

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、計量経済学、社会学、社会心理学など、実験・観察とデータ分析を通じて知見を得る「科学」という営みが、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、教養課程を過ごした後も、「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。

興味のある研究対象が見つからなかったと言っても、統計という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「数学に軸足を置いた応用」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。「数学に軸足を置いた応用」の例として、筆者が貢献したいと考えるのは、セミパラメトリックモデルという数理手法である。セミパラメトリックモデルは、パラメトリックモデルの表現力とノンパラメトリックモデルの事前仮定の少なさとの両方の美点を同時に採ろうとすると自然に辿り着く統計モデルのクラスで、初めは生存時間解析の分野において Cox が 1972 年に開発した比例ハザードモデルからその歴史が始まる。続いて計量経済学や疫学・生物統計学をはじめとして、多くの分野にて同時多発的に生じた統計的因果推論への熱も相まって、さらには天文学など、幅広い科学分野で自然に生じるモデルである ([7],[8])。しかし、ノンパラメトリックモデル同様、そのパラメータ空間が無限次元成分を持つことより、必然的に数学的な困難さが付きまとうこととなる。そこで、数学的な内容を整理し、パラメトリックモデルと同様に、直感が効き、認知が容易な枠組みに整理し直し、有用な共通言語を提供できる可能性は、「数学に軸足を置いた応用」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。

また、「数学に軸足を置いた応用」を目指すために経済学部統計コースを選んだ理由は、次の事柄による。

- (1) まず、真に豊かな数学理論は、真に豊かな応用先と共にあってきた。そして経済学にはそのような応用先にあふれている。マクロ・ミクロの別に拘わらず、経世済民の営みには多くのアクターがおり、これの観察・分析・理論化には統計の手法が欠かせない。それも、考慮に入れるアクターによって種々のモデルが要請され、またデータの性質も様々である。例えば、推測統計学の父といえる 20 世紀英国の統計学者 Ronald Fisher は、Rothamsted 農業研究所におけるデータ分析と密接な関わりを持ち、さらに大局的に Karl Pearson, Jerzy Neyman らを含め、英国の統計学は Charles Darwin の影響を受けた優生学と Francis Galton の優生学と統計学の教授職とを中心に発展した [2]。ほとんどすべての統計モデルがパラメトリックであった時代に Marschak's Maxim, すなわち真に重要な政策問題から推定すべきパラメータを逆算し初めから多くの仮定をモデルに置かない考え方を提唱し、後述のエッセイの主題でもある「セミ／ノンパラメトリックモデル上の汎関数の推定」という枠組みを押しすすめることに一役買った Jacob Marschak は Cowles 財団に所属する計量経済学者であった [27]。そして Neyman の潜在結果 (potential outcome) の概念を一般化し、Pearson 以来放逐された因果推論の言葉を統計学に復活させ、Marschak's Maxim を初めて実装に移した Donald Rubin はもともとはじめは心理学博士であった。
- (2) 次に、

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	2
2	パラメトリックモデルに於ける有効推定量の標準的構成法	3
2.1	パラメトリックモデルの効率限界	3
2.2	パラメトリックモデルにおけるワンステップ推定量	4
3	セミパラメトリックモデルに於ける有効推定量の標準的構成法	5
3.1	統計的汎関数の微分可能性	5
3.2	セミパラメトリックモデルの効率限界	6
3.3	影響関数を用いたワンステップ推定量	7
3.4	汎関数の微分可能性の精査と今後の発展	8
4	頑健統計への応用	9

1 導入

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 $\mathcal{P} := \{P_{\theta, \eta}\}_{(\theta, \eta) \in \Theta \times H}$ 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 Banach 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、単に「統計モデル」と言って \mathcal{P} を指し、「統計的実験」と言って $(\mathcal{X}, \mathcal{A}, \{P_{\theta, \eta}\}_{(\theta, \eta) \in \Theta \times H})$ を指すことも許すこととする。

筆者は、この統計的実験 $(\mathcal{X}, \mathcal{A}, \{P_{\theta, \eta}\}_{(\theta, \eta) \in \Theta \times H})$ 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セミパラメトリックモデルについても拡張できるか。 LeCum のワンステップ推定量の概念のセミパラメトリックな対応物として、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) セミパラメトリックモデルにおいて、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) セミパラメトリックモデルにおける「ロバストな推定量」をどう考えればよいか。他のモデルを用いることに比べてどのような利点があるか。

の3つに関心がある。(3)などの実用的な応用の問題もちろんだが、(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「無限次元の多様体」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、したがって筆者が数学科学部生時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解析学や Malliavin 解析）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することで、数学の分野から特有の貢献が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、さらには特定の手法が使えるかどうか、いつ使えるかの明晰な規準を提供することで種々の統計手法の普及に貢献できるのではないかと考えている。

このエッセイでは、第2～4節にて

第2節 まず、パラメトリックモデルにおいて、漸近有効な一致推定量を構成する標準的方法である LeCum のワンステップ推定量の理論と、この手法がいつ使えるのかに関する明瞭な規準（定理 2.4 における仮定 (E1)～(E5) など）をまとめる。

第3節 第2節で見た古典論を、セミパラメトリックモデルに一般化することを考える。そのためには、定理 2.1 に出現する Jacobi 行列の対応物として、影響関数という概念が自然に出現し、これが無限次元空間上の汎関数の微分に当たることを見る。このように、セミパラメトリックな一般化には、統計モデルが無限次元成分を含むことに起因する数学的な困難があり、これが既存の数学理論の応用によってどのように解決され得るかを最後の 3.4 節で議論する。その萌芽は初期の論文、例えば von Mises (1947, [14]) から見られ、近頃志向されている (Ichimura and Newey 2022, [16] など)。

第4節 上述の理論的な明瞭さを与えるという数理的な野心を持った試みは、実際の応用においても有用な手段を提供し得る例として、頑健統計への応用を考える。セミパラメトリックモデルはパラメトリックな部分とノンパラメトリックな部分との2つの構造を併せ持つ。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。一見二箇所に脆弱性が増えたために、セミパラメトリックモデルを採用する利点はないように思えるが、この問題に対して、二重頑健性と呼ばれる性質を持つ推定量を構成することができる [17]。

記法 1.1. 以下、断りなく用いる記法をここにまとめる。

- (1) $\mathcal{L}^2(\mathcal{G}, P; \mathbb{R}^p)$ により、確率空間 (\mathcal{G}, 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{G} 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。 $L^2(\mathcal{G}, P; \mathbb{R}^p)$ 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) $L_0^2(P) := \{g \in L^2(P) \mid E[g] = 0\}$ を中心化された確率変数のなす部分空間とする。
- (3) $P(\mathcal{G})$ によって、可測空間 $(\mathcal{G}, \mathcal{A})$ 上の確率測度全体のなす空間とする。
- (4) $f: \mathcal{G} \rightarrow \mathbb{R}$ によって、 \mathcal{G} の全体で定義されているとは限らない部分関数 $f|_{\text{Dom} f}: \text{Dom} f \rightarrow \mathbb{R}$ を表す。
- (5) 自然数 $k \in \mathbb{N}$ について、 $[k] := \{1, 2, \dots, k\}$ と表す。
- (6) $C(\mathcal{G})$ によって、位相空間 \mathcal{G} 上の連続関数全体のなす空間とする。

また、パラメトリックモデルに関する仮定は (P1)~(P4)、ノンパラメトリックモデルに関する仮定は (N1)~(N5)、セミパラメトリックモデルに関する仮定は (S0)~(S4) がある。

2 パラメトリックモデルに於ける有効推定量の標準的構成法

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 [5] 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). $U \subset \mathbb{R}^q$ 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 $\psi: U \rightarrow \mathbb{R}^p$ と、その $\theta \in U$ における不偏推定量 $\hat{\psi} \in \mathcal{L}^2(\mathcal{G}^n; \mathbb{R}^p)$ 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 $\mu \in P(\mathcal{G})$ に関して絶対連続とし、その Radon-Nikodym 微分を $p_\theta: \mathcal{G} \rightarrow [0, 1]$ で表す。
- (P1) $p_-(x): U \rightarrow [0, 1]$ 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 $g_i: \mathcal{G} \times U \rightarrow \mathbb{R}$ を $g_i(x, \theta) := \frac{\partial \log p_\theta(x)}{\partial \theta_i} = \frac{1}{p_\theta} \frac{\partial p_\theta}{\partial \theta_i}$ と定めると^{†1}、二次の絶対積率が有限 $g_i(-, \theta) \in \mathcal{L}^2(P_\theta)$ ：

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

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

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

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

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

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

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

$$\text{Var}_\theta[\hat{\psi}] \geq J(\theta) I(\theta)^{-1} J(\theta)^\top.$$

[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,

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

これと $E_\theta[g] = 0$ より,

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

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

証明 すると, Cauchy-Schwartz の不等式の証明と同様に, $\text{Var}_\theta[\delta - JI^{-1}g] \geq 0$ 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,

$$\begin{aligned} 0 &\leq \text{Var}_\theta[\delta - JI^{-1}g] = \text{Cov}[\delta - JI^{-1}g, \delta - JI^{-1}g] \\ &= \text{Var}_\theta[\delta] - JI^{-1} \text{Cov}_\theta[g, \delta] - \text{Cov}_\theta[\delta, g] I^{-1} J^\top + JI^{-1} \text{Var}_\theta[g] I^{-1} g^\top \\ &= \text{Var}_\theta[\delta] - JI^{-1} J^\top. \end{aligned}$$

■

系 2.2. 特に $q = p = 1$ の場合, $\text{Var}_\theta[\hat{\psi}] \geq \frac{\left(\frac{\partial \psi(\theta)}{\partial \theta}\right)^2}{\text{Var}_\theta[g]}.$

2.2 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 2.3. $\Theta \subset \mathbb{R}^p$ を可測集合, $\psi: \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$ を推定関数とし, $\psi(x, -)$ は連続とする. $\psi_n(\theta) := \sum_{j=1}^n \psi(x_j, \theta)$, $\bar{\psi}_n(\theta) :=$

$\mathbb{E}_n[\psi(\theta)] = \frac{1}{n} \psi_n(\theta)$ と定め, 真の平均を $\bar{\psi}(\theta) := \int_{\mathcal{X}} \psi(x, \theta) P(dx)$ と定める.

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

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

によって定めると, 右辺は

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

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 $(\hat{\theta}_n)$ を $(\hat{\theta}_n^0)$ を初期推定量とするワンステップ推定量という.

定理 2.4. $\theta_0 \in \Theta^\circ$ 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 $\psi: \mathcal{X} \times \Theta \rightarrow \mathbb{R}^p$ 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) $\exists M \in \mathcal{L}^1(\mathcal{G}, \mathbb{R}) \forall x \in \mathcal{X}, \theta \in B |\partial_\theta \psi(x, \theta)| \leq M(x)$. さらに, $\Gamma := -P\partial_\theta \psi(x, \theta_0)$ は正則.
 (E5) 確率変数列 $\hat{\theta}_n^0: \mathcal{X}^n \rightarrow \Theta$ が \sqrt{n} -一致性を持つ: $\hat{\theta}_n^0 - \theta_0 = O_p(1/\sqrt{n})$.^{†2}

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

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

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,

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

が成り立つ.

3 セミパラメトリックモデルに於ける有効推定量の標準的構成法

3.1 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 $\mathcal{P} \rightarrow \mathbb{R}$ は, (識別可能性条件などに基づく) Euclid 部分空間との同一視 $\mathcal{P} \simeq U$ 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 $[0, 1] \rightarrow \mathcal{P}$ を取ることに他ならないが, すべての連続写像 $[0, 1] \rightarrow \mathcal{P}$ 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 $p: \mathcal{X} \rightarrow [0, 1]$ で表す.

をおく.

定義 3.1 (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 $(P_t): [0, 1] \rightarrow \mathcal{P}$ であって, ある 2 乗可積分な関数 $g: \mathcal{X} \rightarrow \mathbb{R}$ が存在して,

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

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 $[0, 1] \rightarrow \mathcal{P}$ を簡単に**正則な P -道**といい, その全体を $\text{Sub}_P(\mathcal{P})$ 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:

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

^{†2} 条件 $\hat{\theta}_n^0 - \theta_0 = O_p(1/\sqrt{n})$ は, 弱一貫性 $\hat{\theta}_n^0 - \theta_0 = o_p(1)$ 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.

注 3.2 (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, $q'(t, x) = \frac{\partial q(t, x)}{\partial t} = \frac{\partial \sqrt{p_t(x)}}{\partial t} = \frac{1}{2} \frac{p'_t(x)}{\sqrt{p_t(x)}}$ より,

$$g(x) = \frac{p'_0(x)}{p_0(x)} = \frac{p'(x)}{p(x)} = \frac{\partial}{\partial t} \log p_t(x) \Big|_{t=0}.$$

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

続いて, 汎関数 $\psi: \mathcal{P} \rightarrow \mathbb{R}$ 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 $\psi: U \rightarrow \mathbb{R}$ は単に偏微分可能としたのであった.

定義 3.3 (pathwise differentiable, von Mises differentiable). 汎関数 $\psi: \mathcal{P} \rightarrow \mathbb{R}$ 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 $(P_t) \in \text{Sub}_P(\mathcal{P})$ に対して, 連続線型汎関数 $\dot{\psi}_P: L_2(P) \supset \dot{\mathcal{P}}_P \rightarrow \mathbb{R}$ が存在して,

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

が成り立つことをいう. 汎関数 $\dot{\psi}_P: \dot{\mathcal{P}}_P \rightarrow \mathbb{R}$ を ψ の P における微分または影響関数と呼ぶ.

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

$$\left\{ \tilde{\psi} \in L^2(\mu) \mid \dot{\psi}_P(g) = (\tilde{\psi}_P|g) = \int \tilde{\psi}_P g dP \right\} \neq \emptyset$$

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 $\tilde{\psi}_P$ 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 $\tilde{\psi}(x; P') \in \mathcal{L}_0^2(P), R_2(P, P') \in \mathbb{R}$ が存在して,

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

が成り立つことをいう.^{†3}

3.2 セミパラメトリックモデルの効率限界

ノンパラメトリックモデル \mathcal{P} 上の (次の定理の意味で) 正則な汎関数 $\psi: \mathcal{P} \rightarrow \mathbb{R}$ の一致推定量が持ち得る漸近分散の最小値を, 有効影響関数を用いて表す. セミパラメトリックモデルは, \mathcal{P} が, ある $\Theta \subset \mathbb{R}^p$ と無限次元 Banach 空間 H との直積 $\Theta \times H$ の滑らかな像 $\Theta \times H \ni (\theta, \gamma) \mapsto P_{(\theta, \gamma)} \in \mathcal{P}$ として表せる場合であるから, もちろん次の定理の結果に従う.

定理 3.4 (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 $\tilde{\psi}(x; P) = \tilde{\psi}_P(x)$ として von Mises 可微分で, かつ $\dot{\mathcal{P}}_P$ に関して道毎に可微分な汎関数 $\psi: \mathcal{P} \rightarrow \mathbb{R}$ と, その $P \in \mathcal{P}$ における不偏推定量 $\hat{\psi} \in \mathcal{L}^2(\mathcal{X}^n; \mathbb{R})$ について, (S0) と以下の (N1)~(N3), そして定理 3.6 の (N4), (N5) を仮定する.

(N1) 積分と極限の可換性: 任意の $(P_t) \in \text{Sub}_P(\mathcal{P})$ に関して, $(\tilde{\psi}_P(p_t - p))$ 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 $(P_t) \in \text{Sub}_P(\mathcal{P})$ 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: $\tilde{\psi}_P \in \dot{\mathcal{P}}_P$.

このとき, $\hat{\psi}$ の分散は $\text{Var}_P[\hat{\psi}] \geq \text{Var}_P[\tilde{\psi}_P]$ を満たす.

[証明].

^{†3} 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 $\int_{\mathcal{X}} \tilde{\psi}'(P' - P)^2(dx)$ 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.

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

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

が成り立つ. 両辺を $\epsilon \neq 0$ で割って, $\epsilon \rightarrow 0$ の極限を考えることと仮定 (S1),(S2) より, $\dot{\psi}_P = \int_{\mathcal{X}} \tilde{\psi}_P h P(dx)$. よって

Cramer-Rao の定理 2.2 より, $\text{Var}_P[\tilde{\psi}] \geq \frac{E_P[\tilde{\psi}_P h]^2}{E_P[h^2]}$. 右辺は, 任意の $h \in \mathcal{L}_0(P)$ に関して, Cauchy-Schwartz の不等式から $E_P[\tilde{\psi}_P^2]$ により抑えられる.

等号成立条件について また, $h = \varphi \in \mathcal{L}_0(P)$ と取った時の P -道が $E_P[\tilde{\psi}_P^2]$ を与えることに注意すれば, あとは分散 $E_P[\tilde{\psi}_P^2]$ を持つ一致推定量 $\hat{\psi}$ を構成すれば良い. これは, 定理 3.6 による (ここで仮定 (N4),(N5) を用いる).

■

定理 3.5 (7.2 [4]).

- (S1) $\theta \mapsto P_{\theta,\eta}$ は L^2 -微分可能である.
- (S2) $P_{\theta_n,\eta}[\|\hat{l}_{n,\theta_n} - \hat{l}_{\theta_n,\eta}\|^2] \xrightarrow{P} 0$.
- (S3) 有効情報行列 $\tilde{I}_{\theta,\eta}$ は正則になる.
- (S4) \sqrt{n} -一致性 $\sqrt{n}P_{\theta_n,\eta}[\hat{l}_{n,\theta_n}] \xrightarrow{P} 0$ が成り立つ.

このとき, $(\hat{\theta}_n)$ は (θ, η) において漸近有効である.

3.3 影響関数を用いたワンステップ推定量

セミパラメトリックモデルにおいて, ノンパラメトリックな局外母数をここでは確率分布 \mathbb{P} として, これを推定したのち, この推定量 $\hat{\mathbb{P}}$ を用いて (その一致性と効率性に依らず), $\psi(\mathbb{P})$ の漸近有効推定量が構成できる. そしてこの「ワンステップ推定量」と呼ぶべき推定量は \sqrt{n} -一致性も持ち得る. そのための十分条件の一例を次に示す.

定理 3.6. $\hat{\mathbb{P}}$ を \mathbb{P} の推定量とする. これに対して, 推定量 $\hat{\psi}: \mathcal{X}^n \rightarrow \mathbb{R}$ を $\hat{\psi} := \psi(\hat{\mathbb{P}}) + \mathbb{P}_n[\tilde{\psi}_{\hat{\mathbb{P}}}]$ と定め, (S0), 定理 3.4 の (N1)~(N3) と, 次の (N4),(N5) を仮定する.

- (N4) $\hat{\psi}$ は次の K -標本分割アルゴリズムに沿って構成されたものであること: ある整数 $K \in \mathbb{N}$ について, 標本を K 個に分割し, $\hat{\psi}_k := \psi(\hat{\mathbb{P}}_{-k}) + \mathbb{P}_n^k[\tilde{\psi}_{\hat{\mathbb{P}}_{-k}}]$ を用いて

$$\hat{\psi} := \sum_{k=1}^K \left(\frac{N_k}{n} \right) \hat{\psi}_k$$

と構成されたものとする. ただし, $k \in [K]$ に対して, $N_k \in \mathbb{N}$ とは k 番目の小標本のサイズ, $\hat{\mathbb{P}}_{-k}$ とは k 番目の小標本を除いた標本から構成された推定量とし, \mathbb{P}_n^k とは k 番目の小標本についての経験分布とした. さらに, $\forall k \in [K] \quad \|\tilde{\psi}_{\hat{\mathbb{P}}_{-k}} - \tilde{\psi}_{\mathbb{P}}\|_{L^2(\mu)} = o_{\mathbb{P}}(1)$.

- (N5) $U := \sum_{k \in [K]} R_2(\hat{\mathbb{P}}_{-k}, \mathbb{P}) = o_{\mathbb{P}}(n^{-1/2})$.

このとき,

$$\sqrt{n}(\hat{\psi} - \psi) \xrightarrow{d} N(0, \text{Var}[\tilde{\psi}_{\mathbb{P}}]).$$

特に定理 3.4 と併せれば, $\hat{\psi}$ は \sqrt{n} -一致性を持つ漸近正規な漸近有効推定量である.

[証明].

推定量の標準分解

$$\begin{aligned} \hat{\psi} - \psi(\mathbb{P}) &= \psi(\hat{\mathbb{P}}) + \mathbb{P}_n[\tilde{\psi}_{\hat{\mathbb{P}}}] - \psi(\mathbb{P}) \\ &= \mathbb{P}_n[\tilde{\psi}_{\hat{\mathbb{P}}}] + \int_{\mathcal{X}} \tilde{\psi}_{\hat{\mathbb{P}}}(x)(\hat{\mathbb{P}} - \mathbb{P})(dx) + R_2(\hat{\mathbb{P}}, \mathbb{P}) \end{aligned}$$

$$\begin{aligned}
&= (\mathbb{P}_n - \mathbb{P})[\tilde{\psi}_{\hat{\mathbb{P}}}] + R_2(\hat{\mathbb{P}}, \mathbb{P}) && \because \hat{\mathbb{P}}[\tilde{\psi}_{\hat{\mathbb{P}}}] = 0. \\
&= (\mathbb{P}_n - \mathbb{P})[\tilde{\psi}_{\mathbb{P}}] + (\mathbb{P}_n - \mathbb{P})[\tilde{\psi}_{\hat{\mathbb{P}}} - \tilde{\psi}_{\mathbb{P}}] + R_2(\hat{\mathbb{P}}, \mathbb{P})
\end{aligned}$$

と分解できる．第 1,2,3 項をそれぞれ S_n, T_n, U とおく．

第 1 項 中心極限定理より，分布収束 $S_n \xrightarrow{d} N(0, \text{Var}[\tilde{\psi}_{\mathbb{P}}]/n)$ が成り立つ．

第 2 項 仮定 (N4) の下で， $T_n = o_{\mathbb{P}}(n^{-1/2})$ が成り立つことは [6] Th'm 3 による．また，仮定 (N4) の下では，第 3 項も

$$U = \sum_{k \in [K]} R_2(\hat{\mathbb{P}}_{-k}, \mathbb{P})$$

と表せる．この下で，仮定 (N5) における U と，証明内で上述した U とは一致する．

第 3 項 上述の議論と，r 仮定 (N5) により， $o_{\mathbb{P}}(n^{-1/2})$ ．

■

3.4 汎関数の微分可能性の精査と今後の発展

前節での定理を再検討し，数学方面からどのような貢献があり得るかを議論する．前節の定理の精緻化や，新たな汎関数に対する漸近有効推定量の構成法を考えるなどの応用上の未解決問題は，統計的汎関数 ψ の \mathbb{P} における「微分可能性」に関する数学的理解を深めることによって解決の兆しが見え得ることを議論する．

定理 3.6 においては，第 3 項 U の評価が肝心であった．定理の証明内で叙述した通り，第 2 項 T_n は経験分布 \mathbb{P}_n にも依存しており，標本分割を利用して過適合 (overfitting) を避けることで，その収束性のある程度制御できた．そこで，仮定 (N4) などの， $o_{\mathbb{P}}(n^{-1/2})$ となるための十分条件も見つかっている [6]．しかし第 3 項 U は von Mises 展開における 2 次の剰余項であり，数学的には統計的汎関数 ψ の定義のみからその性質が決まっているはずである．さらに言えば， ψ の \mathbb{P} の周りでの滑らかさの情報が殆どを決めていると考えられる．

ψ の \mathbb{P} の周りでの滑らかさの情報から第 3 項 U の振る舞いを予想する問題は未解決で，取り掛かるのは難しいと考えられるが，これは Malliavin 解析という既存の数学分野の応用を通じて打開策が発見される可能性がある．Weiner 空間 (原点を通る \mathbb{R}_+ 上の連続実関数の全体からなる集合に広義一様収束の位相を入れた Fréchet 空間など) 上の汎関数の微分として自然な概念として，Schwartz 超関数理論に基づく概念である「弱微分」に当たる Malliavin 微分を定義することで始まった Malliavin 解析なる分野は，Wiener 空間自体が無次元空間であることから，しばしば「無限次元空間上の解析学」と説明され，近年では拡散過程論やファイナンスへ応用されている [25]．ノンパラメトリック統計モデル $\mathcal{P} \subset P(X)$ は Wiener 空間同様無限次元であり，その上の汎関数 ψ について，定義 3.3 で考えたような道毎の可微分性や von Mises 可微分性よりも弱い概念になるであろう「Malliavin 微分」をうまく考えることが出来れば，その汎関数 ψ の「Malliavin 可微分性」に応じて，第 3 項 U の振る舞いが分類出来る，というような数学理論が出来る可能性は十分にないと筆者は考える．現代数学における微積分学の舞台は可微分多様体であるが，実際，Malliavin が Wiener 空間を Wiener-Riemann 多様体と名付け直して微分幾何学の手法で分析し，曲率の概念などを定義した ([?], [24])．全く同様に，セミパラメトリックモデル $\mathcal{P} \subset P(\mathcal{X})$ は， \mathcal{X} がコンパクトハウスドルフ空間であるとき，Banach 代数 $C(\mathcal{X})$ の双対空間 $(C(\mathcal{X}))^*$ に埋め込まれた， w^* -コンパクトな Frechet 多様体とみなせる [26]．ここで，例えば $\mathcal{X} = [0, T]$ ($T > 0$) と実数上の閉区間とすれば，確率分布全体の空間 $P(\mathcal{X})$ は，古典的 Wiener 空間 $C(\mathcal{X})$ の双対空間の凸部分集合であるので，深い関わりがあることは間違いないだろう．

すなわち，「定理 3.6 の条件 (N5) がいつ成り立つか？」または「どのような統計的汎関数 ψ に対して，定理 3.6 が成り立つようなワンステップ推定量が構成可能か？」という実用的にも有益な問いに対しては，既存の数学理論を変形・応用することによって解決され得る．統計的汎関数 $\psi: \mathcal{P} \rightarrow \mathbb{R}$ とは，因果推論における平均処置効果 $\psi(P) := E_P[E_P[Y|X, A=1] - E_P[Y|X, A=0]]$ など，各応用の現場において生じる実際的な問題に合わせて設定されたパラメータが想定されている．この考え方は，Cowles 委員会で活躍した計量経済学者である Jacob Marschak が，当時興隆していた構造方程式モデルに疑問を呈する形で提唱した，実際上の問題から関心のあるパラメータを明確にし，これを推定することを考えるという Marschak's Maxim に起源を見いだせる枠組みであるが [27]，真にその枠組みを実現するには，広範な ψ に対して，「漸近有効推定量が構成可能であるか？」，さらには「定理 3.6 が成り立たない場合でも，別の方法で漸近有効推定量が構成可能である場合はあるか？」などの未解決問題 [5] などの知識が重要な課題となる．

統計的汎関数 $\psi: \mathcal{P} \rightarrow \mathbb{R}$ の微分を調べる試みは、 \mathcal{P} を無限次元の場合も含めて一般的に扱われたことは少ないにせよ、より簡単な設定で微分や滑らかさの概念が考えられたことがないわけではない。そもそも影響関数の概念は、本エッセイで取り上げているようなセミパラメトリックモデルに関するコンテキストとは独立に、Hampel の頑健統計に関する博士論文 (1968,[9]) 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応用される。その意味で、非常に安定な数学的概念であることが窺える。そのことも含め、Ichimura and Newey (2022,[16]) で「影響関数とは、滑らかな道に関する Gateaux 微分である」という見方が提示されているように、この方向での研究をさらに押し進める形で、一度「影響関数」の数学的な素性を明らかにしたい。

その試みは、von Mises (1947,[14]) と Hampel (1974,[11]) の仕事の「ノンパラメトリックな変量 \mathbb{P} を最初に推定する必要がある場合」への一般化と [16] で議論されている通り、確率分布全体のなす空間を考えて、その上での「微分」を考える発想は、元を辿れば、von Mises の 1947 年の論文 [14] で定義された「統計的汎関数の Volterra 微分」の概念の精緻化でもある。

以上のように、1947 年以来、漸近分布の分布クラスを統計的汎関数の滑らかさによって分類しようとした試み von Mises (1974,[14]) や、頑健性の定義の提案 Hampel (1968,[9]) など、複数の文脈で独立に着想された統計的汎関数の微分の概念を、「無限次元局所凸線形空間上の汎関数の微分」として数学的自然な一般的定義を用いて捉え直し、

- (1) 定理 3.6 が成り立つような統計的汎関数 ψ のクラスを、Malliavin 解析が確率微分方程式の問題を乗り越えたように、新たに「弱微分」の概念をうまく定義することで、数学的に特徴づけることができないか？
- (2) 特に、von Mises 可微分な統計的汎関数について、2 次以上の高次の剰余項 U の収束性を、応用の現場で重要と理解された ψ について簡単に判別できるような規準を見つけることができないか？

などの観点から整理することが、数学方面から手法論への貢献の 1 つとして考え得る。

歴史 3.7 (von Mises の研究について). 余談であるが、von Mises (1947,[14]) では、Volterra (1913,[13]) 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。これは無限次元空間上の解析学を樹立する試みの萌芽とも思える。

4 頑健統計への応用

前節では、セミパラメトリックモデルにおいて、ノンパラメトリックな局外母数の推定と、これを用いたパラメトリックな推定の 2 段階構成において、どのようにすれば漸近有効な推定量が構成できるかについて 1 つの解答を与え、この研究を Marschak 以来の研究方針の大きな流れの中に位置づけ直し、この先に数学の分野からどのような貢献が可能化についてのグランドデザインを描いた。

この節では、より喫緊な問題に対する応用として、二重に頑健な推定量に関する未解決問題に対する数学分野からの貢献の可能性について考えたい。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。

背景 4.1. セミパラメトリックモデル $\mathcal{P} \subset P(\mathcal{X})$ が十分に大きい場合、第 1 段階のノンパラメトリックな局外母数の推定は当然困難になっていく。ノン・セミパラメトリックモデルは、分布に関する仮定をおかないため、モデルの misspecification が起こりにくく、適用しやすいため応用上非常に魅力的である。しかし応用上の一番の問題がこの次元の呪いで、長い間ノン・セミパラメトリックモデルの応用を阻んでいた実際的な問題であった。

では、なぜ今になって、そのような困難な状況の打開がセミパラメトリックモデルにおいて考えられているのであろうか。これは次のような理由による。まず、パラメトリック部分 $\psi(\mathbb{P})$ の推定量の漸近的な振る舞いは、計量経済学・生物統計学分野などに出現する高次元モデルでの推定量の漸近的な振る舞いに非常に似ている [28] ため、理論的にも重要である。しかしそれだけでなく、セミパラメトリックモデルにおいては、次のような手法が考えられるためである。素朴に「セミパラメトリックモデルにおいて頑健な手法はなにか？」を考えると、ノンパラメトリック部分 \mathbb{P} にパラメトリックな仮定をおく、または特定の正則化手法を採用する第 1 段階と、パラメトリック部分 $\psi(\mathbb{P})$ の推定との 2 段階の推定の、両方を頑健にする旧来の発想の結合が考えられる。しかしそれに留まらず、「2 つの段階のいずれかの推定にバイアスが生じて、全体として統計的汎関数 ψ の推定に大きな影響を生じない」という意味での、セミパラメトリックモデルに特有な頑健性が考え得る。これは二重に頑健な推定量 (Doubly robust / doubly protected estimator) と呼ばれる。

このとき、次の2つの問題が未解決である。

- (1) 二重に頑健な推定量が存在するのはいつか。
- (2) 二重に頑健な推定量の標準的な構成法が存在するか。

この節の残りでは、この2つを議論する。

定義 4.2. セミパラメトリックモデル $\mathcal{P} = (P_{(\theta, \gamma)})_{(\theta, \gamma) \in \Theta \times H}$ ($\Theta \subset \mathbb{R}^p, H \subset \text{Map}(\mathcal{X}; \mathbb{R})$) のパラメータ θ の推定を考える。

- (1) 真値 $\theta_0 \in \Theta, \gamma_0 \in H$ に対して、 $E[g(X, \gamma_0, \theta_0)] = 0$ かつ $\forall \theta \in \Theta \setminus \{\theta_0\} E[g(X, \gamma_0, \theta)] \neq 0$ を満たす関数 $g : \mathcal{X} \times H \times \Theta \rightarrow \mathbb{R}^q$ を推定関数という。
- (2)

参考文献

- [1] 吉田朋広. (2006). 数理統計学. 朝倉書店.
- [2] 芝村良. (2004). 『R. A. フィッシャーの統計理論』. 九州大学出版会.
- [3] van der Vaart, A. W. (1998) *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge.
- [4] Bolthausen, E., Vaart, A., and Perkins, E. (2002). *Lectures on Probability Theory and Statistics*. Springer Berlin, Heidelberg.
- [5] Kennedy, E. H. (2022). Semiparametric Doubly Robust Targeted Double Machine Learning: A Review. arXiv: 2203.06469.
- [6] Kennedy, E. H., Balakrishnan, S., and G' Sell, M. (2020). Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*. 48(4): 2008 – 2030.
- [7] Bickel, P. J., Chris A.J. Klaassen, Ya'acov Ritov, Wellner, J. A. (1998) *Efficient and Adaptive Estimation for Semiparametric Models*. Springer New York.
- [8] Bickel, P. J., Kwon, J. (2001). Inference for Semiparametric Models: Some Questions and an Answer. *Statistica Sinica*. 11: 863-960.
- [9] Hampel, F. R. (1968). *Contributions to the theory of robust estimation*. Unpublished dissertation, Berkeley: University of California.
- [10] Hampel, F. R. (1971). A General Qualitative Definition of Robustness. *The Annals of Mathematical Statistics*. 42(6): 1887-1896.
- [11] Hampel, F. R. (1974). The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, 69(346): 383-393. DOI: 10.1080/01621459.1974.10482962.
- [12] Bliss, G. A. (1915). A Note on Functions of Lines. *Proceedings of the National Academy of Sciences of the United States of America*. 1(3): 173-177.
- [13] Vito Volterra. (1913). *Leçons sur les fonctions de lignes*. Paris: Gauthier-Villars.
- [14] von Mises, R. (1947). On the Asymptotic Distribution of Differentiable Statistical Functions. *The Annals of Mathematical Statistics*. 18(3): 309-348. DOI: 10.1214/aoms/1177730385.
- [15] Newey, W. K., and McFadden, D. (1994). Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics*. Volume 4. North-Holland, an imprint of Elsevier, Amsterdam.
- [16] Ichimura, H., and Newey, W. K. (2022). The Influence Function of Semiparametric Estimators. *Quantitative Economics*. 13(1): 29-61. DOI: 10.3982/QE826.
- [17] Chernozhukov, V., Escanciano, J.C., Ichimura, H., Newey, W. K., and Robins, J. M. (2016). Locally Robust Semiparametric Estimation. arXiv:1608.00033.
- [18] Chernozhukov, J., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. K., and Robins, M. J. (2018). Double/Debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal*. 21(1): C1-C68.

- [19] Chernozhukov, V., Newey, W. K., and Singh, R. (2022). DeBiased Machine Learning of Global and Local Parameters Using Regularized Riesz Representers. *The Econometrics Journal*. utac002, 00: 1-26.
- [20] Rotnitzky, A., Smucler, E., and Robins, J. M. (2020). Characterization of parameters with a mixed bias property. *Biometrika*. 108(1): 231-238.
- [21] Seaman, S. R., and Vansteelandt, S. (2018). Introduction to Double Robust Methods for Incomplete Data. *Statistical Science*. 33(2): 184-197.
- [22] Hines, O., Dukes, O., Diaz-Ordaz, K., and Vansteelandt, S. (2022). Demystifying statistical learning based on efficient influence functions. *The American Statistician*. DOI: 10.1080/00031305.2021.2021984.
- [23] Malliavin, P. (1978). Stochastic Calculus of Variations and Hypoelliptic Operators. Proceedings of the International Symposium on Stochastic Differential Equations (1976 Kyoto). Wiley New York.
- [24] Malliavin, P., and Cruzeiro, A. B. (1996). Renormalized Differential Geometry on Path Space: Structural Equation, Curvature. *Journal of Functional Analysis*. 139: 119-181.
- [25] Shinzo Watanabe. (1987). Analysis of Wiener Functionals (Malliavin Calculus) and its Applications to Heat Kernels. *Annals of Probability*. 15(1): 1-39. DOI: 10.1214/aop/1176992255.
- [26] Pedersen, G. K. (1989). *Analysis Now*. Springer New York.
- [27] van der Laan, M. J., Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer New York.
- [28] Robins, J. M., and Ritov, Y. (1997). Toward a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory For Semi-parametric Models. *Statistics in Medicine*. 16: 285-319.

(1) 漸近展開やめる??回収出来ないならもういい. 漸近展開は知られてない.

(2) Frechet 多様体の話を, Doubly Robust で拾えたら最強. 純粋数学の方法が, Doubly Robust で活かす可能性を見せると売れる. そして, 「だから, 異分野統合に数学科が適任なんだ」

(3) 研究計画は具体的に. 数学の論文を引用しても良いかも. 平易な言葉で書いて, 「エッセイ参照」がいいだろう.

(4) 影響関数はパラメトリックでも使う. Ichimura and Newey.

(5) いつ使えばいいのかわからないことがあるから, misspecification は結構ありえる. 実用上, 普及させる上で非常に大事になる. (根拠となる応用論文を探したい). 経済の人間でも使い方がわかる人は限られる, 統計教育の普及よりも, 手法に埋め込まれた祈りの方が大事. 最終形態は R パッケージとその先であることをよく考えるように.

適用の実務の視点: 適用可能性が非常に広いモデルがほしい, いまはそれがないから, コンサルタントがはびこる. 「この場合はパラメトリックを使いましょう, この場合は正規性の仮定が怪しいので, ノンパラで行きましょう」など. データ数を見て Deep を使う方がいいか, 漸近展開を使わないと怪しいかの境界の判定など.

Doubly Robust を使わなかった場合, どのようなひどいことが起こるか, とかの論文がみつかったらいいね. 「パラメトリックとノンパラメトリックの選択ミス」という脆弱性も新たに解決可能になったわけだから, このミスが何を引き起こし得るか.

(6) 計量のセミパラの人は Newey, 計量の影響関数の人は市村.

(7) van der Vaart の DR が読みやすいかも. 427 の missing data.

(8) 今のところ一番一般化されているのが DR.