7.3. Convergence to a Normal Experiment

Asymptotic Statistics 勉強会 第19回 資料

Minato

2023/12/16

概要

Asymptotic Statistics p.97, 98 (7.10 Theorem の紹介まで) の内容を記載しました。

7.3 Convergence to a Normal Experiment

条件設定とおさらい

$\Theta \subset \mathbb{R}^k$: 開集合

$(\mathcal{X}, \mathcal{A})$: パラメータ $\theta \in \Theta$ で定まる可測空間

$X_1, \dots, X_n$: $(\mathcal{X}, \mathcal{A})$ 上の分布 P_θ から i.i.d. で生成されたサンプル

* 全観測 $(X_1, \dots, X_n)$ は P_θ^n による単一の観測と捉えられる

* 統計モデルはサンプル空間 $(\mathcal{X}^n, \mathcal{A}^n)$ 上の確率測度の族 $\{P_\theta^n : \theta \in \Theta\}$ として記述できる.
本章では統計的実験 (statistical experiment) と呼ぶ.

* $\theta_0 \in \Theta$: 既知のパラメータとして固定.
局所パラメータ (local parameter) h は $h = \sqrt{n}(\theta - \theta_0)$ として定義される.
$\theta = \theta_0 + h/\sqrt{n}$

* P_θ^n は h を用いて $P^n_{\theta_0 + h/\sqrt{n}}$ と書き換えられる.
本章では, 適切な条件下で, 十分大きな n に対し, 実験

$$\left(P^n_{\theta_0 + h/\sqrt{n}} : h \in \mathbb{R}^k \right), \quad \left(\mathcal{N}(h, I_{\theta_0}^{-1}) : h \in \mathbb{R}^k \right)$$

正規実験　　フィッシャー情報行列の逆行列

は類似の統計的性質をもつことを示す.

局所漸近正規性 (local asymptotic normality, 以降 LAN)
の本質は 局所統計実験 (local statistical experiments) が 正規実験
に収束すること.

← 一般論は9章で!

本節: 一般論を迂回し, 局所実験と正規極限実験
の間で成り立つ直接的な関係を導く

極限実験 (limit experiment): 単一の観測 $X \sim \mathcal{N}(h, I_\theta^{-1})$ からなる実験.

この実験の対数尤度比は

$$\log \frac{d\mathcal{N}(h, I_\theta^{-1})}{d\mathcal{N}(0, I_\theta^{-1})} = \underline{h^T I_\theta X - \frac{1}{2} h^T I_\theta h} . \quad \cdots \text{①}$$

Th7.2の対数尤度比 $\frac{dP^n_{\theta+h/\sqrt{n}}}{dP^n_\theta}$ の展開式に類似.

Th7.2. (復習)

$\Theta \subset \mathbb{R}^k$: 開集合, モデル $(P_\theta : \theta \in \Theta)$ は θ 上 二乗平均微分可能.

このとき, $P_\theta \dot{\ell}_\theta = 0$ であり, フィッシャー情報行列 $I_\theta = P_\theta \dot{\ell}_\theta \dot{\ell}_\theta^T$ が存在する.

さらに, $h_n \to h \ (n \to \infty)$ なる任意の収束列 h_n に対し,

$$\underbrace{\log \prod_{i=1}^n \frac{P_{\theta+h_n/\sqrt{n}}}{P_\theta}(X_i)} = \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \dot{\ell}_\theta(X_i) - \frac{1}{2} h^T I_\theta h} + o_{P_\theta}(1) .$$

モデルの対数尤度比　　① の右辺と類似 ⟶ 正規実験への近似も直観的

(∵) ① の証明

k次元正規分布の密度関数 $p(x | h, I_\theta^{-1})$ は

$$p(x | h, I_\theta^{-1}) = (2\pi)^{-\frac{k}{2}} \det(I_\theta^{-1})^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (x-h)^T I_\theta (x-h) \right]$$

$$\therefore \log \frac{p(x | h, I_\theta^{-1})}{p(x | 0, I_\theta^{-1})} = \log \frac{\exp\left[-\frac{1}{2} (x-h)^T I_\theta (x-h) \right]}{\exp\left(-\frac{1}{2} x^T I_\theta x \right)}$$

$$= -\frac{1}{2} (x-h)^T I_\theta (x-h) + \frac{1}{2} x^T I_\theta x$$

$$= -\frac{1}{2} \left(\cancel{x^T I_\theta x} - x^T I_\theta h - h^T I_\theta x + h^T I_\theta h - \cancel{x^T I_\theta x} \right)$$

$$= h^T I_\theta x - \frac{1}{2} h^T I_\theta h .$$

テーマ：統計モデルの正規近似

θを固定し，既知θ，パラメータhによる統計モデル

$$\left(P^n_{\theta+h/\sqrt{n}} : h \in \mathbb{R}^k\right)$$

が，統計モデル$\left(\mathcal{N}(h, I_\theta^{-1}) : h \in \mathbb{R}^k\right)$で近似できることを示す.

本節の近似は本質的に“局所”パラメータhに関するもの.

(This approximation in this section is “local” in nature.)

局所近似のモチベーション：真のパラメータを，通常漸近的に，無制限の精度 (unlimited precision) で知ることができうる.

$\longrightarrow$ 真の値の“小”近傍内のパラメータθにおける分布P_θにより，真値を得ることの難しさが定まる.

真のパラメータをθ_0とすると，θ_0の近くの点θにおいて，分布P_θがP_{θ_0}と大きく異なる分布であると推定が難しい.

前節の設定では“小” (small) は $O(1/\sqrt{n})$ であることが分かっている.

勉強会で頂いたコメント
局所パラメータの定義式 $\theta = \theta_0 + h/\sqrt{n}$ から言えそう.

それらのモデルの関係は統計量の (極限) 分布により記述できる.

$T_n = T_n(X_1, \dots, X_n)$：固定されたユークリッド空間上で値をとる実験$\left(P^n_{\theta+h/\sqrt{n}} : h \in \mathbb{R}^k\right)$上の統計量.

T_nはn, θに依存. θ: 既知. h: パラメータ

仮定：$\{T_n\}$は各局所パラメータhの下で分布収束する.

$$T_n \overset{h}{\rightsquigarrow} L_{\theta, h} , \quad \text{every } h.$$

ここで，

- $\overset{h}{\rightsquigarrow}$ ： パラメータ $\theta + h/\sqrt{n}$ の下での分布収束
- $L_{\theta,h}$ ： 何らかの確率分布

次の定理より，分布 $\{L_{\theta,h} : h \in \mathbb{R}^k\}$ は必ず
正規実験 $\left(\mathcal{N}(h, I_\theta^{-1}) : h \in \mathbb{R}^k\right)$ における統計量 T になる．

→ 任意の分布収束する統計量の列は極限実験において
ある統計量に一致する．

モデルの収束に統計的解釈を与える

補足：質（quality）は（現時点では）数学的に厳密な定義が述べられておらず，
“統計量が推定対象をいかに推定できるか”という文脈で用いられている．

統計量の質の尺度は，異なるパラメータの下での分布により
表現できる．

いかに対象を推測できるか

詳細は後の章で！

例1： 統計的仮説検定

ある仮説が，統計量 T_n が $T_n > c$ なら棄却されるとしたら，
検出力関数 $h \mapsto P_h(T_n > c)$ が関連する．

例2： 二乗誤差（回帰／分類）

T_n が h の推定量とすると，平均二乗誤差
$h \mapsto \mathbb{E}_h[(T_n - h)^2]$ ないしはその類似の量が
T_n の質を決定付ける．

上記のいずれの例でも，質の尺度はその統計量の分布のみに依存する．

Th7.10 は 統計量 T_n の分布は，h の関数として，何らかの統計量 T の分布によりよく近似できることを主張する．

⟶ T への近似の質（精度）は T_n の漸近的な質と等価．

n が大きくなると T_n の分布は T の分布にどう近づくか

その統計量 T がどういったものかを調べることで，T_n の漸近的な精度（asymptotic performance）も明らかになる．

検定や推定への具体的な応用は後の章で扱うとのこと．

Rmk.

- テクニカルな話で，極限実験において randomized statistic を導入する必要がある．

勉強会で頂いたコメント
ランダム統計量を導入して証明するケースは他の定理でもある（決定理論や Karlin-Rubin の定理等）

- 観測 X における ランダム統計量 T：可測写像 $T=T(X,U)$
  ここで，U は X と独立で，一様分布にしたがう確率変数．
  多くの場合非実用的だが，定理の証明は U の手助けを借りて示す．

- それゆえ，極限実験の解析をする際は，
  観測と，パラメータの知識によらず実験できる他のものの結果の両方に基づき検定・推定する．

※ U が一様分布に従うことは本質的ではない．
任意の乱数を用いる手法はこれを満たしている．

Th7.10

- 実験 $(P_\theta : \theta \in \Theta)$：フィッシャー情報行列 I_θ が正則な点 θ 上で二乗平均微分可能．
- T_n：実験 $(P^n_{\theta+h/\sqrt{n}} : h \in \Theta)$ の統計量 s.t. 各 h の下で T_n は分布収束する

このとき，実験 $(\mathcal{N}(h, I_\theta^{-1}) : h \in \mathbb{R}^k)$ におけるある統計量 T が存在し，各 h で $T_n \overset{h}{\rightsquigarrow} T$.