# ノンパラメトリックバウンドについて

情報理工学系研究科　数理情報学専攻
数理第四研究室　博士三年
指導教員：　駒木　文保　准教授
**鈴木　大慈**

2008年8月14日

# 発表の流れ

- 経験過程の一般論
  – ドンスカークラスの十分条件
  – 凸コスト最小化におけるノンパラメトリックバウンド
  – Tsybakovの低雑音条件

# 経験過程の理論
# （一様大数の法則，一様中心極限定理）

$$X_i \sim \text{i.i.d.}$$

$$Pf = \mathrm{E}_P f, \qquad P_n f = \frac{\sum_{i=1}^n f(X_i)}{n}$$

## 大数の法則

$$(P - P_n)f \xrightarrow{\text{a.s.}} 0$$

## 中心極限定理

$$\sqrt{n}(P - P_n)f \xrightarrow{\text{分布収束}} \text{正規分布}$$

# d個の関数

$$\{f_j \mid j = 1, \ldots, d\}$$

$$Pf = \mathrm{E}_P f, \qquad P_n f = \frac{\sum_{i=1}^n f(X_i)}{n}$$

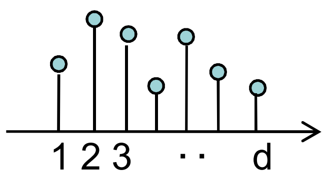
## 大数の法則

$$\max_{1 \le j \le d} |(P - P_n) f_j| \longrightarrow 0$$

## 多変量中心極限定理

$$\sqrt{n}(P - P_n) \begin{pmatrix} f_1 \\ \vdots \\ f_d \end{pmatrix} \longrightarrow$$

多変量
正規分布



# 無限(個)の関数

$$Pf = \mathrm{E}_P f, \qquad P_n f = \frac{\sum_{i=1}^n f(X_i)}{n}$$

$$\mathcal{F} := \{f_\theta \mid \theta \in \Theta\}$$

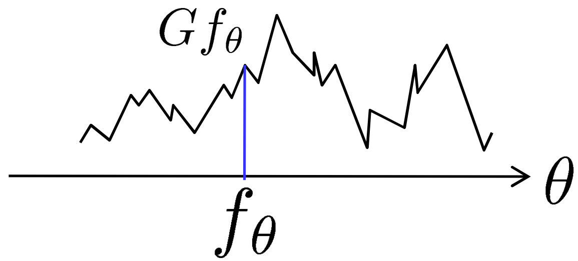
## 一様大数の法則

$$\sup_{f \in \mathcal{F}} |(P - P_n)f| \longrightarrow 0$$

## 一様中心極限定理

$l^\infty(\mathcal{F})$上の分布
として分布収束

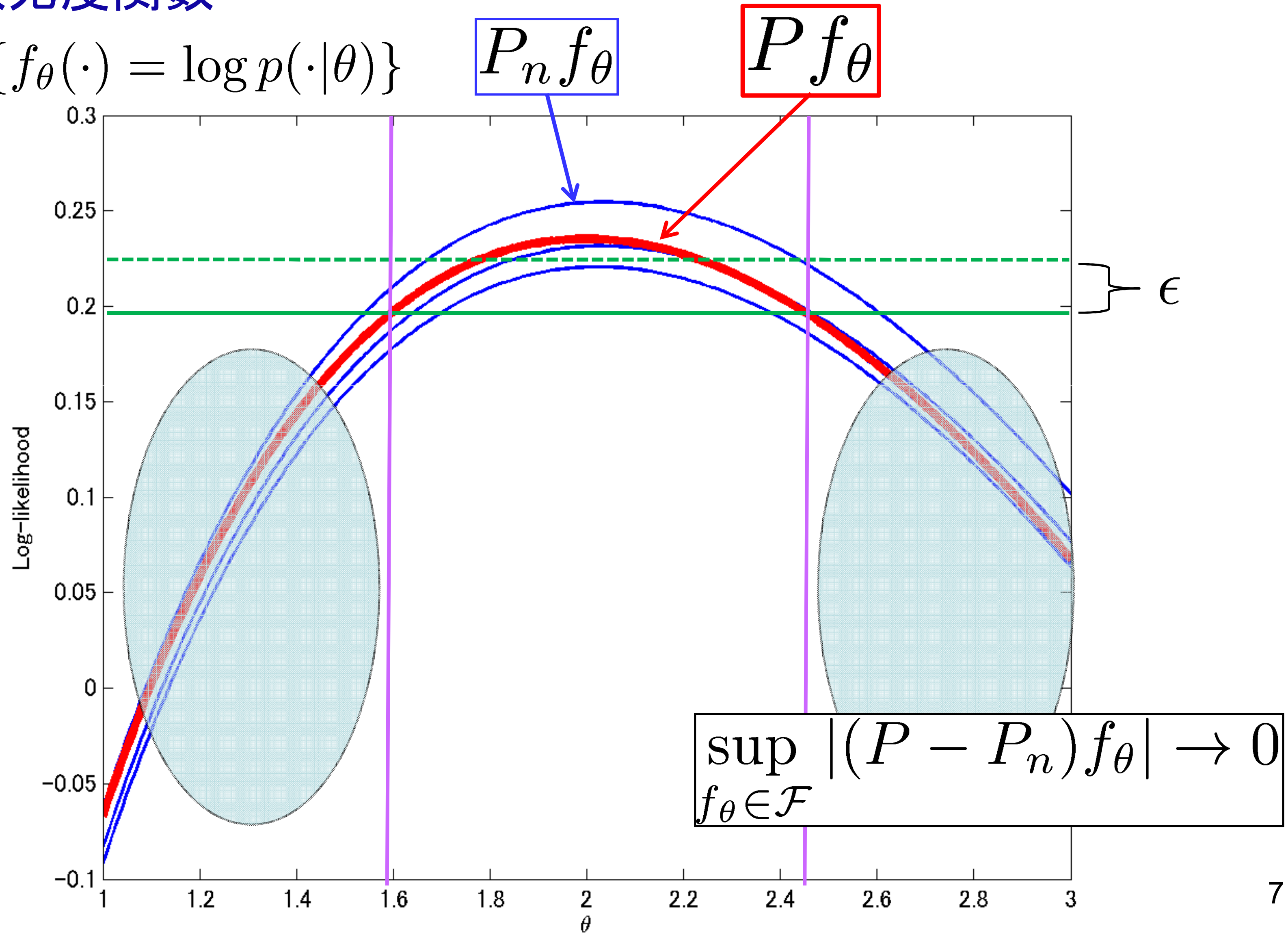
$$\sqrt{n}(P - P_n) \begin{pmatrix} \vdots \\ f_\theta \\ \vdots \end{pmatrix} \longrightarrow \text{ガウシアンプロセス}$$



# 例: 一様大数の法則

- 対数尤度関数

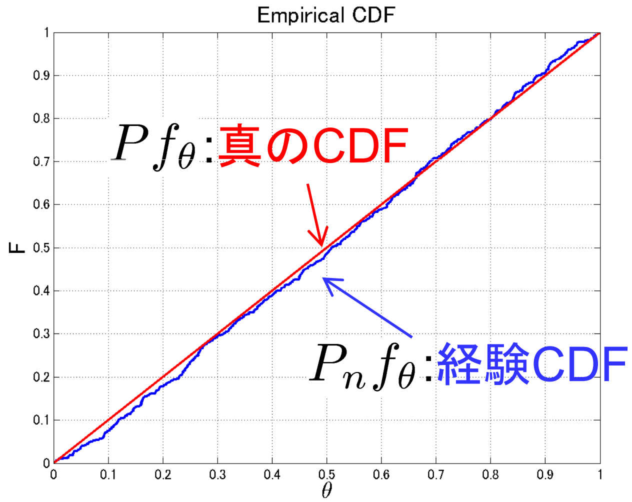
$$\mathcal{F} = \{f_\theta(\cdot) = \log p(\cdot|\theta)\}$$



# 例：一様中心極限定理

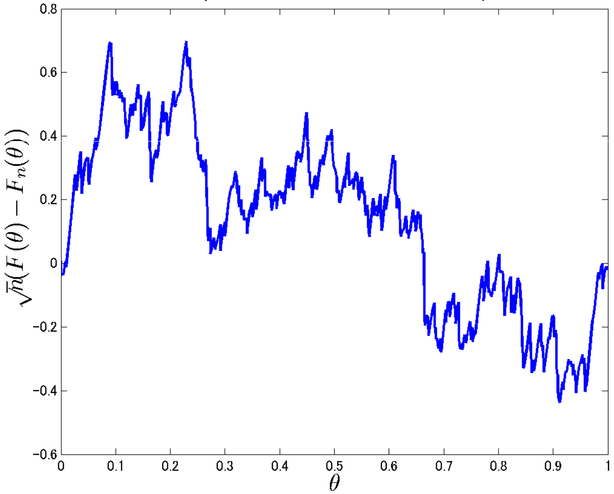
- 経験累積分布関数

$$f_\theta(x) = \mathbf{1}\{x \le \theta\} \quad (0 \le \theta \le 1)$$




一様分布

$$\sqrt{n}(Pf_\theta - P_n f_\theta)$$




経験過程
(Kolmogorov-Smirnov検定)

- 以降，一様中心極限定理が成り立つ関数族を考える.

※ 一様中心極限定理 → 一様大数の法則

- ($P$-)ドンスカークラス
  - 一様中心極限定理が成り立つ関数集合

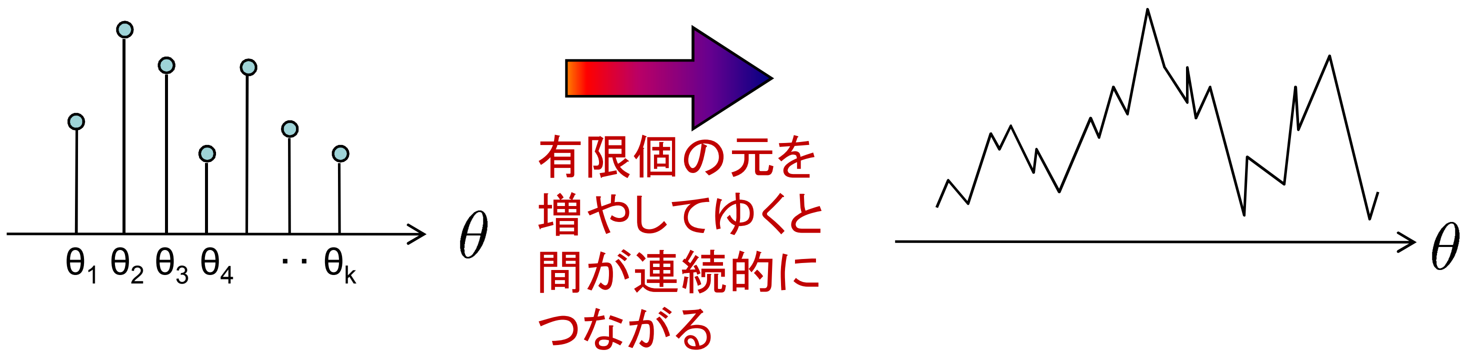
$$\sqrt{n}(P - P_n) \rightsquigarrow G \quad \text{in} \quad l^{\infty}(\mathcal{F})$$

# ドンスカーの必要十分条件

1．有限個の元を任意に持ってきて，
その経験過程がある正規分布に収束する．

$$\sqrt{n}(P_n - P)(f_1, \ldots, f_k) \rightsquigarrow N_k(0, \Sigma)$$

2．有限個の元で$\mathcal{F}$をうまく近似できる
（漸近等連続性⇔漸近タイト）



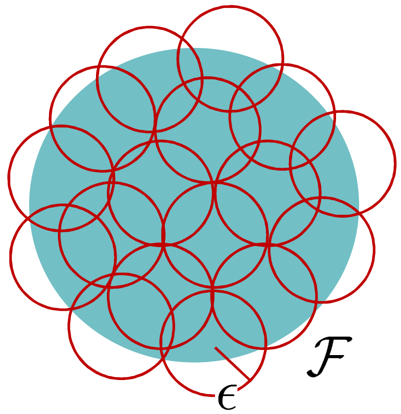
# 2. の十分条件

関数集合: $\mathcal{F}$

- 準備:関数集合の複雑さ

ε-カバリングナンバー

$\boxed{N(\mathcal{F}, \epsilon, d)}$




:ノルム$d$ によるε-ボールで $\mathcal{F}$ を覆うのに必要な最小のボールの個数

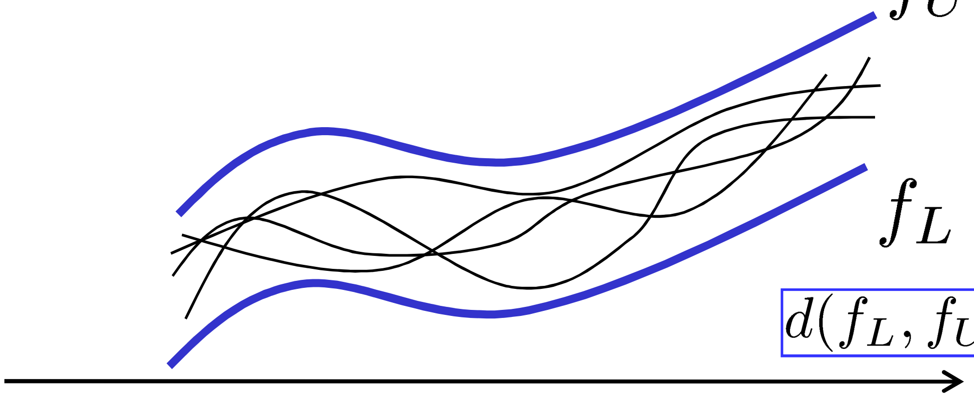
---

ε-ブラケッティングナンバー

$\boxed{N_{[]}(\mathcal{F}, \epsilon, d)}$

$\mathcal{F} \subset \cup_{i=1}^{K} \{ f \mid f_{L_i} \le f \le f_{U_i} \}$

と出来るような最小のK




$\boxed{d(f_L, f_U) \le \epsilon}$

# 2. の十分条件

関数集合: $\mathcal{F}$

- 一様エントロピー条件

$$\int \sup_{Q} \sqrt{\log N(\mathcal{F}, \epsilon, L_2(Q))} \mathrm{d}\epsilon < \infty$$

有限離散確率測度の中でsupとる

または

$$\int \sqrt{\log N_{[]}(\mathcal{F}, \epsilon, L_2(P))} \mathrm{d}\epsilon < \infty$$

Dudley積分という

# カバリングナンバーの例

- *d* 次元，有界

$\mathcal{F} = \{f = \sum_{i=1}^{d} \alpha_i f_i \mid \|f\|_\infty \le 1\}$

$$\boxed{\sup_Q N(\mathcal{F}, \epsilon, L_2(Q)) \le \left(\frac{4+\epsilon}{\epsilon}\right)^d}$$

- VC次元 *V*，有界

$$\boxed{\sup_Q N(\mathcal{F}, \epsilon, L_2(Q)) \le K_V \left(\frac{1}{\epsilon}\right)^{2(V-1)}}$$

それのconvex hull

$\boxed{\rho = 1 - V^{-1}}$

$$\boxed{\sup_Q \log N(\mathcal{F}, \epsilon, L_2(Q)) \le C_V \left(\frac{1}{\epsilon}\right)^{2(1-V^{-1})}}$$

# カバリング/ブラケッティングナンバーの例

•ガウスカーネルにより生成されるRKHSの単位球（d次元，コンパクト集合上）

$$0 < \forall p < 2$$

$$\log N(B_H, \epsilon, \|\cdot\|_\infty) \leq c_{p,d}\sigma^{(1-p/4)d}\epsilon^{-p}$$

[Steinwart, Scovel: A.S. 2005]

•ソボレフ空間の単位球：α階連続微分可能なd次元実数空間上の関数

$$\log N(C_1^\alpha, \epsilon, \|\cdot\|_\infty) \leq K\left(\frac{1}{\epsilon}\right)^{d/\alpha}$$

$\boxed{\text{次元が高いほど複雑，滑らかなほど単純}}$

# 一様なバウンド

関数集合Fの部分集合で二乗ノルムがδ以下の集合

$$\mathcal{F}(\delta) := \{f \in \mathcal{F} \mid Pf^2 \leq \delta^2\}$$

$$\mathrm{E}_P\left[\sup_{f \in \mathcal{F}(\delta)} |(P_n - P)f|\right] \leq \frac{C}{\sqrt{n}} \int_0^{\delta} \sup_Q \sqrt{\log N(\mathcal{F}, \epsilon, L_2(Q))} \mathrm{d}\epsilon$$

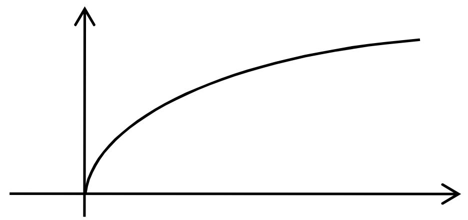
ブラケットでも似たような不等式が成り立つ

$0 < \rho < 1$

$\sup_Q \log N(\mathcal{F}, \epsilon, L_2(Q)) \leq C\epsilon^{-2\rho}$

の場合

$$\leq O\left(\frac{\delta^{1-\rho}}{\sqrt{n}}\right)$$



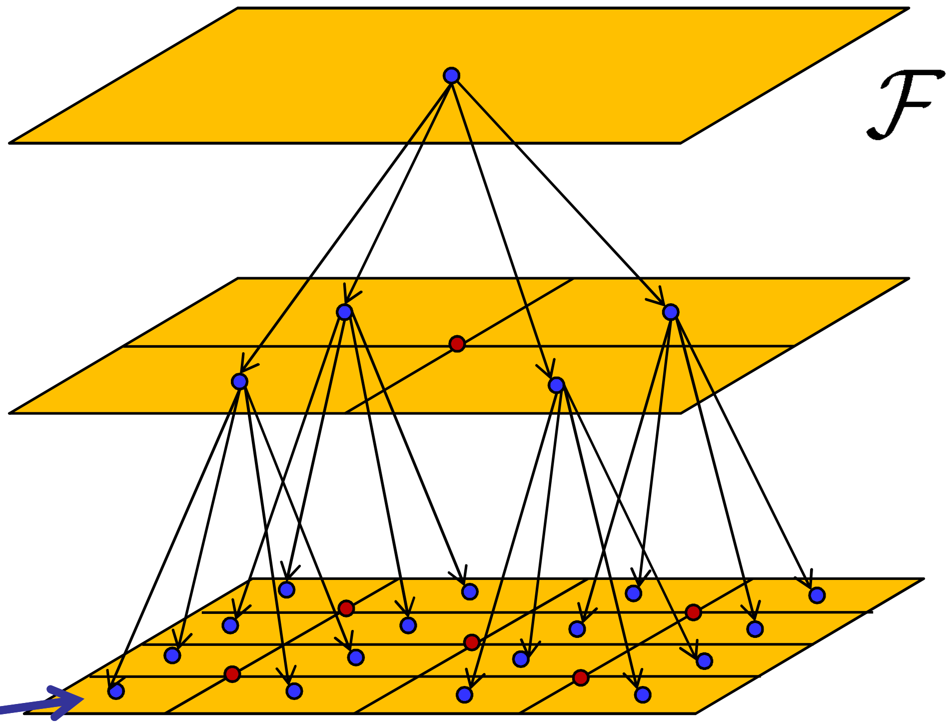
# Dudley積分について

※ カバリングナンバーは関数集合$\mathcal{F}$の複雑さを表す.

関数集合$\mathcal{F}$を有限個の元で近似するのに必要な個数を表している.

積分は解像度を上げてゆくことに対応.

同じε-ボールの中に入っている元は高々2εの距離にある.



# Dudley積分の雰囲気をつかむ

- Hoeffdingの不等式　　→一様カバリングナンバー

$Z_i \ (i = 1, \ldots, n)$ :独立で期待値0の確率変数 s.t. $|Z_i| \leq m_i$

$$P\left(\frac{|\sum_{i=1}^n Z_i|}{\sqrt{n}} > x\right) \leq 2\exp\left(-\frac{x^2}{2\sum_{i=1}^n m_i^2/n}\right)$$

---

- Bernsteinの不等式　　→ブラケッティングナンバー

$Z_i \ (i = 1, \ldots, n)$ :独立で期待値0の確率変数 s.t. $E|Z_i|^2 = v_i, \ |Z_i| \leq M$

$$P\left(\frac{|\sum_{i=1}^n Z_i|}{\sqrt{n}} > x\right) \leq 2\exp\left(-\frac{x^2}{2(v + \frac{Mx}{\sqrt{n}})}\right)$$

ただし, $v = \dfrac{\sum_{i=1}^n v_i}{n}$

# Maximal-Inequality1

- Hoeffdingの不等式に対するMaximal-Ineq.

$\mathcal{F} := \{f_j \ (j = 1, \ldots, m)\}$　有限個の関数集合:どれも期待値0

$$P\left(\frac{|\sum_{i=1}^n f_j(X_i)|}{\sqrt{n}} > x\right) \leq 2\exp\left(-\frac{x^2}{2\|f_j\|_\infty^2}\right)$$

Maximal Inequality

$$E\left(\max_{1\leq j\leq m}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n f_j(X_i)\right|\right) \leq C\max_j \|f_j\|_\infty \underline{\sqrt{\log(1+m)}}$$

これを，解像度を細かくしていって
積分（チェイニング）したのがDudley積分

# Maximal–Inequality2

- Bernsteinの不等式に対するMaximal-Ineq.

$\mathcal{F} := \{f_j \ (j = 1, \dots, m)\}$　有限個の関数集合：どれも期待値0

$$P\left(\frac{|\sum_{i=1}^n f_j(X_i)|}{\sqrt{n}} > x\right) \le 2\exp\left(-\frac{x^2}{2(\|f_j\|_2^2 + \frac{\|f_j\|_\infty x}{\sqrt{n}})}\right)$$

Maximal Inequality

$$E\left(\max_{1\le j\le m}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n f_j(X_i)\right|\right)$$

$$\lesssim \max_j \frac{\|f_j\|_\infty}{\sqrt{n}}\log(1+m) + \max_j \|f_j\|_2 \underline{\sqrt{\log(1+m)}}$$

# バウンドを出してみる

# 設定

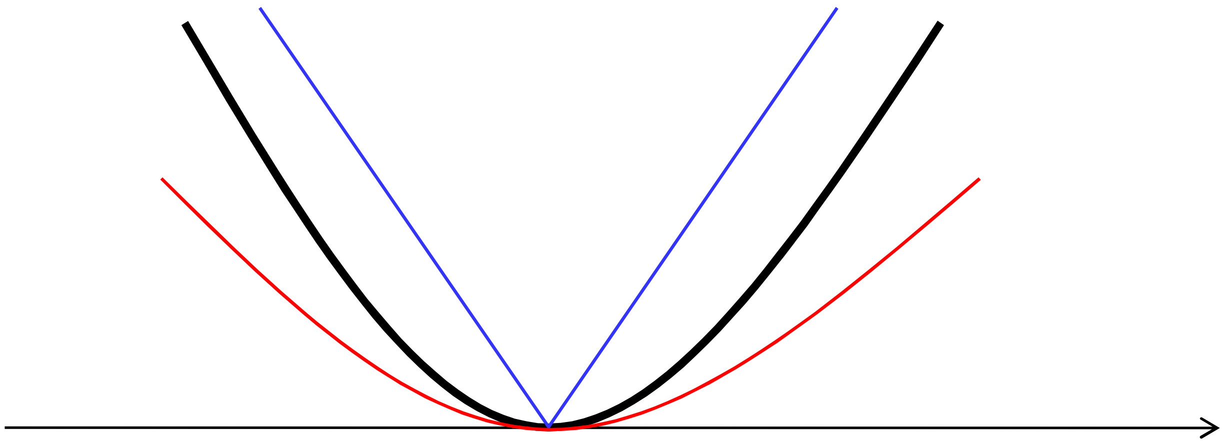
- 凸コスト最小化：正則化項なし

$$\min_{f\in\mathcal{F}} P[l(y,f(x))]$$

ロス関数$l$の条件： $\forall y,u,v$

- リプシッツ連続 $$|l(y,u)-l(y,v)| \le L|u-v|$$

- Modulus of convexity $$c|u-v|^2 \le \frac{l(y,u)+l(y,v)}{2} - l\left(y,\frac{u+v}{2}\right)$$



$f^*$: 最適解

## 関数集合$\mathcal{F}$の条件

1. 一様有界 $\|f\|_\infty \le M \quad (\forall f \in \mathcal{F})$

2. 多項式複雑さ $0 < \exists \rho < 1, \quad \sup_Q \log N(\mathcal{F}, \epsilon, L_2(Q)) \le C\epsilon^{-2\rho}$

$\Rightarrow$ $\mathcal{G} := \{l(\cdot, f(\cdot)) - l(\cdot, f^*(\cdot)) \mid f \in \mathcal{F}\}$ とすると

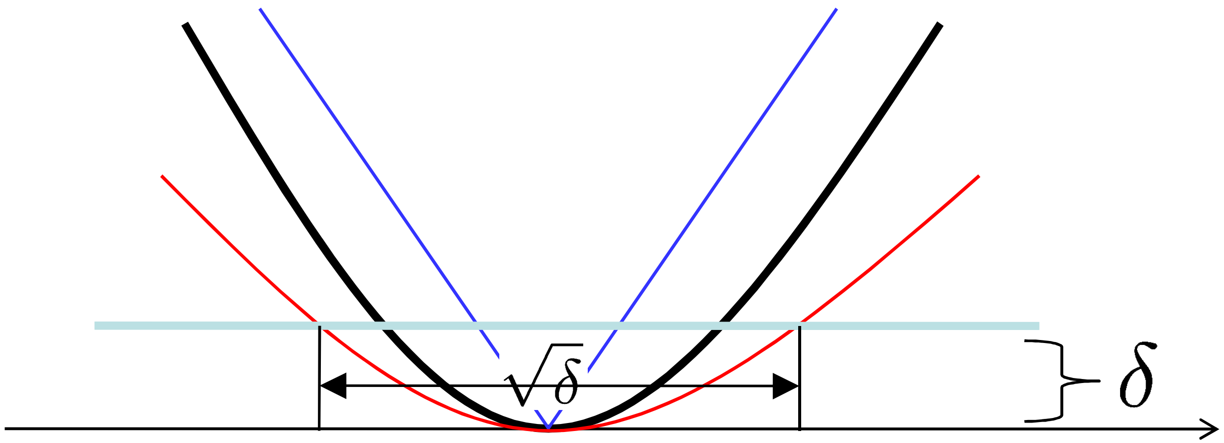
リプシッツ連続性より $\sup_Q \log N(\mathcal{G}, \epsilon, L_2(Q)) \le C'\epsilon^{-2\rho}$

---

Modulus of convexity

$$cP(\hat{f} - f^*)^2 \le P(l(\hat{f}) + l(f^*))/2 - Pl\left(\frac{\hat{f} + f^*}{2}\right) \le P(l(\hat{f}) - l(f^*))/2$$

$\Rightarrow$ $P(l(\hat{f}) - l(f^*))^2 \lesssim P(l(\hat{f}) - l(f^*))$



$$\mathcal{G}(\delta) := \{g \in \mathcal{G} \mid \sqrt{Pg^2} \le \delta\}$$

$$P(l(\hat{f}) - l(f^*))^2 \lesssim P(l(\hat{f}) - l(f^*))$$

$$= (P - P_n)(l(\hat{f}) - l(f^*)) + \underbrace{P_n(l(\hat{f}) - l(f^*))}_{\leqq 0}$$

$$\le (P - P_n)(l(\hat{f}) - l(f^*))$$

$$\le \delta^2$$

$$\boxed{l(\hat{f}) - l(f^*) \in \mathcal{G}(\delta)}$$

$$\mathrm{E}\left[\sup_{g \in \mathcal{G}(\delta)} |(P_n - P)g|\right] \le \frac{C}{\sqrt{n}}\delta^{1-\rho}$$

$$\boxed{\delta^2 = \frac{1}{\sqrt{n}}\delta^{1-\rho} \Rightarrow \delta = n^{-\frac{1}{2(1+\rho)}}}$$

$$\boxed{P(l(\hat{f})) - P(l(f)) = O_p(n^{-\frac{1}{1+\rho}})}$$

---

**Talagrand's Concentration Inequality**

$$P\left[\sup_{g \in \mathcal{G}(\delta)} (P_n - P)(g) \ge C\left(\mathrm{E}[\sup_{g \in \mathcal{G}(\delta)} (P_n - P)(g)] + \sqrt{\frac{t}{n}\delta} + \frac{t}{n}\right)\right] \le e^{-t}$$

$$\boxed{\sim \delta^2}$$

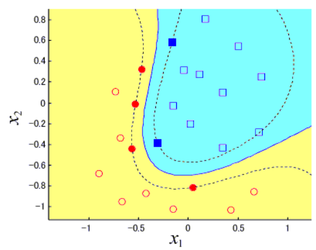
$$\mathrm{E}\left[\sup_{g \in \mathcal{G}(\delta)} |(P_n - P)g|\right] \le \frac{C}{\sqrt{n}}\delta^{1-\rho}$$ 一様バウンド

# Tsybakovの低雑音条件

# Tsybakovの低雑音条件

$\mathcal{X}$ :入力変数の空間　　　$\mathcal{Y} = \{\pm 1\}$ :出力変数の空間

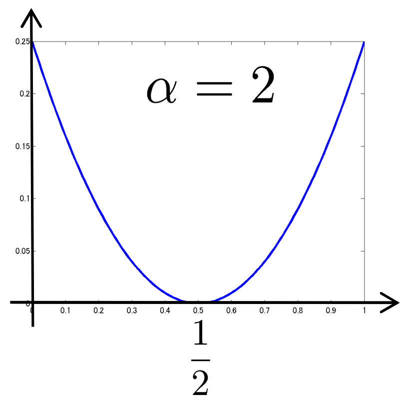
$P(X,Y) \quad (X \in \mathcal{X}, Y \in \{\pm 1\})$ :$\mathcal{X}$,$\mathcal{Y}$上の確率分布



Tsybakovの低雑音条件 (Noise Exponent α)

$$\eta(X) := P(Y=1|X)$$

$$\exists \alpha \geq 0 \qquad P_X\left(\left|\eta(X) - \frac{1}{2}\right| \leq t\right) \leq Ct^{\alpha}$$




$$\kappa = \frac{1+\alpha}{\alpha} \quad \text{としておく}$$

[Tsybakov: 2004, A.S.]

# 二値判別・経験誤差最小化

N個のサンプル $\{(x_i, y_i)\}_{i=1}^N$

用いる判別機の集合（モデル） $\mathcal{F}: \mathcal{X} \to \{\pm 1\}$

$\boxed{f \in \mathcal{F}}$

経験リスク

$$\hat{R}(f) = \frac{\sum_{i=1}^N \mathbf{1}\{f(x_i) \neq y_i\}}{N}$$

真のリスク (誤り確率)

$$R(f) = P(f(X) \neq Y)$$

経験リスク最小化

$$\hat{f} := \arg\min_{f \in \mathcal{F}} \hat{R}(f)$$

真の最適解

$$f^* := \arg\min_{f \in \mathcal{F}} R(f)$$

$\sup_Q \log N(\mathcal{F}, \epsilon, L_2(Q)) \leq C' \epsilon^{-2\rho}$ :モデルFの複雑さはρ (0≦ ρ ≦ 1)

$\arg\min_{f:\mathcal{X}\to\{\pm 1\}} R(f) = f^* \in \mathcal{F}$ :真の解はモデルに含まれている

# 低雑音条件における汎化誤差

経験リスク最小化元の真のリスク

$$R(\hat{f}) - R(f^*) = R(\hat{f}) - \hat{R}(\hat{f}) + \overbrace{\hat{R}(\hat{f}) - \hat{R}(f^*)}^{\leqq 0} + \hat{R}(f^*) - R(f^*)$$

$$\leq \underbrace{R(\hat{f}) - \hat{R}(\hat{f})}_{O_p(1/\sqrt{N})} + \underbrace{\hat{R}(f^*) - R(f^*)}_{O_p(1/\sqrt{N})} \quad = O_p(1/\sqrt{N})$$

経験過程の理論

しかし

Tsybakovの低雑音条件のもとでは

Fast Learning Rate

$$\boxed{R(\hat{f}) - R(f^*) = O_p\left(N^{-\frac{\kappa}{2\kappa+\rho-1}}\right)}$$

$O_p(1/\sqrt{n})$より速い $\left(\frac{\kappa}{2\kappa+\rho-1} = \frac{1}{2-\frac{1-\rho}{\kappa}}\right)$

[Tsybakov: 2004, A.S.]

# 論理の概要

Tsybakovの低雑音条件

$$P(f-f^*)^2 \le C_\alpha (R(f)-R(f^*))^{\frac{1}{\kappa}}$$ :リスクが低いなら，真との距離も近い

$$\mathcal{F}(\delta) := \{f \mid P(f-f^*)^2 \le \delta^2\}$$

一様なバウンド

$$\mathrm{E}_P\left[\sup_{f\in\mathcal{F}(\delta)} |\hat{R}(f)-R(f)-(\hat{R}(f^*)-R(f^*))|\right] \le C\left(\frac{\delta^{1-\rho}}{\sqrt{N}}\right)$$

$R(\hat{f})-R(f^*)\le\delta$ ならば $\hat{f}\in\mathcal{F}(\sqrt{C_\alpha}\delta^{\frac{1}{2\kappa}})$

すると高い確率で

$$R(\hat{f})-R(f^*) \le R(\hat{f})-R_n(\hat{f})-(R(f^*)-R_n(f^*)) \le C\frac{1}{\sqrt{N}}\delta^{\frac{1-\rho}{2\kappa}}$$

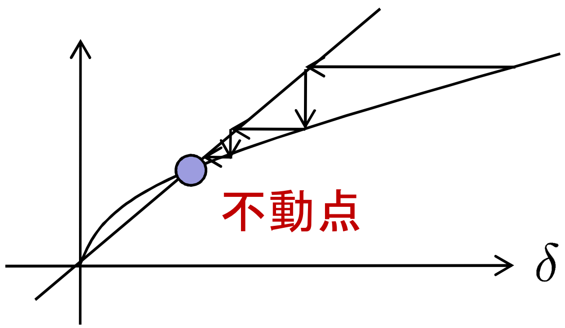
$\hat{f}$ の最適性　　一様バウンド

（cf:Talagrand's Concentration Inequality）

$\delta=1$ から同じ論理を繰り返すと

$$\delta = \frac{1}{\sqrt{N}}\delta^{\frac{1-\rho}{2\kappa}}$$

$$\Rightarrow \delta = N^{-\frac{\kappa}{2\kappa+\rho-1}}$$




$C\frac{1}{\sqrt{N}}\delta^{\frac{1-\rho}{2\kappa}}$

$$R(\hat{f}) - R(f^*) = R(\hat{f}) - \hat{R}(\hat{f}) + \hat{R}(\hat{f}) - \hat{R}(f^*) + \hat{R}(f^*) - R(f^*)$$

$$\leq \boxed{R(\hat{f}) - \hat{R}(\hat{f})} + \boxed{\hat{R}(f^*) - R(f^*)}$$

これらを別々に扱ってはいけない.
近いものは似た振る舞いをする.

# 終了