研究計画書　要綱

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。

志望動機

筆者は東京大学前期教養学部理科一類にて学んだあと，理学部数学科に進学したが，これは元々数学科のガイダンスにて「数学を軸足に置いた応用」という言葉に出会ったためであった．筆者は物理学，経済学，社会学，心理学など，実験・観察とデータ分析を通じて知見を得る科学が，数理科学・社会科学・人文科学の別に拘らず満遍なく大好きであったが，教養過程を過ごした後も，何か特定の分野で特定の対象を研究したいという興味が醸成されたわけではなかった．

しかし，統計的手法を通じて我々はどのような知見を引き出し得るかの限界には興味があったため，「研究対象」ではなく「研究手法」の方からアプローチすることを目指し，統計学を学ぶことを考えた．その際に経済学部ではなく数学科を選んだのは，先述のように「数学を軸足に置いた応用」で初めて可能になる視点も多いのではないかと考えたためである．その例として現時点での筆者が持っているものは，セミパラメトリックモデルという，初めは生存時間解析の手法において Cox が 1972 年に開発した比例ハザードモデルなどから始まり，続いて疫学・生物統計学の分野にて生じた統計的因果推論の問題など，さらには経済学・天文学の応用の現場で自然に生じるモデルであるが，そのパラメータ空間が無限次元になることにより必然的に数学的な困難さが付きまとう統計手法に対して，有用な共通言語を提供できる可能性は，「数学を軸足に置いた応用」の良い例になるのではないかと考える．その内容については，エッセイにて詳述した．

研究計画

古典的に使われていたモデルはパラメトリックモデルであった．その最初の問題点としては，「推論前に置いた仮定が正しくなかった場合」への脆弱性である．そこで 1960 年代に，そもそも推論以前の仮定をなるべく減らす方法として「分布に仮定を置かなくても使えるモデル」であるノンパラメトリックモデルが志向されたのと同様に，仮定が間違っていたとしても，そのことによって推論結果が大きな影響を受けにくいモデルが探求された（頑健統計）．さらに，1970 年代にはパラメトリックモデルの表現力と両立させるために，モデルに部分的にパラメトリックな仮定を含ませたセミパラメトリックモデルが考えられた．

エッセイ：統計モデルの標準理論

1 導入

このエッセイでは,「統計モデル」と「セミパラメトリックモデル」とをほぼ同義語として用いる.セミパラメトリックモデルとは,統計的実験 $(\mathcal{X}, \mathcal{A}, \{P_{\theta,\eta}\}_{(\theta,\eta)\in\Theta\times H})$ であって,確率分布の族 $\{P_{\theta,\eta}\}_{(\theta,\eta)\in\Theta\times H}$ のパラメータ空間 $\Theta\times H$ が,有限次元 Euclid 空間 Θ と無限次元 Hilbert 空間 H との直積で表せるものをいう.$H = 0$ のときがパラメトリックモデル,$\Theta = 0$ のときがノンパラメトリックモデルにあたると考えれば,史上登場した統計モデルの全てを含んだ最も一般的な枠組みともみなせるために,筆者はそう考える.

筆者は,この最も一般的な枠組み $(\mathcal{X}, \mathcal{A}, \{P_{\theta,\eta}\}_{(\theta,\eta)\in\Theta\times H})$ の下で,

(1) パラメトリックモデルについて知られてる標準理論がどの程度一般的なセミパラメトリックモデルについても拡張できるか.特に,どのようにして有効な一致推定量を構成すればよいか.
(2) 有効な一致推定量が標準的に構成できるための必要条件・十分条件は何か.
(3) ロバストな推定量をどう考えればよいか.

の 3 つがどのように考えられるかに関心がある.(3) などの実用的な応用の問題ももちろんだが,(1), (2) の問題は統計モデルを統一的に扱うための数学的枠組み(確率分布の族の「可微分 Fréchet 多様体」としての構造,その上の関数の微分可能性の定義と扱い,推定量の級数展開可能性など)自体へのより深い理解と深く関連しており,筆者が数学科学部時代に注力した関数解析や確率解析(特に無限次元空間上での微積分)の方法を用いて,無限次元空間やその上での微積分という非常に扱いが難しい対象に対して,見通しの良い概念と枠組みが提供できないか,ということにも関心がある.もしそのような枠組みが提供できれば,数学者を初めとした理論研究者の積極的な参入や,実践・応用の議論においても見通しの良い共通言語を提供できるのではないかと考えている.

このエッセイでは,第 2〜5 節にて

第 2 節 まず,影響関数を用いた有効な一致推定量の標準的構成法をみて,パラメトリックモデルの古典理論との比較を通じて,共通点と相違点,特に無限次元空間の出現によってどのような点に困難が生じるかを考える.

第 3 節 次に,その手法で最も重要な概念である影響関数の数学的性質をよく考え,今後どのような発展が考え得るかを議論する.特に,セミパラメトリックモデルでは解析的な一般論はその難しさ故に今まで見られなかったため,標本の数が十分大きくない場合はシミュレーションなどによってパフォーマンスが測られたが,漸近展開の手法がこの点についても強力な一般論を提供できる可能性,または漸近論的手法が適用可能であるための必要十分条件を考える.

第 4 節 さらに,応用の可能性を 2 つ議論する.一つは頑健統計への応用である.セミパラメトリックモデルはパラメトリックな部分もあるので,ここに相変わらずパラメトリックな仮定が現実と違ってしまうこと (misspecification) に対する脆弱性があるが,これがどのように解消され得るかを考える.

第 5 節 2 つ目に,必要な正則性条件を満たさないために漸近理論が適用できない場合などに於ては,他の方法で分布を近似することを考えることとなる.このような,漸近理論と両輪となるような,分布近似法による理論展開の可能性を議論する.

記法 1.1. 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 影響関数を用いたワンステップ推定に関するレビュー

滑らかな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．この議論における対応物が，影響関数を用いてセミパラメトリックモデルにおいても展開できる．この観点は [4] を参考にした．
パラメトリックモデルに於て，漸近論が展開できるための正則性条件はいくつかが知られている一方，セミパラメトリックモデルに於てはそれらが明確でない場合が多い．
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．

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 2.1 (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 $g_i:\mathcal{X}\times U\twoheadrightarrow\mathbb{R}$ 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，
$$\mathrm{Cov}_\theta[g,\delta]=E_\theta[g\delta^\top]=J^\top.$$
$$I=E_\theta[gg^\top]=\mathrm{Var}_\theta[g].$$

[†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．

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 2.2. 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　統計的汎関数の微分可能性

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，その上限を求める，という方針が考え得る（「全ての部分パラメトリックモデル」を考慮に入れていない限り，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 2.3 (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 2.4 (定義式の意味). 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 2.5 (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，

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

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 $\widetilde{\psi}(x; P') \in \mathcal{L}^2_0(P), R_2(P, P') \in \mathbb{R}$ 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]

2.3　セミパラメトリックモデルの効率限界

定理 2.6 (ノンパラメトリック一致推定量の最小分散)**.** 任意のノンパラメトリックモデル $\mathcal{P} \subset P(\mathcal{X})$ と，有効影響関数を $\widetilde{\psi}(x; P) = \widetilde{\psi}_P(x)$ として von Mises 可微分かつ $\dot{\mathcal{P}}_P$ に関して道毎に可微分な汎関数 $\psi : \mathcal{P} \to \mathbb{R}$ と，その $P \in \mathcal{P}$ における不偏推定量 $\widehat{\psi} \in \mathcal{L}^2(\mathcal{X}^n; \mathbb{R})$ について，(S0) と (N1)〜(N4) を仮定する．

(N1) 積分と極限の可換性：任意の $(P_t) \in \mathrm{Sub}_P(\mathcal{P})$ に関して，$(\widetilde{\psi}_P(p_t - p))$ は非負値の可積分な優関数を持つ．

(N2) von Mises 展開の正当性：任意の $(P_t) \in \mathrm{Sub}_P(\mathcal{P})$ に関して，von Mises 展開の 2 次の剰余項は $R_2(P_t, P) = o(t)$.

(N3) 接正錐が十分に大きい：$\widetilde{\psi}_P \in \dot{\mathcal{P}}_P$.

このとき，$\widehat{\psi}$ の分散は $\mathrm{Var}_P[\widehat{\psi}] \geqslant \mathrm{Var}_P[\widetilde{\psi}_P^2]$ を満たす．

［証明］．

部分パラメトリックモデル　中心化された確率変数 $h \in \mathcal{L}_0(P)$ について，$p_\epsilon := p(1 + \epsilon h)$ により定まる P-道 $(p_\epsilon)_{\epsilon \in [0,1]}$ は，スコア関数 h を持つ．von Mises 可微分性より，任意の $\epsilon \in [0, 1]$ に関して

$$\psi(P_\epsilon) - \psi(P) = \int_{\mathcal{X}} \widetilde{\psi}_P(P_\epsilon - P)(dx) + R_2(P_\epsilon, P)$$

が成り立つ．両辺を $\epsilon \neq 0$ で割って，$\epsilon \to 0$ の極限を考えることと仮定 (S1),(S2) より，$\dot{\psi}_P = \int_{\mathcal{X}} \widetilde{\psi}_P h P(dx)$. よって Cramer-Rao の定理 2.2 より，$\mathrm{Var}_P[\widetilde{\psi}] \geqslant \dfrac{E_P[\widetilde{\psi}_P h]^2}{E_P[h^2]}$. 右辺は，任意の $h \in \mathcal{L}_0(P)$ に関して，Cauchy-Schwartz の不等式から $E_P[\widetilde{\psi}_P^2]$ により抑えられる．また，$h = \varphi \in \mathcal{L}_0(P)$ と取った時の P-道が $E_P[\widetilde{\psi}_P^2]$ を与えることに注意すれば，あとは分散 $E_P[\widetilde{\psi}_P^2]$ を持つ一致推定量 $\widehat{\psi}$ を構成すれば良い．これは，定理 2.10 による．

■

[†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．

定理 2.7 (7.2 [3]).

(S1) $\theta \mapsto P_{\theta,\eta}$ は L^2-微分可能である．
(S2) $P_{\theta_n,\eta}[\|\widehat{l}_{n,\theta_n} - \widehat{l}_{\theta_n,\eta}\|^2] \xrightarrow{P} 0$.
(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．

2.4 パラメトリックモデルにおけるワンステップ推定量

記法 2.8. $\Theta \subset \mathbb{R}^p$ を可測集合，$\psi : \mathcal{X} \times \Theta \to \mathbb{R}^p$ を推定関数とし，$\psi(x,-)$ は連続とする．$\psi_n(\theta) := \sum_{j=1}^n \psi(x_j,\theta)$，$\overline{\psi}_n(\theta) := \mathbb{E}_n[\psi(\theta)] = \dfrac{1}{n}\psi_n(\theta)$ と定め，真の平均を $\overline{\psi}(\theta) := \int_{\mathcal{X}} \psi(x,\theta)P(dx)$ と定める．

Θ-値確率変数の列 $(\widehat{\theta}^0_n)$ に対して，新しい推定量の列を

$$\widehat{\theta}_n := \widehat{\theta}^0_n - [\partial_\theta \psi_n(\widehat{\theta}^0_n)]^{-1}\psi_n(\widehat{\theta}^0_n)$$

によって定めると，右辺は

$$\mathcal{X}^*_n := \left\{x \in \mathcal{X}^n \mid \widehat{\theta}^0_n\text{において}\psi_n\text{は微分可能で } \partial_\theta\psi_n(\widehat{\theta}^0_n) \text{ は正則で}\widehat{\theta}_n \in \Theta\right\}$$

上で定まるから，これを $\mathcal{X}$ 上に可測に延長する．延長の仕方は任意で良い．こうして得た $(\widehat{\theta}_n)$ を $(\widehat{\theta}^0_n)$ を初期推定量とするワンステップ推定量という．

定理 2.9. $\theta_0 \in \Theta^\circ$ とし，$\theta_0 \in B \subset \Theta$ を開近傍とする．次の条件を仮定する．

(E1) 推定関数 $\psi : \mathcal{X} \times \Theta \to \mathbb{R}^p$ は B 上 C^1 級．
(E2) $\psi(-,\theta)$ は可測．
(E3) $\psi(x,\theta_0) \in \mathcal{L}^2(P;\mathbb{R}^p)$ で，$P\psi(x,\theta_0) = 0$.
(E4) $\exists_{M\in\mathcal{L}^1(\mathcal{X},\mathbb{R})}\ \forall_{x\in\mathcal{X},\theta\in B}\ |\partial_\theta\psi(x,\theta)| \leqslant M(x)$．さらに，$\Gamma := -P\partial_\theta\psi(x,\theta_0)$ は正則．
(E5) 確率変数列 $\widetilde{\theta}^0_n : \mathcal{X}^n \to \Theta$ が $\sqrt{n}$-一致性を持つ：$\widehat{\theta}^0_n - \theta_0 = O_p(1/\sqrt{n})$.[†3]

このとき，ワンステップ推定量 $\widehat{\theta}_n$ に対して，

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = -\Gamma^{-1}\frac{1}{\sqrt{n}}\psi_n(\theta_0) + o_P(1).$$

特に，$\Sigma := \Gamma^{-1}\Phi(\Gamma^\top)^{-1}$，$\Phi := \int_{\mathcal{X}} \phi\phi^\top(x,\theta_0)P(dx)$ について，

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N_p(0,\Sigma).$$

が成り立つ．

2.5 影響関数を用いたワンステップ推定量

定理 2.10. $\widehat{\mathbb{P}}$ を $\mathbb{P}$ の推定量とする．$\widehat{\psi} := \psi(\widehat{\mathbb{P}}) + \mathbb{P}_n[\widetilde{\psi}_{\widehat{\mathbb{P}}}]$ と定め，(S0),(N1)∼(N4) を仮定する．

(N4) $R_2(\widehat{\mathbb{P}},P) \xrightarrow{n\to\infty} 0$.

[†3] 条件 $\widehat{\theta}^0_n - \theta_0 = O_p(1/\sqrt{n})$ は，弱一致性 $\widehat{\theta}^0_n - \theta_0 = o_p(1)$ を含意することに注意．これに加えて，$n^{-1/2}$ 倍の範囲で真値 θ_0 を捕まえていることは要求する．

3 影響関数の概念について

影響関数の概念は Hampel の頑健統計に関する博士論文 (1968) にて登場するが，前節のように，セミパラメトリックモデルにおける漸近有効性の文脈でも自然に登場した．これはさらに元を辿れば，von Mises の 1947 年の論文で"Volterra 微分"の概念の精緻化でもある．このように，独立の文脈で着想された無限次元微分の概念を，

(1) 微分可能性を特徴付けられないか．
(2) 微分可能であっても，Taylor 展開可能であるのはどのような場合か．
(3) 特に，各項（特に 2 次以上の高次の剰余項）の収束性について，一般的に成り立つ命題はないか．

などの観点から整理することが考え得る．

歴史 3.1. von Mises (1947) では，Volterra (1913) に於て，無限個の変数を持つ関数を「線の関数 (fonction de ligne)」とみなし，その連続性や微分可能性，さらには Taylor 級数展開を考えたことに着想を得て，統計的汎関数にも同様の微分と Taylor 級数展開が考え得ることを議論している．

条件 (N4) はいつ成り立つか？

3.1 von Mises 展開

3.2 漸近展開の可能性？

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] 吉田朋広『数理統計学』(朝倉書店，2006)
[2] van der Vaart - Asymptotic Statistics
[3] Bolthausen - Lectures on Probability Theory And Statistics
[4] Edward Kennedy - Semiparametric Doubly Robust Targeted Double Machine Learning