

# 機械学習×計量経済学：Double/Debiased Machine Learning

- [source](#)

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018a). "Double/Debiased Machine Learning for Treatment and Structural Parameters", *Econometrics Journal*, 21, pp.C1–C68.

矢田紘平

## 機械学習と計量経済学

本稿では、計量経済学モデルの推定に機械学習の手法を応用する方法を提案した論文Chernozhukov et al. (2018a)を紹介する。ここでは、機械学習をLASSO、ランダムフォレスト、勾配ブースティング、ニューラルネットワーク等に代表される「予測・分類アルゴリズム」と捉えて議論する(厳密に「機械学習」とは何かを定義するのは、かなりハードな問題である)。

さて、機械学習と計量経済学の違いの1つとして、機械学習は予測精度を高めることを目的とするのに対し、計量経済学はモデルのパラメータの推定を目的とするということがしばしば挙げられる。

たとえば、個人の賃金 $Y$ と特徴 $X$ のデータがあるとしよう。機械学習は、データに存在しない別の個人の賃金 $Y$ をその人の特徴 $X$ から高精度で予測する手法(たとえば、平均二乗誤差 $E[(Y - f(X))^2]$ が小さい $f$ を求めるアルゴリズム)の開発を行う。とりわけ、 $X$ が高次元(データサイズと比して $X$ の次元が大きい)の場合でも性能がよいことが求められる。

一方、計量経済学は、変数間の関係を記述するモデルを立て、そのモデルのパラメータを推定することを目指す。たとえば、モデルを $Y = X_1\beta_1 + X_2'\beta_2 + U$  (ここで、 $X_1$ は教育年数、 $X_2$ はそれ以外の特徴を示す変数が入ったベクトル、 $U$ は「能力」等の観察できない $Y$ の決定因子)とし、教育のリターンを表す $\beta_1$ を推定したい、といった具合である。

このように、機械学習と計量経済学は異なった問題に取り組むが、近年、計量経済学モデルの推定に機械学習の手法を応用する試みが進展している。本稿はそこからChernozhukov et al.(2018a)を取り上げて紹介しよう。

## 設定

あるモデルの推定したいパラメータ $\theta_0$ が、モーメント条件

$$E[\psi(W; \theta_0, \eta_0)] = 0$$

を満たすとする。ここで $W$ は確率変数ベクトル、 $\psi$ はスコア関数と呼ばれる、既知の関数である。 $\eta_0$ は未知の関数である。関数 $\eta_0$ は線形などのパラメトリックな形を仮定されておらず、分析者の興味の対象ではないパラメータである。。このモデルのように、推定したいパラメータが有限次元であるが、興味の対象でない関数パラメータを含むモデルはセミパラメトリック・モデルと呼ばれ、政策効果・処置効果の推定、産業組織論や労働経済学における構造モデルの推定等で使われている。

例として、以下の部分線形モデルと呼ばれるセミパラメトリック・モデルを考えよう。

$$Y = D\theta_0 + g_0(X) + U, E[U|X, D] = 0$$

ここで、 $Y$ はアウトカム、 $D$ は処置、 $X$ は共変量、 $U$ は観察できないノイズである。 $g_0$ は未知の関数である。このモデルでは、処置の効果 $\theta_0$ を推定したいが、関数 $g_0$ は興味の対象ではない。このモデルにおいて、モーメント条件を満たすスコア関数 $\psi$ は、複数存在する。たとえば、 $W = (Y, X, D)$ ,  $\psi_1(W; \theta, g) = (Y - D\theta - g(X))D$  とすると、真のパラメータ $(\theta_0, g_0)$ において、モーメント条件 $E[\psi_1(W; \theta_0, \eta_0)] = 0$ を満たす。また、 $m_0(X) = E[D|X]$ と定義し、 $\psi_2(W; \theta, \eta) = (Y - D\theta - g(X))(D - m_0(X))$  とすると、真のパラメータ $(\theta_0, \eta_0)$ において、モーメント条件 $E[\psi_2(W; \theta_0, \eta_0)] = 0$ を満たす。このように、モーメント条件を満たすスコア関数は複数存在する。

## Double/Debiased Machine Learning

サイズ $N$ のデータセット $W_{i=1}^N$ が与えられたとする。このデータセットから、パラメータ $\theta_0$ を推定するための手順を考える。通常、 $\theta_0$ を推定するためには、まず未知の関数 $\eta_0$ を推定する必要がある。関数の推定の古典的な手法として、カーネル回帰やシリーズ推定が知られるが、関数の引数が高次元である場合は、関数の推定を予測タスクに置き換え、機械学習を用いて推定するという方法が新たに提案されている。たとえば、上記の部分線形モデルにおける $m_0$ の推定は、 $D$ を $X$ で予測するタスクとみなせるため、LASSOやランダムフォレスト等を用いて行うことが可能である[<sup>1</sup>]。機械学習を用いた推定は、とりわけ関数の引数が高次元の場合、推定誤差や収束レートの面で優れていることが理論的・実証的に示されている。データサイズと比べて膨大な数の変数がある場合や、関数の引数にどの変数を入れるべきかわからない場合は、機械学習を用いることが妥当であろう。Chernozhukov et al. (2018a)は、機械学習を用いた $\theta_0$ の推定方法として、Double/Debiased Machine Learning(以下、DML)と呼ばれる以下の方法を提案し、その統計的性質を示した[<sup>2</sup>]。

1. 無作為にデータセット $\{W_i\}_{i=1}^N$ をサイズ $n = N/2$ の2つのサンプル $\{W_i\}_{i \in I_1}$ と $\{W_i\}_{i \in I_2}$ に分割する。
2. 2つめのサンプル $\{W_i\}_{i \in I_2}$ を使い、機械学習を用いて $\eta_0$ の推定量 $\hat{\eta}_{0,1}$ を得る。次に、1つめのサンプル $\{W_i\}_{i \in I_1}$ を使い、機械学習を用い、 $\frac{1}{n} \sum_{i \in I_1} \psi(W_i; \theta, \hat{\eta}_{0,1}) = 0$ を $\theta$ について解き、推定量 $\hat{\theta}_{0,1}$ を得る。ただし、 $\psi$ はモーメント条件に加えて、後述するネイマン直行条件を満たすスコア関数である。
3. ステップ2を、2つのサンプルを入れ替えて行い、推定量 $\hat{\theta}_{0,2}$ を得る。
4. 最終的な推定量を $\hat{\theta}_0 = \frac{1}{2}(\hat{\theta}_{0,1} + \hat{\theta}_{0,2})$ とする。

ここで、ステップ2および3において $\eta_0$ の推定のためにどのような機械学習アルゴリズムを使うのかについての指定はなく、その選択は分析者に委ねられる。 $\eta_0$ の推定量の収束レートなどに関する条件が満たされる限り、使用するアルゴリズムによらず、DML推定量 $\hat{\theta}_0$ は $\frac{1}{\sqrt{N}}$ のレートで $\theta_0$ に収束し、漸近的に正規分布にしたがう。

これらの統計的性質を得るための鍵となるのが、①ネイマン直交条件を満たすスコア $\psi$ を用いること、②cross-fittingを用いることである。この2つについてそれぞれ説明する。

## ネイマン直交条件

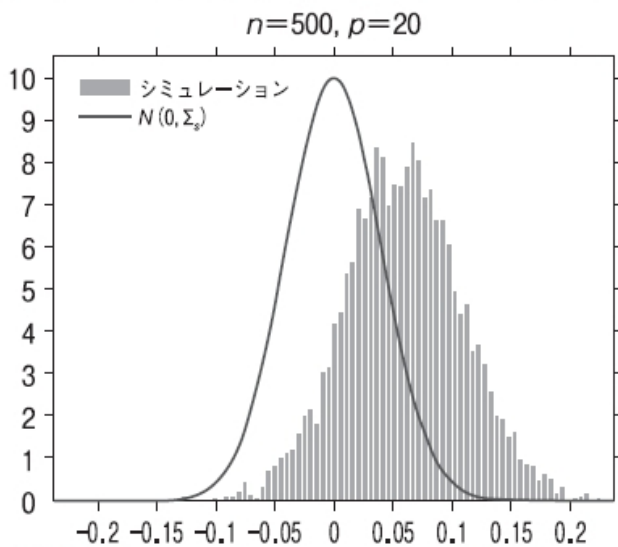
機械学習アルゴリズムは、予測モデルがデータに過剰にフィットすることを防ぐため、正則化と呼ばれる、モデルの複雑さに制約を加える処理を行う。これにより、 $\eta_0$ の推定量の分散が小さく抑えられる一方で、その代償としてバイアスが大きくなる。そのバイアスは、モーメント条件を通じて、 $\hat{\theta}_0$ に影響を与える。以下で与えられるネイマン直交条件は、その影響の程度が軽減されるための条件である[<sup>3</sup>]。

$$\left. \frac{\partial E[\psi(W; \theta_0, \eta_0 + r(\eta - \eta_0))]}{\partial r} \right|_{r_0} = 0$$

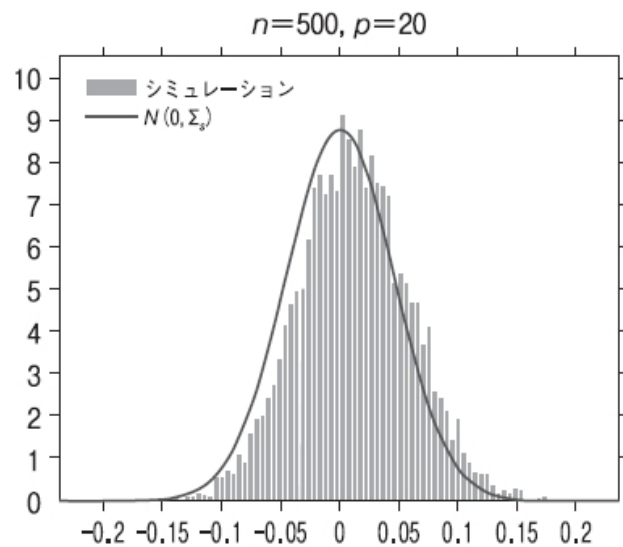
左辺は関数 $\eta_0$ を $\eta - \eta_0$ の方向に微小に動かしたときの $E[\psi(W; \theta_0, \eta_0)]$ の変動を表す。この変動がゼロであるということは、モーメント条件が $\eta_0$ の推定誤差に対して頑健であることを意味する。その結果、バイアスのある $\eta_0$ の推定量を用いたとしても、 $\hat{\theta}_0$ への影響が抑えられる。

部分線形モデルの例を挙げた際、モーメント条件を満たす2つのスコア( $\psi_1$ と $\psi_2$ )を紹介したが、簡単な計算により、スコア $\psi_1$ は直行条件を満たさない一方、スコア $\psi_2$ は直行条件を満たすことが示される。したがって、前者のスコアを用いた推定量の方が、 $\eta_0$ の推定のバイアスの影響を受けやすいと考えられる。[[fig:sim](#)] はこの2つのスコアを用いた推定量をシミュレーションにより比較したもので、ヒストグラムはシミュレーションで得られた $\hat{\theta}_0 - \theta_0$ の分布を表している。 $\psi_1$ を用いた推定量は、 $\psi_2$ を用いた推定量に比べて、バイアスが大きくなっていることがわかる。

(a) ネイマン直交条件を満たさないスコア $\psi_1$ による



(b) ネイマン直交条件を満たすスコア $\psi_2$ による

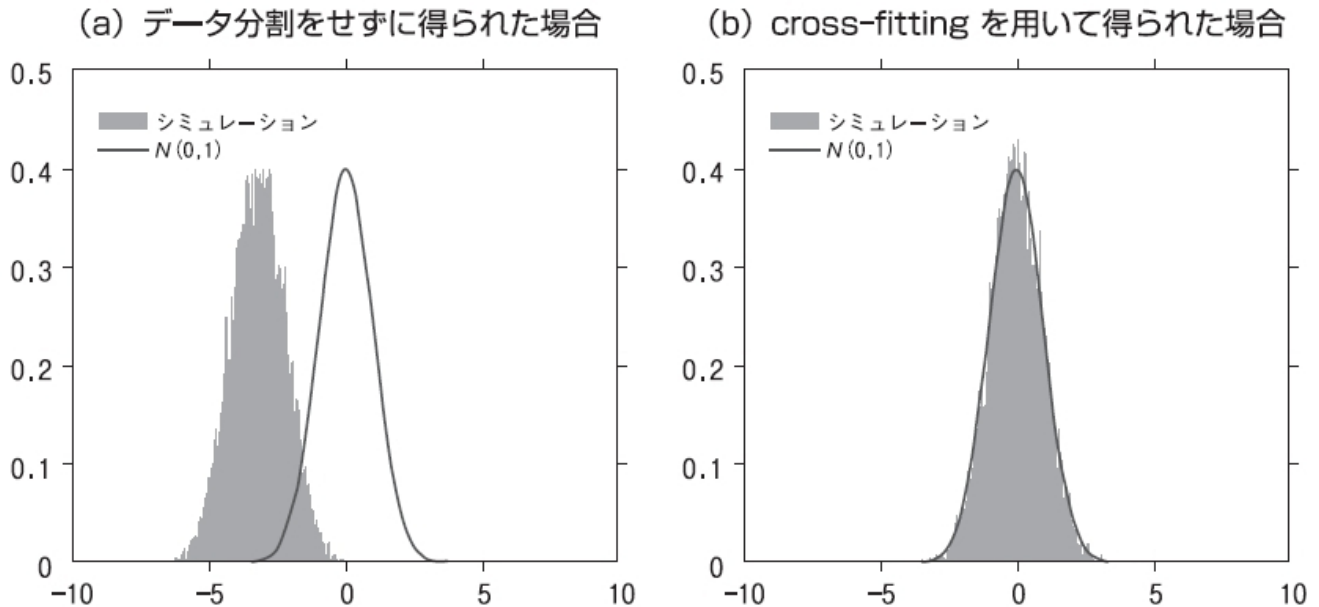


出所) Chernozhukov et al. (2018a).

{[fig:sim](#)}

## Cross-fitting

DMLのもう1つの特徴は、データ分割を行い、 $\eta_0$ の推定と $\theta_0$ の推定を別々のサンプルで行うことである。データ全体を用いて $\eta_0$ と $\theta_0$ の両方を推定した場合、過学習によるバイアスが生じる ([[fig:sim2](#)])。一方で、データ分割を行うと、 $\eta_0$ の推定量が $\theta_0$ の推定に用いるサンプルと独立になるため、バイアスを軽減することができる。理論的には、データ分割を行うことで、しばしばデータ分割を行わない場合より緩い条件の下で、推定量の $\frac{1}{\sqrt{n}}$ の収束レートや漸近正規性を示せることがわかっている。



出所) Chernozhukov et al. (2018a).

{#fig:sim2}

また、DMLでは、ステップ2で $\theta_0$ の推定量 $\hat{\theta}_{0,1}$ を得た後、2つのサンプルを入れ替えて $\theta_0$ の2つめの推定量 $\hat{\theta}_{0,2}$ を求め、その二つの平均を最終的な推定量としている(cross-fitting)。これは、データすべてを使うため効率性を損なわずにバイアスを取り除くことが可能になる。

## DMLの応用可能性と今後の展望

本稿で紹介したDMLは、セミパラメトリック・モデルの関数パラメータ( $\eta_0$ )が高次元な場合においても、関心のあるパラメータ( $\theta_0$ )をバイアスの小さい形で推定することを可能にする。代表的な応用例は、部分線形モデルの推定や、平均処置効果等の処置効果の推定であり、因果推論やプログラム評価において、多数の共変量を適切にコントロールするためにDMLを適用する実証研究が登場してきた。一方、その他のモデルにDMLを応用するには、直交条件を満たすスコア $\psi$ をみつけることが必要である。たとえば、本稿で紹介したものとは別の研究であるChernozhukov et al. (2018b)はスコアの構築方法を提示しており、例として動学的離散選択モデルの推定のためのスコアを導出している。今後は、処置効果の推定だけでなく、産業組織論や労働経済学等の分野における構造モデルの推定にもDMLが応用されることが考えられる。

[^1]:部分線形モデルにおいて、関数 $g_0$ はたとえば次のような方法で推定できる。 $\hat{\theta}_0$ の初期値を設定し、以下のプロセスを収束するまで繰り返す：①ランダムフォレストにより $Y - D\hat{\theta}_0$ を $X$ で予測し、 $g_0$ の推定値 $\hat{g}_0$ を得る。② $Y - \hat{g}_0(X)$ を $D$ で線形回帰し、 $\hat{\theta}_0$ を得る。

[^2]:論文では、データを $K$ 個のサンプルに分割し、 $\eta_0$ の推定に使うサンプルと $\theta_0$ の推定に使うサンプルを動かしていき、得られた $K$ 個の $\theta_0$ の推定量を平均する方法が紹介されているが、本稿では単純化のため $K = 2$ のケースを考える。推定量の漸近的性質は $K$ によらない。

[^3]:紙面の都合上、ここではややインフォーマルな直交条件の定義を述べている。正確な定義については、論文を参照されたい。

[https://en.wikipedia.org/wiki/Donsker%27s\\_theorem](https://en.wikipedia.org/wiki/Donsker%27s_theorem)