Bootstrap / Yuchi Matsuoka

23.0 Notation

$\hat{\theta}$ は観測の従う分布 P のあるパラメータ θ の推定量である．分布 P の推定として分布 $\hat{P}$ を考える．bootstrap 推定量は分布 P を $\hat{P}$ に置き換えたときの"plug-in"推定量のことである．実際には $\hat{P}$ は，経験分布 $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$（emprical bootstrap），時にあるパラメトリックモデル $P_{\hat{\theta}}$ が用いられる（parametric bootstrap）．そのことを示すための，$\hat{P}$ からの観測により計算された推定量は $\hat{\theta}^*$ のようにスターをつける．ただし，この場合 $\hat{\theta}$ と同様の方法で計算されている．

また，$P(\cdot|\hat{P})$ は，$\hat{\theta}^*$ の分布が，オリジナルの観測が与えられた元での $\hat{P}$ からのサンプルで評価されることを示す．つまり，$P(g(\hat{\theta}^*, \hat{\theta})|\hat{P})$ のような場合 $\hat{\theta}$ はランダムでない．

23.2 Consistency

ここから empirical bootstrap について考える．つまり，$\hat{P}_n = \mathbb{P}_n$ とする．Bootstrap 推定量の分布に対する Kolmogorov-Smirnov 距離に基づく consistency は，

$$\sup_x \left| P\left(\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \le x|P\right) - P\left(\frac{\hat{\theta}^*_n - \hat{\theta}_n}{\hat{\sigma}^*_n} \le x|\hat{P}_n\right)\right| \xrightarrow{\mathrm{P}} 0$$

で定義された．

これは，任意の x に対して

$$P\left(\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \le x|P\right) \to F(x), \quad P\left(\frac{\hat{\theta}^*_n - \hat{\theta}_n}{\hat{\sigma}^*_n} \le x|\hat{P}_n\right) \xrightarrow{P} F(x)$$

が成り立つことと同値であった．

これを，F が正規分布のときのより広いクラスの統計量に対して，確証する．我々の方法ではまず標本平均と同等である $\hat{\theta}_n$ に対する consistency を示し，次に consistency がデルタ法の適用化で保たれることを示す．

これらの結果を合わせることで多くの bootstrap 推定量の consistency を示すことができる．例えば，相関係数に対する信頼区間の consistency などである．

Slutsky の補題から，中心化列 $\sqrt{n}(\hat{\theta}_n - \theta)$ の弱収束と $\hat{\sigma}_n/\sqrt{n}$ の確率収束と合わせると，$(\hat{\theta}_n - \theta)/\hat{\sigma}_n$ の弱収束が得られる．bootstrap 統計量に対しても，このアナロジーが言える．このとき，$\hat{\sigma}^*_n/\sqrt{n}$ の確率収束はオリジナルの観測で条件付けて（つまり $\mathbb{P}_n$ で条件付けたもとで）示さなければならない．

$\hat{\sigma}_n/\sqrt{n}$ と $\hat{\sigma}^*_n/\sqrt{n}$ の（条件付き）一致性を示すのは大抵難しくない．それゆえ，スチューデント化されてない統計量にのみ議論を制限することにする．

$\bar{X}_n$ は有限な平均ベクトル μ と共分散行列 Σ に従う n 個の確率ベクトル $X_1, ..., X_n$ の標本平均とする．多変量の中心極限定理により，$\sqrt{n}(\bar{X}_n - \mu)$ は漸近的に正規分布 $N(0, \Sigma)$ となる．これと同じことを $\sqrt{n}(\bar{X}^*_n - \bar{X}_n)$ に対しても示したい．$\bar{X}^*_n$ は $\mathbb{P}_n$ からの n 個の標本の平均である．すなわち，オリジナルの観測 $\{X_1, ..., X_n\}$ からの n 個の復元抽出である．

Theorem 23.4 (Sample mean)．$X_1, X_2, ...$ は平均 μ, 共分散行列 Σ の i.i.d. 確率ベクトルとする．ほとんどすべての列 $X_1, X_2, ...$ に対して $X_1, X_2, ...$ で条件付けるとき，

$$\sqrt{n}(\bar{X}^*_n - \bar{X}_n) \rightsquigarrow N(0, \Sigma)$$

Proof. fix された列, $X_1, X_2, ...$ に対して, 変数 $\bar{X}_n^*$ は経験分布 $\mathbb{P}_n$ からサンプリングされた n 個の観測 $X_1^*, ..., X_n^*$ の平均である．条件付き平均および共分散行列は

$$\begin{aligned} E(X_i^*|\mathbb{P}_n) &= \sum_{i=1}^{n} \frac{1}{n} X_i = \bar{X}_n, \\ E((X_i^* - \bar{X}_n)(X_i^* - \bar{X}_n)^T|\mathbb{P}_n) &= \sum_{i=1}^{n} \frac{1}{n}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \\ &= \overline{X_n X_n^T} - \bar{X}_n \bar{X}_n^T \end{aligned}$$

大数の強法則により，ほとんどすべての列 $X_1, X_2, ...$ に対し，条件付き共分散は Σ に収束する．

$\bar{X}_n^*$ の漸近分布は中心極限定理により得られる．$X_1^*, ..., X_n^*$ は任意の n に対して異なる分布 $\mathbb{P}_n$ からサンプリングされているので，triangular array に対する中心極限定理が必要である．Theorem 2.27 Lindeberg の中心極限定理を使うことができる．

Proposition 2.27 (Lindeberg-Feller central limit theorem)**.** 各 n に対して，$Y_{n,1}, ..., Y_{n,k_n}$ は有限な分散を持つ独立な確率変数で,

$$\sum_{i=1}^{k_n} E||Y_{n,i}||^2 \{||Y_{n,i}|| > \epsilon\} \to 0, \quad \text{every } \epsilon > 0,$$

$$\sum_{i=1}^{k_n} \mathrm{Cov} Y_{n,i} \to \Sigma$$

を満たすとする．このとき，$\sum_{i=1}^{k_n}(Y_{n,i} - EY_{n,i}) \rightsquigarrow N(0, \Sigma)$.

そのためには任意の $\epsilon > 0$ に対して,

$$E||X_i^*||^2 1\{||X_i^*|| > \epsilon\sqrt{n}\} = \frac{1}{n} \sum_{i=1}^{n} ||X_i||^2 1\{||X_i|| > \epsilon\sqrt{n}\} \overset{\text{a.s}}{\to} 0$$

の収束を示す必要がある．左辺は $\epsilon\sqrt{n} \geq M$ のとき，$\frac{1}{n}\sum_{i=1}^{n} ||X_i||^2 1\{||X_i|| > M\}$ より小さくなる．大数の強法則により，ほとんどすべての列 $X_1, X_2, ...$ に対して，これは，$E||X_i||^2 1\{||X_i|| > M\}$ に収束する．十分大きな M に対して，これは任意に小さくなる．結局左辺の上界は任意の $\eta > 0$ より almost sure で小さくなり，ほとんどすべての $X_1, X_2, ...$ で 0 に概収束することが示せた．よって Lindeberg-Feller の中心極限定理が適用でき，題意が示せた． □

デルタ法を復習しておく．

Theorem 3.1 (デルタ法)**.** $\phi : \mathbb{R}^k \mapsto \mathbb{R}^m$ で ϕ は θ で微分可能とする．T_n は ϕ の定義域内で値をとる確率変数であり，$T_n \in \mathbb{R}^k$ である．

もし $r_n(T_n - \theta) \rightsquigarrow T(as\ r_n \to \infty)$ ならば, $r_n(\phi(T_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(T)$ が成り立つ．更に，$r_n(\phi(T_n) - \phi(\theta)) - \phi'_\theta(r_n(T_n - \theta))$ は 0 に確率収束する．

ここで，定理の中の $\phi'_\theta : h \mapsto \phi'_\theta(h)$ は

$$\phi'_\theta(h) = \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(\theta) \cdots \frac{\partial \phi_1}{\partial x_k}(\theta) \\ \vdots \ddots \vdots \\ \frac{\partial \phi_m}{\partial x_1}(\theta) \cdots \frac{\partial \phi_m}{\partial x_k}(\theta) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_k \end{pmatrix}$$

で定義される．

$\hat{\theta}_n$ は統計量とし，ϕ 与えられた微分可能の関数とする．デルタ法より列 $\sqrt{n}(\hat{\theta}_n - \theta)$ が分布収束するならば，列 $\sqrt{n}(\phi(\hat{\theta}_n) - \phi(\theta))$ も分布収束する．

$\phi(\hat{\theta}_n) - \phi(\theta)$ の分布に対する bootstrap 推定量は $\phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n)$ である．bootstrap が $\hat{\theta}_n - \theta$ に対して consistent であれば，$\phi(\hat{\theta}_n) - \phi(\theta)$ の分布の推定に対する bootstrap も consistent である．

Theorem 23.5 (Delta method for bootstrap). $\phi : \mathbb{R}^k - \mathbb{R}^m$ は θ の近傍で定義された可測で連続微分可能な写像とする．$\hat{\theta}_n$ は ϕ の定義域に値を取り，θ に概収束する確率ベクトルとする．

このとき，$\sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow T$ かつ $\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) \rightsquigarrow T$ が条件付きで almost sure で成り立つならば，

$$\begin{aligned} \sqrt{n}(\phi(\hat{\theta}_n) - \phi(\theta)) &\rightsquigarrow \phi'_\theta(T) \\ \sqrt{n}(\phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n)) &\rightsquigarrow \phi'_\theta(T) \end{aligned}$$

が条件付きで almost sure で成り立つ．

Proof. 平均値の定理により，$\hat{\theta}^*_n, \hat{\theta}_n$ が ϕ が連続微分可能となるような θ の近傍に含まれていれば，ある $\tilde{\theta}_n \in [\hat{\theta}^*_n, \hat{\theta}_n]$ が存在して，

$$\phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n) = \phi'_{\tilde{\theta}_n}(\hat{\theta}^*_n - \hat{\theta}_n)$$

と書ける．微分の連続性により，任意の $\eta > 0$ に対して，ある定数 $\delta > 0$ が存在して，任意の $||\theta' - \theta|| \leq \delta$ で，

$$||\phi'_{\theta'}h - \phi'_\theta h|| < \eta||h||, \quad \forall h.$$

n を十分大きく，δ を十分小さくとれば，$\sqrt{n}||\hat{\theta}^*_n - \hat{\theta}_n|| \leq M$ で $||\hat{\theta}_n - \theta|| \leq \delta$ となり，

$$\begin{aligned} R_n &:= ||\sqrt{n}(\phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n)) - \phi'_\theta\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n)|| \\ &= |(\phi'_{\tilde{\theta}_n} - \phi'_\theta)\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n)| \leq \eta M. \end{aligned}$$

$\epsilon > 0$ を固定し，M を十分大きくとる．$\eta M < \epsilon$ となるように十分小さく η をとる．すると

$$P(R_n > \epsilon|\hat{P}_n) \leq P(\sqrt{n}||\hat{\theta}^*_n - \hat{\theta}_n|| > M \text{ or } ||\hat{\theta}_n - \theta|| > \delta|\hat{P}_n)$$

が成り立つ．

$\hat{\theta}_n$ は θ に概収束するので，右辺は，$||T||$ の任意の連続点 M で $P(||T|| \geq M)$ に概収束する．M の選び方でこれは任意に小さくすることができる．以上により左辺は 0 に概収束する．定理は Slutsky の補題を当てはめることで示される．つまり，$\hat{P}_n$ で条件付けたもとで，

$$\begin{aligned} \sqrt{n}(\phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n)) &= \{\sqrt{n}(\phi(\hat{\theta}^*_n) - \phi(\hat{\theta}_n)) - \phi(\hat{\theta}_n)) - \phi'_\theta\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n)\} + \phi'_\theta\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) \\ &\rightsquigarrow \phi'_\theta(T). \end{aligned}$$

以上により示せた． □