6 Rates of Convergence

M 推定量の収束レートについて議論する．

$\mathbb{P}_n$ を分布 P からのサイズ n のサンプルによる経験分布とする．任意の $\theta \in \Theta$ に対して，$x \mapsto m_\theta(x)$ は可測関数であるとする．$\hat{\theta}_n$ は規準関数 $\theta \mapsto \mathbb{P}_n m_\theta$ をほとんど最大にするような推定量であるとする．

規準関数は $\theta \mapsto Pm_\theta$ と $\theta \mapsto \mathbb{P}_n m_\theta - Pm_\theta$ の和となっている．$\hat{\theta}_n$ の収束レートはこれらの写像の組み合わせ次第である．もし，前者が θ が最大点から離れるにつれ速く変化し，後者が小さいとすれば，$\hat{\theta}_n$ の収束レートは高くなる．

簡便のため，変動を経験過程 $\mathbb{G}_n m_\theta = \sqrt{n}(\mathbb{P}_n m_\theta - pm_\theta)$ で測ることにする．

Theorem 6.1 (Rate of convergence)**.** 固定された定数 C と $\alpha > \beta$ について，任意の n と十分小さな $\delta > 0$ に対して，

$$\sup_{d(\theta,\theta_0)<\delta} P(m_\theta - m_{\theta_0}) \le -C\delta^\alpha,$$
$$E^* \sup_{d(\theta,\theta_0)<\delta} |\mathbb{G}_n(m_\theta - m_{\theta_0}) \le C\delta^\beta$$

が成り立つと仮定する．もし，$\hat{\theta}_n$ が

$$\mathbb{P}_n m_{\hat{\theta}_n} \ge \mathbb{P}_n m_{\theta_0} - O_P(n^{\alpha/(2\beta-2\alpha)})$$

を満たし，θ_0 に外確率収束するならば

$$n^{1/(2\alpha-2\beta)} d(\hat{\theta}_n, \theta_0) = O_P^*(1)$$

となる．

Proof. $r_n = n^{1/2\alpha-2\beta}$ とする．$\hat{\theta}_n$ は規準関数 $\theta \mapsto \mathbb{P}_n m_\theta$ を $R_n = O_P(r_n^{-\alpha})$ まで最大化しているとする．

※ これは

$$\mathbb{P}_n - \sup_\theta \mathbb{P}_n m_\theta \ge -O_P(r_n^{-\alpha})$$

を意味している．

各 n に対して，$\Theta - \{\theta_0\}$ は $j \in \mathbb{Z}_+$ として，$S_{j,n} = \{\theta : 2^{j-1} < r_n d(\theta, \theta_0) \le 2^j\}$ に分割できる．

ある $M \in \mathbb{Z}_+$ について，$r_n d(\hat{\theta}_n, \theta_0) > 2^M$ であれば，$\hat{\theta}_n$ は $j > M$ なる $S_{j,n}$ のどれか 1 つに含まれる．このとき，$\hat{\theta}_n$ に対する仮定より，

$$\sup_{\theta \in S_{j,n}} (\mathbb{P}_n m_\theta - \mathbb{P}_n m_{\theta_0}) \ge -R_n$$

である．これより，任意の $\epsilon > 0$ に対して，

$$\begin{aligned} P^*(r_n d(\hat{\theta}_n, \theta_0) > 2^M) \le & \sum_{j>M, 2^j \le \epsilon r_n} P^* \left(\sup_{\theta \in S_{j,n}} (\mathbb{P}_n m_\theta - \mathbb{P}_n m_{\theta_)}) \ge -\frac{K}{r_n^\alpha} \right) \\ & + P^*(2d(\hat{\theta}_n, \theta_0) \ge \epsilon) + P(r_n^\alpha R_n \ge K) \end{aligned}$$

が成り立つ．これを示す．

$$\begin{aligned} P(A) &\le P(A \cap B) + P(B^c) \\ &\le P(A \cap B \cap C) + P(C^c) + P(B^c) \end{aligned}$$

で，$A = \{r_n d(\hat{\theta}_n, \theta_0) > 2^M\}$, $B = \{2d(\hat{\theta}_n, \theta_0) < \epsilon\}$, $C = \{r_n^\alpha R_n < K\}$ とおく．このとき，$A \cap B \cap C$ という事象を考える．この事象は $\{2^M < r_n d(\hat{\theta}_n, \theta_0) < \frac{\epsilon r_n}{2}$ かつ $r_n^\alpha R_n < K\}$. このとき先ほどの議論により，ある j $(j > M, 2^j \leq \epsilon r_n)$ が存在して，$\theta \in S_{j,n}$ で，このとき，

$$\sup_{\theta \in S_{j,n}} (\mathbb{P}_n m_\theta - \mathbb{P}_n m_{\theta_0}) \geq -R_n \geq -\frac{K}{r_n^\alpha}$$

以上により先の式が示せた．

ここで，$\hat{\theta}_n \xrightarrow{\mathrm{P}} \theta_n$ より，$P^*(2d(\hat{\theta}_n, \theta_0) \geq \epsilon) \to 0$.

次に，$P(r_n^\alpha R_n \geq K)$ は $R_n = O_P(r_n^{-\alpha})$ から $r_n^\alpha R_n = O_P(1)$ より，K を十分小さくとれば，n について一様に，任意の小さくすることができる．

最後に，$\sum_{j>M, 2^j \leq \epsilon r_n} P^* \left(\sup_{\theta \in S_{j,n}} (\mathbb{P}_n m_\theta - \mathbb{P}_n m_{\theta_)}) \geq -\frac{K}{-r_n^\alpha} \right)$ について．

まず，$\epsilon > 0$ を定理の条件が $\delta \geq \epsilon$ で成り立つように十分小さくとる．すると和に現れる任意の j で

$$\sup_{\theta \in S_{j,n}} P(m_\theta - m_{\theta_0}) \leq -C \frac{2^{(j-1)\alpha}}{r_n^\alpha}.$$

結局，$\frac{1}{2} C 2^{(M-1)\alpha} \geq K$ に対して，

$$\begin{aligned}
\sum_{j>M, 2^j \leq \epsilon r_n} P^* \left(||\mathbb{G}_n (m_\theta - m_{\theta_0})||_{S_{j,n}} \geq C\sqrt{n} \frac{2^{(j-1)\alpha}}{2r_n^\alpha} \right) &\leq \sum_{j \geq M} \frac{C(2^j/r_n)^\beta}{C\sqrt{n} 2^{(j-1)\alpha}/2r_n^\alpha} \\
&= \sum_{j \geq M} \frac{(2^j/r_n)^\beta 2r_n^\alpha}{\sqrt{n} 2^{(j-1)\alpha}}.
\end{aligned}$$

$M = M_n \to \infty$ で右辺はゼロに収束する．これを用いれば，

$$\begin{aligned}
&\sum_{j>M, 2^j \leq \epsilon r_n} P^* \left(\sup_{\theta \in S_{j,n}} (\mathbb{P}_n m_\theta - \mathbb{P}_n m_{\theta_0}) \geq -\frac{K}{r_n^\alpha} \right) \\
\leq &\sum_{j>M, 2^j \leq \epsilon r_n} P^* \left(\sup_{\theta \in S_{j,n}} \{\mathbb{P} m_\theta - \mathbb{P} m_{\theta_0} - (P m_\theta - P m_{\theta_0}) + (P m_\theta - P m_{\theta_0})\} \geq -\frac{K}{r_n^\alpha} \right) \\
\leq &\sum_{j>M, 2^j \leq \epsilon r_n} P^* \left(\sup_{\theta \in S_{j,n}} \{\mathbb{P} m_\theta - \mathbb{P} m_{\theta_0} - (P m_\theta - P m_{\theta_0})\} + \sup_{\theta \in S_{j,n}} \{(P m_\theta - P m_{\theta_0})\} \geq -\frac{K}{r_n^\alpha} \right) \\
\leq &\sum_{j>M, 2^j \leq \epsilon r_n} P^* \left(\sup_{\theta \in S_{j,n}} \{\mathbb{P} m_\theta - \mathbb{P} m_{\theta_0} - (P m_\theta - P m_{\theta_0})\} - C\frac{2^{(j-1)\alpha}}{r_n^\alpha} \geq -\frac{C 2^{(j-1)\alpha}}{2r_n^\alpha} \right) \\
= &\sum_{j>M, 2^j \leq \epsilon r_n} P^* \left(||\mathbb{G}_n (m_\theta - m_{\theta_0})||_{S_{j,n}} \geq C\sqrt{n} \frac{2^{(j-1)\alpha}}{2r_n^\alpha} \right) \to 0
\end{aligned}$$

が言えた．

定理の条件をお手軽チェックする方法を考える．θ がユークリッド空間上のベクトルとする．写像 $\theta \mapsto P m_\theta$ が θ_0 で 2 階連続微分可能とする．1 階微分は消えるので，テイラー展開すると

$$P(m_\theta - m_{\theta_0}) = \frac{1}{2}(\theta - \theta_0)^T V (\theta - \theta_0) + o(||\theta - \theta_0||^2)$$

よって 1 つ目の条件

$$\sup_{d(\theta, \theta_0) < \delta} P(m_\theta - m_{\theta_0}) \leq -C\delta^\alpha$$

は V が非特異であれば，$\alpha = 2$ で成り立つことが分かる．

この定理の 2 つ目の条件（Maximal indequality）

$$E^* \sup_{d(\theta,\theta_0)<\delta} |\mathbb{G}_n(m_\theta - m_{\theta_0})| \leq C\delta^\beta$$

はチェックするのが難しい．

$\beta = 1$ が’’よくある’’ $\sqrt{n}$ の収束レートを得るための regular cases. $\alpha = 2, \beta = 1/2$ で収束レート $n^{1/3}$ になる普通でないケースもあるらしい．

19 章では $\{m_\theta - m_{\theta_0} : d(\theta, \theta_0) < \delta\}$ という関数のクラスの複雑さ（エントロピー）を測り，それを元に左辺の上界を得る補題や系が紹介される．$\theta \mapsto m_\theta$ に対するリプシッツ条件はエントロピーのシンプルに推定し，多くの場面で応用される． □

Corollary 6.2. ユークリッド空間の開部分集合の各 θ について，$x \mapsto m_\theta(x)$ は可測関数で，θ_0 の近傍の任意の θ_1, θ_2 と $P\dot{m}^2 < \infty$ なる可測関数 $\dot{m}$ について

$$|m_{\theta_1}(x) - m_{\theta_2}(x)| \leq \dot{m}(x)||\theta_1 - \theta_2||$$

が成り立つとする．さらに，写像 $\theta \mapsto Pm_\theta$ は θ_0 で非特異な 2 階導関数により 2 次のテイラー展開可能とする．このとき，$\mathbb{P}_n m_{\hat{\theta}_n} \geq \mathbb{P} m_{\theta_0} - O_P(n^{-1})$ で $\hat{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$ であれば，$\sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1)$ である．

Proof. テイラー展開に関する仮定と $P\dot{m}_{\theta_0} = 0$ から

$$P(m_\theta - m_{\theta_0}) = \frac{1}{2}(\theta - \theta_0)^T V(\theta - \theta_0) + o(||\theta - \theta_0||^2)$$

で V は非特異となり，

$$\sup_{d(\theta,\theta_0)<\delta} P(m_\theta - m_{\theta_0}) \leq -C\delta^2.$$

定理 5.52 の 1 つ目の条件は成り立つ．次に 2 つ目の条件

$$E^* \sup_{d(\theta,\theta_0)<\delta} |\mathbb{G}_n(m_\theta - m_{\theta_0})| \leq C\delta^\beta$$

が，$\beta = 1$ で成り立つことを示す．そのために Corollary 19.35

Corollary 6.3. envelope function F を持つ可測関数のクラス $\mathcal{F}$ に対して

$$E_P^* ||\mathbb{G}_n||_{\mathcal{F}} \lesssim J_{[]}(||F||_{p,2}, \mathcal{F}, L_2(P))$$

が成り立つ．

を，使いたい．まず，この定理に出てくる用語や記号を説明する．

まず，$x \mapsto f(x)$ $(x \in \mathbb{R}^p, f(x) \in \mathbb{R})$ の関数クラス $\mathcal{F}$ があって F がその envelope function であるとは，任意の x, f について

$$|f(x)| \leq F(x) < \infty$$

となることをいう．

また，’$\lesssim$’ は smaller than up to a universal constant, つまり右辺は定数倍以上で，その定数は問題によらず一定であることを意味する．

$J_{[]}$ は $\mathcal{F}$ の複雑さの指標である bracketing entropy $N_{[]}$ から定義される量である．１９章を参照されたい．

$||\mathbb{G}_n||_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathbb{G}_n f|$ である．

Corollary 6.2 に戻る．まず，$\mathcal{F} = \{m_\theta - m_{\theta_0} : ||\theta - \theta_0|| < \delta\}$ とすれば，仮定から envelope function $F = \dot{m}\delta$ をとれる．このとき，Corollary 19.35 を適用すれば，

$$E^* \sup_{||\theta-\theta_0||<\delta} |\mathbb{G}_n(m_\theta - m_{\theta_0})| \lesssim \int_0^{||\dot{m}||_{P,2}\delta} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon$$
$$\leq C \int_0^{||\dot{m}||_{P,2}\delta} \sqrt{\log\left(\frac{\delta}{\epsilon}\right)}.$$

ここで，$\log\left(\frac{\delta}{\epsilon}\right) = u$ と変数変換すれば，

$$\begin{aligned}
\int_0^{||\dot{m}||_{P,2}\delta} \sqrt{\log\left(\frac{\delta}{\epsilon}\right)} &= \int_\infty^{-\log||\dot{m}||_{P,2}} u^{1/2}(-\delta)e^{-u}du \\
&= \delta \int_{||\dot{m}||_{P,2}\delta}^\infty u^{1/2}e^{-u}du \\
&\leq \delta \int_0^\infty u^{1/2}e^{-u}du \\
&= \delta\Gamma(3/2)
\end{aligned}$$

結局 δ の定数倍で押さえられることが分かり，Theorem 5.52 の条件 2 が $\beta = 1$ で成り立つことが分かる．以上により，$\sqrt{n}d(\hat{\theta}_n, \theta_0) = O_P^*(1)$ が示せた． □