概要

統計学コースの研究計画書は、A4 版の用紙で 10 ページ以内とし、最初の 3 ページに統計コースを志望した理由、動機とともに入学後の研究計画を記載し、残りの枚数で統計学または計量経済学に関するエッセーを記載すること。エッセーは、例えば卒論の内容でも、今興味を持って取り組んでいる課題を小論文として執筆したものでもよい。ただし、数学的な能力の高さを評価するので、エッセーは数式に基づいて論理的に記述された内容である必要がある。なお、使用言語は日本語又は英語とする。

1　研究計画

1.1　志望動機

高校時代は自然科学と社会科学が好きで，受験数学の点数は伸び悩んでも，物理は大学教養過程の教科書も用いて完全な答案を書くことに精を出した．一方で受験とは関係ないながらも，隙間時間で心理学や行動経済学などの分野で学会をリードする学者の書籍を読み，ときには巻末の参考文献を参照して，将来はこのような研究者になりたいと考えていた．特に読みふけたのは Daniel Kahneman "Thinking Fast and Slow"の翻訳書 [1] で，

1.2　研究計画について

2　影響関数を用いたワンステップ推定

滑らかなノンパラメトリックモデルにおける最適な推定量は，ワンステップ推定量として構成できる．その（滑らかな）セミパラメトリックモデルにおける対応物としては，影響関数なる無限次元解析の言葉が自然に現れる．そこで，標準的な構成法を得るには，Malliavin 解析からの知見が生きる可能性が十分にある．

記法 2.1. 以下，断りなく用いる記法をここにまとめる．

(1) $\mathcal{L}^2(\mathcal{X},P;\mathbb{R}^p)$ により，確率空間 $(\mathcal{X},P)$ 上の 2 乗可積分な $\mathbb{R}^p$-値関数の全体のなす線型空間を表す．$\mathcal{X}$ が明らかな場合は省略し，$p=1$ の場合は $\mathbb{R}^p$ も省略する．$L^2(\mathcal{X},P;\mathbb{R}^p)$ で，零集合の差を除いて等しい関数を同一視して得る Hilbert 空間を表し，ノルム $\|-\|$ は特に断りがない限り，この Hilbert 空間において考える．内積を $(-|-)$ で表す．

(2) $L^2_0(P):=\left\{g\in L^2(P)\mid E[g]=0\right\}$ を中心化された確率変数のなす部分空間とする．

(3) $P(\mathcal{X})$ によって，可測空間 $(\mathcal{X},\mathcal{A})$ 上の確率測度全体のなす空間とする．

(4) $f:\mathcal{X}\rightharpoonup\mathbb{R}$ によって，$\mathcal{X}$ の全体で定義されているとは限らない部分関数 $f|_{\mathrm{Dom}f}:\mathrm{Dom}f\to\mathbb{R}$ を表す．

2.1　パラメトリックモデルの効率限界

推定量の有効性を漸近分散の小ささで測ることとしよう．すると，正則性条件を満たす滑らかなパラメトリックモデルについては，Fisher 情報量が下界を与えるのであった (Cramer-Rao の不等式)．後ほどセミパラメトリックモデルの場合と比較するため，Cramer-Rao の不等式が成立するための十分条件をなるべく弱い形で述べる．

定理 2.2 (Cramer-Rao). $U\overset{\text{open}}{\subset}\mathbb{R}^q$ で添字付けられたパラメトリックモデル $(P_\theta)_{\theta\in U}$ と，偏微分可能な関数 $\psi:U\to\mathbb{R}^p$ と，その $\theta\in U$ における不偏推定量 $\widehat{\psi}\in\mathcal{L}^2(\mathcal{X}^n;\mathbb{R}^p)$ とについて，次を仮定する：

(P0) 分布族 $(P_\theta)_{\theta\in U}$ はある σ-有限な参照測度 $\mu\in P(\mathcal{X})$ に関して絶対連続とし，その Radon-Nikodym 微分を $p_\theta:\mathcal{X}\to[0,1]$ で表す．

(P1) $p_-(x):U\to[0,1]$ は P_θ-a.e. x に関して偏微分可能である．

(P2) $p_\theta(x)>0$ を満たす点 (x,θ) の上で，スコア関数の第 i 成分 $g_i:\mathcal{X}\times U\rightharpoonup\mathbb{R}$ を $g_i(x,\theta):=\dfrac{\partial\log p_\theta(x)}{\partial\theta_i}=\dfrac{1}{p_\theta}\dfrac{\partial p_\theta}{\partial\theta_i}$ と定める

と[†1]，二次の絶対積率が有限 $g_i(-,\theta)\in\mathcal{L}^2(P_\theta)$：

$$E_{P_\theta}[|g_i|^2]=\int_{\mathcal{X}}|g_i(x,\theta)|^2p_\theta(x)\mu(dx)<\infty\qquad(\forall_{i\in[q]}).$$

(P3) 各 $\theta\in U$ における Fisher 情報行列 $I=(I_{ij})$ を，第 (i,j)-成分を

$$I_{ij}(\theta):=\mathrm{Cov}_{P_\theta}[g_i,g_j]=\int_{\mathcal{X}}\frac{\partial\log p_\theta(x)}{\partial\theta_i}\frac{\partial\log p_\theta(x)}{\partial\theta_j}\cdot p_\theta(x)\mu(dx)\qquad(\forall_{i,j\in[q]}).$$

とすることによって定めると，正定値な対称行列となる．

(P4) 不偏推定量 $\widehat{\psi}$ は $g:=(g_1,\cdots,g_q)^\top$ が定める推定方程式の解であり：$E_{P_\theta}[g]=\displaystyle\int_{\mathcal{X}}g(x,\theta)p_\theta(x)\mu(dx)=0\in\mathbb{R}^q$，かつ，次の微分と積分の可換性が成り立つ：

$$\int_{\mathcal{X}}\delta(x)g_i(x,\theta)p_\theta(x)\mu(dx)\left(=\int_{\mathcal{X}}\delta(x)\frac{\partial p_\theta}{\partial\theta_i}\mu(dx)\right)=\frac{\partial}{\partial\theta_i}\int_{\mathcal{X}}\delta(x)p_\theta(x)\mu(dx)\qquad(i\in[q]).$$

このとき，関数 $\psi:\mathbb{R}^q\to\mathbb{R}^p$ の Jacobi 行列を $J(\theta):=\dfrac{\partial\psi}{\partial\theta}\in M_{p,q}(\mathbb{R})$ とすると，次の行列不等式が成り立つ：

$$\mathrm{Var}_\theta[\widehat{\psi}]\geqslant J(\theta)I(\theta)^{-1}J(\theta)^\top.$$

[証明]．任意に $\theta\in U$ を取り，$I:=I(\theta),J:=J(\theta)$ と略記する．

共分散への翻訳 仮定 (P4) より，

$$J=\frac{\partial\widehat{\psi}(\theta)}{\partial\theta}=\frac{\partial}{\partial\theta}E_\theta[\delta]=E_\theta[\delta g^\top].$$

これと $E_\theta[g]=0$ より，

$$\mathrm{Cov}_\theta[g,\delta]=E_\theta[g\delta^\top]=J^\top.$$

$I=E_\theta[gg^\top]=\mathrm{Var}_\theta[g]$.

証明 すると，Cauchy-Schwartz の不等式の証明と同様に，$\mathrm{Var}_\theta[\delta-JI^{-1}g]\geqslant O$ であることから，Cov の双線型性のみから従う．また，対称行列 S に対して，$S^{-1}=(S^{-1})^\top=(S^\top)^{-1}$ であることと Fisher 情報行列が対称であることに注意すると，

$$\begin{aligned}O&\leqslant\mathrm{Var}_\theta[\delta-JI^{-1}g]=\mathrm{Cov}[\delta-JI^{-1}g,\delta-JI^{-1}g]\\&=\mathrm{Var}_\theta[\delta]-JI^{-1}\mathrm{Cov}_\theta[g,\delta]-\mathrm{Cov}_\theta[\delta,g]I^{-1}J^\top+JI^{-1}\mathrm{Var}[g]I^{-1}g^\top\\&=\mathrm{Var}_\theta[\delta]-JI^{-1}J^\top.\end{aligned}$$

■

系 2.3. 特に $q=p=1$ の場合，$\mathrm{Var}_\theta[\widehat{\psi}]\geqslant\dfrac{\left(\frac{\partial\psi(\theta)}{\partial\theta}\right)^2}{\mathrm{Var}_\theta[g]}$.

2.2 セミパラメトリックモデル

パラメトリックモデル上の汎関数 $\mathcal{P}\to\mathbb{R}$ は，（識別可能性条件などに基づく）$\mathcal{P}\simeq U$ の同一視によって，Euclid 空間上の関数とみなせたため，正則性条件の記述も議論も初等的であった．しかしセミパラメトリックモデルでは $\mathcal{P}$ は局所的に無限次元線型空間となる，Frechet 多様体というべき対象である．そこで，正則性条件の記述自体も数学的な準備が必要となり，適切な正則性条件を見つけることも困難となる．そのうち一つの自然な概念には，局所漸近正規性の成立を一つの規準とした「L^2-微分可能性」がある．

モデル $\mathcal{P}\subset P(\mathcal{X})$ 上の点 $P\in\mathcal{P}$ を真の分布として，$\psi(P)$ の値を推定する際の効率限界を得るためには，モデル $\mathcal{P}$ も有限次元に制限すればパラメトリックモデルであるから，そのような「部分パラメトリックモデル」のそれぞれについて Cramer-Rao の下界を求め，その下限を求める，という方針が考え得る（最良の下界を与えるかは議論の余地が残る）．ここで，１次元の部分パラメトリックモデルとは，幾何学的には道 $[0,1]\to\mathcal{P}$ を取ることに他ならないが，すべての連続写像 $[0,1]\to\mathcal{P}$ が定める部分パラメ

[†1] $\{x\in\mathcal{X}\mid p_\theta(x)=0\}$ 上で $g_i(-,\theta)$ は定義されていないことに注意．

トリックモデルが正則性条件を満たすわけではないから，適切なクラスを定める必要がある．これが次の，正則なパラメトリックモデルと付随するスコア関数の概念である．

なお，引き続き，(P0) と同様の仮定：

(S0) ある σ-有限な参照測度 $\mu \in P(\mathcal{X})$ に関してモデル $\mathcal{P}$ の元はすべて絶対連続とし，その Radon-Nikodym 微分を対応する小文字 $p : \mathcal{X} \to [0,1]$ で表す．

をおく．

定義 2.4 (regular parametric model / regular path, tangent set). モデル $\mathcal{P} \subset P(\mathcal{X})$ と分布 $P \in \mathcal{P}$ に対して，

(1) **正則な部分パラメトリックモデル**または**スコア関数 g に関して L^2-微分可能な P-道**とは，$P_0 = P$ を満たす写像 $(P_t) : [0,1] \to \mathcal{P}$ であって，ある 2 乗可積分な関数 $g : \mathcal{X} \to \mathbb{R}$ が存在して，

$$\left\| \frac{\sqrt{p_t} - \sqrt{p}}{t} - \frac{1}{2} g\sqrt{p} \right\|^2 = \int_{\mathcal{X}} \left(\frac{\sqrt{p_t} - \sqrt{p}}{t} - \frac{1}{2} g\sqrt{p} \right)^2 d\mu \xrightarrow{t \to 0} 0 \tag{1}$$

ただし，ノルムは Hilbert 空間 $L^2(\mathcal{X}, \mu)$ 上で考えた．以降，この条件を満たす写像 $[0,1] \to \mathcal{P}$ を簡単に**正則な P-道**といい，その全体を $\mathrm{Sub}_P(\mathcal{P})$ で表す．

(2) 次の集合をモデル $\mathcal{P}$ の P における**接集合**または**接正錐**という：

$$\dot{\mathcal{P}}_P := \left\{ g \in \mathcal{L}^2(\mathcal{X}, \mu) \;\middle|\; \exists_{(P_t) \in \mathrm{Sub}_P(\mathcal{P})}\ g \text{ は部分モデル } (P_t) \text{ のスコア関数である} \right\}.$$

注 2.5. 式 (1) は，関数 $q(t,x) := \sqrt{p_t(x)}$ が，$L_2(\mu)$ の意味で $t=0$ において微分係数 $q'(0,x)$ を持ち，これを用いて

$$g(x) := 2\frac{q'(0,x)}{q(0,x)} = 2 \left.\frac{\partial \log q(t,x)}{\partial t}\right|_{t=0}$$

と定めることに等しい．これを $p_t(x)$ で書き直すと，$q'(t,x) = \dfrac{\partial q(t,x)}{\partial t} = \dfrac{\partial \sqrt{p_t(x)}}{\partial t} = \dfrac{1}{2} \dfrac{p_t'(x)}{\sqrt{p_t(x)}}$ より，

$$g(x) = \frac{p_0'(x)}{p_0(x)} = \frac{p'(x)}{p(x)} = \frac{\partial}{\partial t} \log p_t(x)|_{t=0}.$$

したがってこれは，存在すれば，部分モデル $(P_t) \in \mathrm{Sub}_P(\mathcal{P})$ のスコアに他ならない．

続いて，汎関数 $\psi : \mathcal{P} \to \mathbb{R}$ が満たすべき正則性条件について考える．パラメトリックな場合では，関数 $\psi : U \to \mathbb{R}$ は単に偏微分可能としたのであった．

定義 2.6 (pathwise differentiable, von Mises differentiable). 汎関数 $\psi : \mathcal{P} \to \mathbb{R}$ について，

(1) 接正錐 $\dot{\mathcal{P}}_P$ に関して**道ごとに微分可能**とは，任意のスコア関数 $g \in \dot{\mathcal{P}}_P$ と対応する部分モデル $(P_t) \in \mathrm{Sub}_P(\mathcal{P})$ に対して，連続線型汎関数 $\dot{\psi}_P : L_2(P) \supset \dot{\mathcal{P}}_P \to \mathbb{R}$ が存在して，

$$\frac{\psi(P_t) - \psi(P)}{t} \xrightarrow{t \to 0} \dot{\psi}_P(g) \;\in \mathbb{R}$$

が成り立つことをいう．汎関数 $\dot{\psi}_P : \dot{\mathcal{P}}_P \to \mathbb{R}$ を ψ の P における**微分**と呼ぶ．

(2) 連続線型汎関数 $\dot{\psi}_P : \dot{\mathcal{P}}_P \to \mathbb{R}$ に対して，

$$\left\{ \widetilde{\psi} \in L^2(\mu) \;\middle|\; \dot{\psi}_P(g) = (\widetilde{\psi}_P | g) = \int \widetilde{\psi}_P g \, dP \right\} \neq \emptyset$$

が成り立つ．この集合の点で，$\dot{\mathcal{P}}_P$ が生成する閉部分空間との距離が最も近い元 $\widetilde{\psi}_P$ を**有効影響関数**という．

(3) **von Mises 可微分**であるとは，任意の $P' \in \mathcal{P}$ に対して，対応する $\widetilde{\psi}(x;P') \in \mathcal{L}_0^2(P), R_2(P,P') \in \mathbb{R}$ が存在して，

$$\psi(P') - \psi(P) = \int_{\mathcal{X}} \widetilde{\psi}(x;P')(P'-P)(dx) + R_2(P,P')$$

が成り立つことをいう．[†2]

[†2] この剰余項 R_2 に高次の影響関数とその線型作用素による像，いわば $\int_{\mathcal{X}} \widetilde{\psi}'(P'-P)^2(dx)$ とも書くべき項が入っており，その解析的な性質についての理論はありえるだろう．

2.3 セミパラメトリックモデルの効率限界

定理 2.7 (7.2 [4]). (S1) $\theta \mapsto P_{\theta,\eta}$ は L^2-微分可能である.
(S2) $P_{\theta_n,\eta}[\|\widehat{l}_{n,\theta_n} - \widetilde{l}_{\theta_n,\eta}\|^2] \xrightarrow{P} 0$.
(S3) 有効情報行列 $\widetilde{I}_{\theta,\eta}$ は正則になる.
(S4) $\sqrt{n}$-一致性 $\sqrt{n}P_{\theta_n,\eta}[\widehat{l}_{n,\theta_n}] \xrightarrow{P} 0$ が成り立つ.
このとき，$(\widehat{\theta}_n)$ は (θ,η) において漸近有効である.

3 Malliavin 解析による汎関数推定量の漸近展開

4 頑健統計への応用

近似精度の改良へ至る前に可能な応用として，頑健な推定量の構成があり得る．特に，二重に頑健な推定量は，セミパラメトリックモデル特有の性質である．

5 正規近似への応用

参考文献

[1] ダニエル・カーネマン『ファスト＆スロー』（村井章子著・訳，早川書房，2012 年）.
[2] 吉田朋広『数理統計学』(朝倉書店，2006)
[3] van der Vaart - Asymptotic Statistics
[4] Bolthausen - Lectures on Probability Theory And Statistics
[5] Edward Kennedy - Semiparametric Doubly Robust Targeted Double Machine Learning