|\nu(dx) + \int_{E\setminus A_b} f(x)^2\nu(dx)$$

$$\leqq 2\int_{E\setminus A_b} |f(x)|^2\nu(dx) + \sup_{x\in E}|g_1^{(m)}(x,\vec{a})|^2\nu(E\setminus A_b)$$

$\mathcal{F}_{b,r} = \{F_b(x;\vec{a});\ \phi_0(\vec{a}) \leqq r\}$ とおく。先と同様にして、命題 3.4 より

$$\mathcal{R}_n(\mathcal{F}_{b,r};\nu) \leqq 2(b+r)\mathcal{R}_n(\tilde{\mathcal{F}}_r;\nu) \leqq \frac{2^{m+1}C_1}{n^{1/2}}(b+r)r^{m+1}$$

であり、

$$\sup\{|g(x)|; x\in E,\ g\in \mathcal{F}_r\} \leqq (b+r)^2$$

となる。

よって、命題 5.1 より確率 $1-2\exp(-4a^2)/(1-\exp(-2a^2))$ で

$$|\hat{F}_{b,n}(\vec{a}) - F_{b,0}(\vec{a})|$$

$$\leqq \frac{a}{n^{1/2}}(\phi_0(\vec{a})+b+1)^2(\phi_0(\vec{a})+1)^{1/2} + \frac{2^{m+2}C_1}{n^{1/2}}(\phi_0(\vec{a})+b+1)(\phi_0(\vec{a})+1)^{m+1}.$$

この時、$\omega \in \tilde{\Omega}_{b,n}$ であれば

$$|\hat{F}_n(\vec{a}) - F_0(\vec{a})| \leqq |\hat{F}_{b,n}(\vec{a}) - F_{b,0}(\vec{a})| + 2\int_{E\setminus A_b}|f(x)|^2\nu(dx) + \phi_0(\vec{a})^2\nu(E\setminus A_b)$$

であるので、確率 $1-2\exp(-4a^2)/(1-\exp(-2a^2)) - n\nu(E\setminus A_b)$ で

$$|\hat{F}_n(\vec{a}) - F_0(\vec{a})|$$

$$\leqq \frac{a}{n^{1/2}}(\phi_0(\vec{a})+b+1)^2(\phi_0(\vec{a})+1)^{1/2} + \frac{2^{m+2}C_1}{n^{1/2}}(\phi_0(\vec{a})+b+1)(\phi_0(\vec{a})+1)^{m+1}$$

$$+2\int_{E\setminus A_b}|f(x)|^2\nu(dx) + \phi_0(\vec{a})^2\nu(E\setminus A_b)$$

となる。

例えば、$\gamma_2 > 0$ が存在して

$$C_2 = \int_E \exp(\gamma_2^{-1}f(x)^2)\nu(dx) < \infty$$

であるのならば、 $b = \log n + \tilde{a}$ とおけば

$$n\nu(E\setminus A_b) \leqq n\exp(-\gamma_2^{-1}(\log n + \tilde{a})^2)C_2$$

$$\leqq C_2 n\exp(-\gamma_2^{-1}(\log n)^2)\exp(-\gamma_2^{-1}\tilde{a}^2)$$

また、

$$\int_{E\setminus A_b}|f(x)|^2\nu(dx) \leqq 2\gamma_2\int_{E\setminus A_b}\exp(\frac{1}{2\gamma_2}|f(x)|^2)\nu(dx)$$

$$\leqq 2\gamma_1(\int_{E\setminus A_b}\exp(\frac{1}{\gamma_2}|f(x)|^2)\nu(dx))^{1/2}\nu(E\setminus A_b)^{1/2}$$

$$\leqq 2\gamma_2 C_2 n^{1/2} \exp(-\frac{1}{2\gamma_2}(\log n)^2)\exp(-\frac{1}{2\gamma_2}\tilde{a}^2)$$

よって、確率 $1-2\exp(-4a^2)/(1-\exp(-2a^2)) - C_2 n\exp(-\gamma_2^{-1}(\log n)^2)\exp(-\gamma_2^{-1}\tilde{a}^2)$ で

$$|\hat{F}_n(\vec{a}) - F_0(\vec{a})|$$

$$\leqq \frac{a}{n^{1/2}}(\phi_0(\vec{a}) + \log n + \tilde{a})^2(\phi_0(\vec{a})+1)^{1/2} + \frac{2^{m+2}C_1}{n^{1/2}}(\phi_0(\vec{a}) + \log n + \tilde{a} + 1)(\phi_0(\vec{a})+1)^{m+1}$$

$$+4\gamma_2 C_2 n^{1/2}\exp(-\frac{1}{2\gamma_2}(\log n)^2)\exp(-\frac{1}{2\gamma_2}\tilde{a}^2) + \phi_0(\vec{a})^2 C_2 n\exp(-\gamma_2^{-1}(\log n)^2)\exp(-\gamma_2^{-1}\tilde{a}^2)$$

となる。