

異質性治療効果を推定するための統計手法・機械学習手法の現況 と医療分野への応用について

下川敏雄

和歌山県立医科大学 医学部 / 附属病院 臨床研究センター

今回のお話における問題設定と統計的課題

■ はじめに

樹木モデルを基調とした治療効果モデル

Meta-learner

- 変換アウトカム/修正共変量法
 共有基底条件付き平均回帰
 おわりに
- おわりに

はじめに

今回のお話における問題設定と統計的課題





個別化医療に向けての統計的課題



今回のお話における問題設定と統計的課題



RWD等を活用した個別化医療に対するアプローチ



治療効果モデルの諸型 (主なものだけに限定しています)



樹木モデルを基調とした治療効果モデル

因果樹木法 (Athey & Imbens, 2016)

Athey & Inbens(2016)は、CART法(Breiman et al., 1984)では、ふし内平均(回帰パラメータの推定)が 樹木構築(終結ふしの推定)の影響を受けることから、パラメータの独立性が担保できないことを問題 視した.

このことを解決するための方策として、Honest推定を提案している.



Athey, S. & Imbens, G.W. (2016). Recursive partitioning for heterogeneous treatment effect. Proceedings of the National Academy of Science, 113(27), 735–7360.

因果樹木法 (Athey & Imbens, 2016)

Athey & Imbens (2016)は、樹木モデルを構築するための損失関数として、修正MSEを提案するとともに、その期待値(EMSE;期待MSE)の負値を最大にすることを提案している。そのもとで、HTE推定に修正している。

因果樹木の樹木構築で用いられる損失関数(EMSE)

$$\widehat{EMSE} = \frac{1}{N^{\mathrm{tr}}} \sum_{n \in \mathcal{L}^{\mathrm{tr}}} \widehat{\tau}(\boldsymbol{X}_n, \widehat{\Pi}) - \left(\frac{1}{N^{\mathrm{tr}}} + \frac{1}{N^{\mathrm{est}}}\right) \sum_{t_m \in \Pi} \left(\frac{s_{\mathrm{tr},m,1}^2}{p_m} + \frac{s_{\mathrm{tr},m,0}^2}{1 - p_m}\right)$$

 N^{tr} :学習標本の標本サイズ, N^{est} :推定標本の標本サイズ, p_m :ふし $t_m \in \Pi$ における $Z_m = 1$ の割合 $\hat{\tau}(\boldsymbol{X}_n, t_m \in \hat{\Pi}) = \bar{Y}_{1,m}^{\text{tr}} - \bar{Y}_{0,m}^{\text{tr}}$: ふし $t_m \in \Pi$ における推定HTE(応答平均の差) $\bar{Y}_{k,m}^{\text{tr}}$ (k = 1, 0): ふし $t_m \in \Pi$ における群kの応答平均 $s_{\text{tr},m,k}^2$ (k = 1, 0): ふし $t_m \in \Pi$ における群kの不偏分散

EMSEの負値の最大化を行うことが、因果樹木法の樹木構築の目的である.

因果樹木法の構成アルゴリズム

最適樹木の選定 推定HTEの計算 樹木モデルの構築 テスト標本Steを用いる 推定標本 S^{est} を用いる 学習標本 S^{tr} を用いる 推定標本を樹木モデルに当 樹木系列のそれぞれに対し てはめることで,各終結ふ 学習標本においてEMSEが最 て、テスト標本のもとで <u>し(leaf)での推定標本におけ</u> 小になるようにモデルを構 EMSEを評価し、最小になる る推定HTE(終結ふし内での 築する 樹木を最適サイズ樹木(right 応答平均の差)を計算し,因 sized tree)とする. 果樹木での推定値とする。

Athey, S. & Imbens, G.W. (2016). Recursive partitioning for heterogeneous treatment effect. Proceedings of the National Academy of Science, 113(27), 735–7360.

因果樹木法の医学研究への応用:Look AHEAD Study

これは、米国16施設において2型糖尿病を伴う肥満患者5145人を対象にカロリー制限および運動といった生活習慣に対する介入 群とコントロール群に無作為に割り付けた無作為化比較試験である。主要評価項目は、Cause specific survival time(心血管系が 原因の死亡、致死的心筋梗塞、非致死的脳卒中、狭心症)である。



主論文(The Look AHEAD research group, 2013)では、介入によるダイエット効果が認められたが、それがイベント発現に差異を及ぼすには至らなかったと結論付けている

The Look AHEAD Research Group. (2013). Cardiovascular effect of intensive lifestyle intervention in type 2 diabetes, NEJM, 369, 145-154.

因果樹木法の医学研究への応用:Look AHEAD Study

疑問:どのような患者には積極的に栄養・運動介入を行うべきか?



Baum, A. et al. (2017). Targeting weight loss interventions to reduce cardiovascular complications of type 2 diabetes: a machine learning-based post-hoc analysis of heterogeneous treatment effects in the Look AHEAD trial, Lancet Diabetes Endocrinol., 5(10), 808-815.

因果樹木法を基本学習器とするアンサンブル学習法: Causal Forest法

Wager & Athey (2018)は、因果樹木に基づくRandomForest法を提案している。そこでは、樹木モデル(base learner)として、 Double-sample trees及び、Propensity treesを提案している:

Double-sample trees

Double-sample treesでは、観測値 $L = \{X_n, Y_n, Z_n\}_{n=1}^N$ をランダムに半分に分け($L = \{I, J\}$)、次のようなアルゴリズムによってモデルを構築している.

STEP.1: 樹木を成長させる. このとき,標本 Jを用いる分岐候補の選定には, Causal treeのEMSEを用いる.

STEP.2: 各終結ふしにおける応答の予測値 $\hat{\tau}_{O}$ 計算には、Iのみを用いる.

→ したがって, Athey & Imbens (2016)のHonest型のCausal treesと同じ

Propensity trees

Propensity treesでは、 群 Z_n を応答と見做したもとで樹木を構築したもとで、 終結ふしに対する治療効果の予測値 $\hat{\tau}_l$ を応答 に基づいて推定する.

STEP.1: 樹木を成長させる. このとき、(X_n,Z_nに対して、分類樹木(Breiman et al., 1984)によって樹木を構築する.

STEP.2: 各終結ふしにおける応答の予測値 \hat{t}_i を計算する.

なお、オリジナルでは、サブサンプルを用いるような記法になっている(後述のCausal forestに拡張するため).

Causal forestは, double-sample treeあるいはpropensity treeをbase learnerとしたもとでrandom forestのアルゴリズムによって でアンサンブルさせた方法である。ただし、ブートストラップ標本は、<u>重複なし</u>とする。

さらに, Wager & Athey (2018)では, Causal Forest法での漸近正規性および, 分散推定量を提案している.

因果樹木法を基本学習器とするアンサンブル学習法:因果Boosting法

Powers et al. (2018)は、因果樹木を基本学習器(base learner)に用いるBoostingとして、因果Boosting法を提案している。アルゴリズムは、通常の勾配ブースティング(Friedman,2001)と同じ

因果Boosting法のアルゴリズム

Require: $\nu, B \triangleright \nu$: 基底関数の重み, *B*: Boosting回数 Initialize: $\tilde{y}_n = y_n$, $\hat{\Delta}_{CB}^{(0)} = 0 \triangleright \tilde{y}_n$: 疑似残差, $\hat{\Delta}_{CB}^{(0)}$: 因果Boostingによる推定治療効果モデル For *b* in1,..., *B* do $\vec{\tau} - \mathcal{P}$ 集合 { $\tilde{y}_n, \boldsymbol{x}_n, z_n$ }に因果樹木 $\hat{f}_{ctree}(\boldsymbol{x}, z)$ を当てはめる $\tilde{y}_n \leftarrow \nu \hat{f}_{ctree}(\boldsymbol{x}_n, z_n) \triangleright$ 疑似残差の更新 $\hat{\Delta}_{CB}^{(b)}(\boldsymbol{x}) \leftarrow \hat{\Delta}_{CB}^{(b-1)}(\boldsymbol{x}) + \nu f_{ctree}(\boldsymbol{x}, z) \triangleright$ 推定治療効果モデルの更新 End for Return $\hat{\Delta}_{CB}^{(B)}(\boldsymbol{x})$

Powers et al. (2018)は、最適な樹木数を計算するための方法として、hold-out validationを提案している.

Friedman, J.H. (2001). Greedy function approximation: a gradiet boosting machine. Ann. Statist., 29(5), 1189–1232.

Powers, S., et al. (2018). Some methods for heterogeneous treatment effect estimation in high dimensions. Statistics in Medicine, 37(11), 1767-1787.

因果樹木法を基本学習器とするアンサンブル学習法:因果Boosting法

hold-out validation

 S^{tr} を用いてBoosting樹木を構築

を選定する。

Causal Forest法と異なり, Causal BoostingではBase learner (因果樹木)が増えるほど過学習に陥る可能性がある. ただし,通常の交差確認法を用いることができないため, Powers et al.(2018)では,新たに, hold-out validation法 を提案している.



Powers, S., et al. (2018). Some methods for heterogeneous treatment effect estimation in high dimensions. Statistics in Medicine, 37(11), 1767–1787.

meta-learner

S (Single)-learner(Künzel et al., 2019)

いま, $f_{all}(x, z)$ を共変量x および治療指標z (z = 1:新規治療, z = 0:既存治療)による応答y に対する 回帰モデルとするとき、S-learnerによるHTEは

$$\tau_{\rm SL}(\boldsymbol{x}) = f_{\rm all}(\boldsymbol{x}, z = 1) - f_{\rm all}(\boldsymbol{x}, z = 0)$$

で与えられる.

つまり、S-learnerとは、共変量 xに治療指標 zを加えた回帰モデル $f_{all}(x, z)$ を推定し、そのもとで、HTEを推定するだけの卑近な方法である。

S-learnerに基づく方法としては、

BART (Bayesian Additive Regression Trees; Chipman et al., 2007, 2010)を用いた方法 (Hill, 2011)
 Random Forest (Breiman, 2001)を用いた方法: Virtual Twins(Foster et al., 2011)
 などがある.

Künzel, S.R., et al. (2019) Metalearners for estimating heterogeneous treatment effects using machine learning, PNAS, 116(19), 4156-4165. Hill, J. L. (2011) Bayesian Nonparametric Modeling for Causal Inference, J. Compt. Graph. Statist., 20(1), 217-240. Foster, J.C. et al. (2011) Subgroup identification from randomized clinical trial data, Statist. Med., 30(24), 2867-80.

T (Two) -learner(Künzel et al., 2019)

いま、新規治療群z = 1に対する回帰モデル $f_1(\mathbf{x})$ 、既存治療群z = 0に対する回帰モデル $f_0(\mathbf{x})$ が与えられたとき、T-learnerによるHTEは、

$$au_{\mathrm{TL}}(\boldsymbol{x}) = f_1(\boldsymbol{x}) - f_0(\boldsymbol{x})$$

で与えられる.

つまり、T-learnerとは、治療指標*z*によりデータを2群に分けたうえで、各群において、回帰モデル $f_1(x), f_0(x)$ を推定したうえで差をとるだけの卑近な方法である。

Powers et al.(2018)は、T-learnerを条件付き平均(conditional mean regression)と称している.

- S/T-learnerの問題点

■ S/T-learnerはHTEを推定するのみであり、(推定)HTEに対する要因構造が提示できない。

S/T-learnerでは、治療法に対する選択バイアスを調整する手段は考慮されていない。したがって、いわゆる観察研究と同様に取り扱われるRWD(Real World Data)では不適切な結果をもたらす恐れがある。

X-learner(Künzel et al., 2019)

■ 新規治療群 z = 1に対する回帰モデル: $f_1(\boldsymbol{x})$,既存治療群 z = 0に対する回帰モデル: $f_0(\boldsymbol{x})$ ■ 新規治療群 z = 1のアウトカム・共変量: $y^{(1)}$, $\boldsymbol{x}^{(1)}$,既存治療群 z = 0のアウトカム・共変量: $y^{(0)}$, $\boldsymbol{x}^{(0)}$



Künzel, S.R., et al. (2019) Metalearners for estimating heterogeneous treatment effects using machine learning, PNAS, 116(19), 4156-4165.

DA-learner(EconMLに実装)

X-learnerの問題点
 ■ X-learnerでは*f*₁(*x*)および*f*₀(*x*)の推定において選択バイアスが考慮されていない.
 ■ X-learnerの治療効果モデルは、(推定)HTEに対する影響要因を評価できない.

DA-learnerは、上記の問題点を解消する形式で開発されたX-learnerの派生である。



治療効果モデル: $\Delta_{\mathrm{DA}}(\boldsymbol{x}) = g(\tau; \boldsymbol{x})$: τ に対して何らかの回帰モデルを当てはめ

DR-learner(Kennedy, 2022)

T-learnerの問題点

T-learnerでは $f_1(x)$ および $f_0(x)$ の推定において選択バイアスが考慮されていない.

■ T-learnerの治療効果モデルは、(推定)HTEに対する影響要因を評価できない。

DR-learnerは、2重ロバスト推定の枠組みで拡張された方法である.

■ 新規治療群z = 1に対する回帰モデル: $f_1(\boldsymbol{x})$,既存治療群z = 0に対する回帰モデル: $f_0(\boldsymbol{x})$ ■ 傾向スコア $e(\boldsymbol{x})$

とするとき,HTE(疑似アウトカム)における2重ロバスト推定は

$$\mathbf{r} = \left[f_1(\mathbf{x}) + \frac{z \{ y - f_1(\mathbf{x}) \}}{e(\mathbf{x})} \right] - \left[f_0(\mathbf{x}) + \frac{(1-z) \{ y - f_0(\mathbf{x}) \}}{1 - e(\mathbf{x})} \right]$$

で与えられる. DR-learnerにおいても、DA-learnerと同様に疑似アウトカム τ に対して何らかの回帰 モデルを当てはめ($\Delta_{DR}(x) = g(\tau; x)$)、それを治療効果モデルとしている.

なお、DR-learnerの2重ロバスト性等の数理的な議論については、Kennedy (2022)を参照

R-learner

Nie & Wager (2021)は、Robinson変換に基づく新たなmeta-learnerとしてR-learnerを提案している。

モデル: $Y_n = \mu_0(\mathbf{X}_n) + Z_n \Delta(\mathbf{X}_n) + \epsilon_n$ ただし, $\mathbf{E}[\epsilon_n | \mathbf{X}_n, Z_n] = 0$

(unconfounderedness assumption)

このとき、上記モデルの条件付き平均関数(両辺に期待値をとる)とき、 $m(x) = \mathbb{E}\left[Y|\mathbf{X} = \mathbf{x}\right] = \mu_0(X_n) + e(\mathbf{X}_n)\Delta(\mathbf{X}_n)$

である、ここに、 $e(X_n)$ は傾向スコアである、この2つの式の差分を取ることで、 $Y_n - m(X_n) = (Z_n - e(X_n)) \Delta(X_n) + \epsilon_n$

となる(Robinson変換). したがって、R-learnerによる治療効果モデルは、

$$\Delta_{\mathrm{RL}}(\cdot) = \operatorname{argmin}_{\Delta} \left\{ \operatorname{E} \left[\left\{ (Y_n - m(X_n)) - (Z_n - e(X_n)) \Delta(X_n) \right\}^2 \right] \right\}$$

である. Nie & Wager (2020)は,正則化項 $\Lambda_n\{\Delta_{\mathrm{RL}}(\cdot)\}$ を付与したもとで,経験的推定値を

$$\hat{\Delta}_{\mathrm{RL}}(\cdot) = \operatorname{argmin}_{\Delta} \left\{ \sum_{n=1}^{N} \left\{ \frac{Y_n - m(X_n)}{Z_n - e(X_n)} - \Delta_{\mathrm{RL}}(X_n) \right\}^2 + \Lambda \{\Delta_{\mathrm{RL}}(\cdot)\} \right\}$$

のように定義している。

Nie, X. & Wager, S. (2021). Quasioracle estimation of heterogeneous treatment effects. Biometrika, 108(2), 299-319.

Meta-learnerにおける過学習に対する対応: Cross-fitting approach

DA- /DR- /R- learnerでは,

- アウトカムを推定するための回帰モデル (アウトカム・モデル)
- 傾向スコアを推定するための回帰モデル (傾向スコアモデル)
- HTE (疑似アウトカム)を推定するための回帰モデル (治療効果モデル)
- の3種類の回帰モデルを推定している。このとき、過学習によるバイアス(overfitting bias)を解決するための方法として、cross-fitting approachが存在する。



Estimator

single

cross-fit

Jacob, D. (2020) Cross-fitting and averaging for machine learning estimation of heterogeneous treatment effects, ITRG 1792 Discussion Paper, 2020-014.

Meta-learnerに対するシミュレーション評価 (Okasa, 2022)

データ生成 Okasa, G 2022). Meta-learners for estimation of causal effect: finite sample cross-fit performance, arXiv:2201.12692 共変量: $x_{pn} \sim U(0,1)$ P:共変量の数(100個) 治療指標: $z_n \sim \text{Bern}(e(\boldsymbol{x}_n))$ ← 傾向スコア: $e(\boldsymbol{x}) = \alpha (1 + u(f(\boldsymbol{x})), u(a) \sim \text{beta}(2,4)$ α:バランスを調整するパラメータ(モデルで異なる) $f(\boldsymbol{x}) = \sin\left(\pi \cdot x_1 \cdot x_2 \cdot x_3 \cdot x_4\right)$ 誤差: $\epsilon_n \sim N(0,1)$ 既存治療での真のアウトカム: $\mu_0(\boldsymbol{x}) = \sin(\pi \cdot x_1 \cdot x_2) + 2 \cdot \left(x_3 - \frac{1}{2}\right) + x_4 + \frac{1}{2} \cdot x_5$ 新規治療での真のアウトカム: $\mu_1(\boldsymbol{x}) = \mu_0(\boldsymbol{x}) + \tau(\boldsymbol{x})$

<u>HTE・治療指標のバランスのパターン</u>

Simulation 1: $\tau(\mathbf{x}) = 0$, $\alpha = 1/4$ 新規治療群の割合: 50% Simulation 3: $\tau(\mathbf{x}) = 1$, $\alpha = 1/12$ 新規治療群の割合: 15% $\tau(\mathbf{x}) = 1 + \frac{1}{2}x_1 + \frac{1}{2}x_2$, $\alpha = 1/8$ 新規治療群の割合: 25% Simulation 2: $\tau(\boldsymbol{x}) = \mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x})$ 」。 uoれ 中国 加加 4 新規治療群の割合: 50%

Simulation 4: $\tau(\mathbf{x}) = 1 + I(x_1 > 0.5), \ \alpha = 1/8$ 新規治療群の割合: 25%

Simulation 6:

$$\tau(\mathbf{x}) = 1 + \frac{1}{P'} \sum_{p=1}^{P'} \left(\frac{1}{1 + \exp(-12(x_p - 0.5))} \right) \quad \alpha = 1/8$$

新規治療群の割合: 25%

評価尺度 C: シミュレーション回数(10,000回)

RMSE

$$RMSE(\hat{\tau}(\boldsymbol{x}_n)) = \sqrt{\frac{1}{C}\sum_{c=1}^{C} \left(\tau(\boldsymbol{x}_n) - \hat{\Delta}_c(\boldsymbol{x}_n)\right)^2}$$

ドイアス
$$|BIAS(\hat{\tau}(\boldsymbol{x}_n))| = \frac{1}{C} \sum_{c=1}^{C} |\tau(\boldsymbol{x}_n) - \hat{\Delta}_c(\boldsymbol{x}_n)|$$



推定HTEの標準偏差

$$SD(\hat{\tau}(\boldsymbol{x}_n)) = \sqrt{\frac{1}{C}\sum_{c=1}^{C} \left(\hat{\Delta}_c(\boldsymbol{x}_n) - \frac{1}{C}\sum_{c'=1}^{C} \hat{\Delta}_{c'}(\boldsymbol{x}_n)\right)^2}$$

Jarque-Bera統計量 尖度, 歪度に関する統計量 $JB(\hat{\tau}(x_n)) = \frac{C}{6} \left(\frac{S(\hat{\Delta}(x_n))^2}{\frac{2}{\mathbb{E}\mathfrak{E}}} + \frac{1}{4} \frac{K(\hat{\Delta}(x_n) - 3)^2}{\frac{2}{\mathbb{E}\mathfrak{E}}} \right)$

- ・標本を分割してのX-learner (X-S)
- ・標本を分割してのDR-learner (DR-S)
- ・標本を分割してのR-learner (DR-S)
- ・Cross fittingを伴うX-learner (X-C)
- ・Cross fittingを伴うDR-learner (DR-C)
- Cross fittingを伴うR-learner (DR-C)
- ・初期分岐に治療指標を用いるS-learner (S-W)

手法

S-learner (S)

• T-learner (T)

• X-learner (X-F)

• DR-learner (DR-F)

• R-learner (DR-F)

Note: すべての回帰モデルにはRandomForest法(樹木数100)を用いる.

Okasa, G 2022). Meta-learners for estimation of causal effect: finite sample cross-fit performance, arXiv:2201.12692

Meta-learnerに対するシミュレーション評価 (Okasa, 2022)

Model.6 での平均

	\overline{RMSE}				BIAS				\overline{SD}						\overline{JB}			
標本サイ	ズ 500	2000	8000	32000	500	2000	8000	32000	50	0	2000	8000	32000	500	2000	8000	32000	
S	0.878	0.749	0.651	0.570	0.867	0.739	0.641	0.560	0.10	8	0.096	0.091	0.088	7.140	2.888	2.173	1.936	
S-W	0.765	0.634	0.533	0.462	0.717	0.602	0.508	0.443	0.26	51	0.190	0.149	0.125	2.086	2.106	2.019	1.931	
Т	0.766	0.634	0.533	0.462	0.719	0.602	0.509	0.442	0.26	60	0.190	0.149	0.125	2.603	2.085	2.016	1.924	
X-F	0.743	0.618	0.517	0.442	0.711	0.597	0.500	0.427	0.20	00	0.141	0.117	0.103	3.490	2.230	2.034	1.857	
X-S	0.820	0.707	0.591	0.499	0.779	0.684	0.574	0.484	0.24	4	0.164	0.125	0.107	5.146	2.680	2.157	1.929	
X-C	0.794	0.693	0.582	0.494	0.770	0.680	0.571	0.482	0.17	'1	0.114	0.097	0.092	3.984	2.322	1.964	1.827	
DR-F	0.817	0.659	0.542	0.463	0.764	0.627	0.518	0.443	0.28	35	0.194	0.149	0.126	141.106	40.528	5.490	2.172	
DR-S	1.053	0.825	0.579	0.445	0.906	0.731	0.521	0.403	0.64	0	0.433	0.281	0.206	567.501	458.729	159.041	36.504	
DR-C	0.880	0.727	0.523	0.409	0.809	0.680	0.490	0.383	0.35	59	0.255	0.179	0.143	52.224	38.216	12.644	3.162	
R-F	0.815	0.679	0.590	0.529	0.746	0.632	0.554	0.499	0.34	6	0.251	0.201	0.172	4.583	3.499	2.225	1.983	
R-S	0.932	0.788	0.659	0.580	0.833	0.721	0.613	0.546	0.46	8	0.333	0.243	0.195	3.959	3.365	2.666	2.028	
R-C	0.825	0.725	0.621	0.554	0.779	0.694	0.597	0.533	0.26	51	0.196	0.155	0.136	2.416	2.184	2.036	1.959	

その他のモデルはSupplementary

- ・標本サイズが小さい場合には、通常のX-learner(X-F)での性能が最良であるが、標本サイズが大きい 場合には、Cross-Fittingを伴うDR-learner(DR-C)が最良だった。
- ・S-learnerにおいて初期分岐に治療指標を含んだ場合、T-learnerと性能がほぼ変わらなかった。
- ・標本サイズが小さい場合にサブサンプルに分ける(因果樹木の形式)では、性能が著しく低下する。

Okasa, G 2022). Meta-learners for estimation of causal effect: finite sample cross-fit performance, arXiv:2201.12692

変換アウトカム/修正共変量法

変換アウトカム法

Tian, L., Alizadeh, A.A., Gentles, A.J., Tibshirani, R.: A simple method for estimating interactions between treatment and a large number of covariates, Journal of the American Statistical Association, 109, 1517-1532, 2014.

ここでは議論を簡略化するため、線形モデルで考える.

いま、治療指標Zを以下のように表す: $Z \in \{1, -1\}$, $\Pr(Z = 1) = \Pr(Z = -1) = \frac{1}{2}$

p=1

交互作用を含んだ次のような線形モデルを考える

$$Y = \beta_0 + \sum_{\substack{p=1\\ p \neq 1}}^{P} \gamma_p X_p + \frac{1}{2} \left\{ \underbrace{\gamma_0 Z}_{\substack{p \neq 1}} + \sum_{\substack{p=1\\ \beta \notin 0}}^{P} \gamma_p X_p Z \right\} + \epsilon$$

共変量の影響
治療の影響
治療×共変量の影響

任意の共変量X=xが与えられたときのHTE $\Delta(x)$ は次のように表すことができる.

 $\Delta(\mathbf{x}) = E(Y^{(1)} - Y^{(-1)} | \mathbf{X} = \mathbf{x}) = E(Y | \mathbf{X} = \mathbf{x}, Z = 1) - E(Y | \mathbf{X} = \mathbf{x}, Z = -1) = \gamma_0 + \sum_{p=1}^{\infty} \gamma_p X_p$

となり、治療効果を予測するための線形モデルになる。また、Zが確率1/2で1,あるいは-1をとることから、

$$\Delta(\boldsymbol{x}) = \mathrm{E}(Y^{(1)} - Y^{(-1)} | \boldsymbol{X} = \boldsymbol{x}) = \mathrm{E}(2YZ | \boldsymbol{X} = \boldsymbol{x}) = \gamma_0 + \sum \gamma_p X_p$$

で表すことができる。このことは、アウトカムを2YZに変換したうえで、通常の重回帰モデルの取り扱いで計算できることを意味する。そのため、この方法は、修正アウトカムモデルと呼ばれる。

変換アウトカム法

観測値 $\{y_n, x_n, z_n\}$ が与えられたとき, Tian et al. (2014)による変換アウトカム法では,

$$\tilde{y}_n = 2z_n y_n$$

を応答として回帰モデルf(x)を推定する:

$$\hat{\Delta}_{\mathrm{TO}}(\boldsymbol{x}) = \operatorname{argmin}_{f} \left[\sum_{n=1}^{N} \left\{ \tilde{y}_{n} - f(\boldsymbol{x}_{n}) \right\}^{2} \right]$$

ただし、Tian et al. (2014)の方法は、確率1/2で割り付けられたRCTのみに焦点が当てられている。

Künzel et al. (2019) は、傾向スコアを用いることで、観察研究に対する変換アウトカム法として、 二つの方法を紹介している。

F-learner: $\tilde{y}_n = \frac{z_n - \hat{e}(\boldsymbol{x}_n)}{\hat{e}(\boldsymbol{x}_n)(1 - \hat{e}(\boldsymbol{x}_n))} \cdot y_n$, $z_n \in \{1, 0\}$, $\hat{e}(\boldsymbol{x}_n)$:推定傾向スコア U-learner: $\tilde{y}_n = \frac{y_n - \hat{f}_{all}(\boldsymbol{x}_n)}{z - \hat{e}(\boldsymbol{x}_n)}$, $\hat{f}_{all}(\boldsymbol{x})$:治療指標を伴わないxのみでの回帰モデル

なお、U-learnerは、R-learnerの縮小推定を伴わない場合と同様である.

Künzel, S.R., et al. (2019) Metalearners for estimating heterogeneous treatment effects using machine learning, PNAS, 116(19), 4156-4165.

修正共変量法

Tian et al.(2014)は,修正アウトカムモデルの拡張系として,2値データ,生存時間データに利用できる修正共変量手法を提案している.

いま、次のような回帰モデルを考える:

$$Y = \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{X} \cdot Z/2 + \epsilon, \quad \boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_P)^{\mathrm{T}}, \ \boldsymbol{X} = (1, X_1, \dots, X_P)^{\mathrm{T}}$$

観測値が与えられたとき、最小2乗損失Lossは以下のように定義される.

$$Loss = \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{x}_n \cdot \frac{z_n}{2} \right)^2$$

緑色の部分に注目すると、応答 *y_n*に対して、±1/2を掛け合わせて最小2乗法を行ったとみることができる。つまり、治療効果は、



つまり,修正共変量法では修正アウトカムモデルの半分 で考えていることと同じである.

なので、最小2乗損失Lossに基づくパラメータ推定は、治療効果を推定する修正アウトカムモデルと同じになる.

$$Loss = \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \gamma^{\mathrm{T}} \boldsymbol{x}_n \cdot \frac{z_n}{2} \right)^2 = \frac{1}{N} \sum_{n=1}^{N} \left\{ \frac{1}{2} \left(2y_n z_n - \gamma^{\mathrm{T}} \boldsymbol{x}_n \right) \right\}^2 = \frac{1}{4} \cdot \frac{1}{N} \sum_{n=1}^{N} \left(2y_n z_n - \gamma^{\mathrm{T}} \boldsymbol{x}_n \right)^2$$

アウトカムを円関する代わりに、共変量に重みづけする方法を修正共変量法という.

修正共変量法

修正共変量法の特徴は、2値応答あるいは生存時間データへの拡張が可能な点である.基底関数をW(x) とするとき、各アウトカムに対する修正共変量法のモデルは下記の通り:

連続応答:
$$f(\boldsymbol{x}) = \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{W}(\boldsymbol{x}) \cdot \frac{Z}{2}$$
 HTE: $\tau(\boldsymbol{x}) = \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{W}(\boldsymbol{x})$
2値応答: $\Pr(Y = 1 | \boldsymbol{X}, Z) = \frac{\exp\left\{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{W}(\boldsymbol{X}) \cdot Z/2\right\}}{1 - \exp\left\{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{W}(\boldsymbol{X}) \cdot Z/2\right\}}$
HTE: $\Pr(Y = 1 | \boldsymbol{X}, Z = 1) - \Pr(Y = 1 | \boldsymbol{X}, Z = -1) = \frac{\exp\left\{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{W}(\boldsymbol{X})/2 - 1\right\}}{\exp\left\{\boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{W}(\boldsymbol{X})/2 + 1\right\}}$

Cox比例ハザード・モデルに基づく生存時間データへの拡張は, Tian et al.(2014)を参照されたい.

Chen et al. (2017)は、傾向スコアを用いることで、修正共変量法を観察データに拡張した方法として、 A-learnerを提案している。

$$LOSS = \frac{1}{N} \sum_{n=1}^{N} \left[y_n - \left\{ \frac{1+z_n}{2} - \hat{e}(\boldsymbol{x}_n) \right\} \boldsymbol{\gamma}^{\mathrm{T}} \boldsymbol{W}(\boldsymbol{x}) \right]$$

ここに、 $\hat{e}(\boldsymbol{x}_n)$ は傾向スコアである.

Chen, S., et al. (2017). A general statistical framework for subgroup identification and comparative treatment scoring, Biometerics, 73(4), 1199-1209.

実際例(2):局所進行性上咽頭癌患者に対する化学放射線療法に対する導入化学療法の上乗せ効果のレスポンダー探索

局所進行性上咽頭癌患者に対する導入化学療法の効果には、個人差があることが無作為化比較試験で明らかになった。この研究では、コホート データを用いるとともに、修正共変量手法を用いて導入化学療法スコアを構築するとともに、その妥当性について検討している。



共有基底条件付き平均回帰

治療効果モデルに対する新たなアプローチ:共有基底条件付き平均回帰

条件付き平均回帰はHTEを推定するのみであり、(推定)HTEに対する要因構造が提示できない。
 条件付き平均回帰では、治療法に対する選択バイアスを調整する手段は考慮されていない。

Powers et al. (2018)は、条件付き平均回帰(T-learner)の欠点を解決するためのフレームワークとして、共有基底条件付き平均回帰を提案している。



共有基底条件付き平均回帰では、(1)基底関数の推定、(2)パラメータ推定、2段階での推定を行われる。

Powers et al. (2018) Some methods for heterogeneous treatment effect estimation in high-dimensions, Statist. Med., 37(11), 1767-1787.

共有基底条件付き平均回帰の諸型: PTO Forest法 (Powers, et al., 2018) 基底関数(複数の樹木)の推定



回帰パラメータ(leafにおけるHTE)の推定
共有基底条件付き平均回帰の諸型:因果MARS法 (Powers, et al., 2018) MARS(Multivariate Adaptive Regression Spline: Friedman, 1996)法とは、打ち切りべき垂基底関数に基づく非

MARS(Multivariate Adaptive Regression Spline; Friedman, 1996)法とは、打ち切りベキ乗基底関数に基づく非線形回帰手法である.

■MARS法のモデル:
$$f_{\text{MARS}}(\mathbf{x}) = \sum_{m=1}^{M} \beta_m \prod_{j=1}^{J_m} H(s_{(j,m)}(x_{p(j,m)} - c_{p(j,m)}))$$

↓ 詳し<は付録参照
打ち切りべき乗関数(括弧内が負値ならば0,正値ならばそのまま)
 $H(s_{(j,m)}(x_{p(j,m)} - c_{p(j,m)})) = \begin{cases} [x_{p(j,m)} - c_{p(j,m)}]_+ & , s_{(j,m)} = 1 \\ [c_{p(j,m)} - x_{p(j,m)}]_+ & , s_{(j,m)} = -1 \end{cases}$

MARS法は, (1) 前進ステップワイズ法, (2)後退ステップワイズ法(モデル系列の構成), (3) 最適モデルの選択(通常はGCV(Generalized Cross-Validation)), で構成される. 前進ステップワイズ法におけるモデル構築課程では,以下の制約がある. (1) 交互作用を伴うため,基底関数の積 $\prod_{j=1}^{J_m} H(s_{(j,m)}(x_{p(j,m)} - c_{p(j,m)}))$ のなかに,同じ変数は含まない. (2) 基底関数は $[x_{p(j,m)} - c_{p(j,m)}]_+, [c_{p(j,m)} - x_{p(j,m)}]_+$ の2個を追加する.

(3) 各ステップにおける基底関数の選定および回帰パラメータの推定の損失関数には残差平方和が用いられる.

$$LOSS = \sum_{n=1}^{N} \left\{ y - \left(\sum_{m=1}^{M-2} \beta_m \prod_{j=1}^{J} H(S_{(j,m)}(x_{p(j,m)} - c_{p(j,m)})) + \beta_{M-1} H(x_{p(j,M)} - c_{p(j,M)}) + \beta_M H(c_{p(j,M)} - x_{p(j,M)}) \right) \right\}^2$$

Friedman, J.H.: Multivariate adaptive regression splines, The Annals of Statistics, 19(1), 1-67, 1991.

共有基底条件付き平均回帰の諸型:因果MARS法 (Powers, et al., 2018) 因果MARS法のモデル 新規治療群

$$f_{\text{CMARS}}(\boldsymbol{x}, z) = \sum_{m=1}^{M} \beta_m^{(1)} \prod_{j=1}^{J_m} H(s_{(j,m)}(x_{p(j,m)} - c_{p(j,m)})) I(z=1) + \sum_{m=1}^{M} \beta_m^{(0)} \prod_{j=1}^{J_m} H(s_{(j,m)}(x_{p(j,m)} - c_{p(j,m)})) I(z=0)$$

既存治療群

因果MARS法においても,(1)前進ステップワイズ法,(2)後退ステップワイズ法,の順番でモデル推定を行う.ただし,前進ステップワイズ法において,基底関数の推定と回帰パラメータの推定は,個別に行われる(2段階推定).

因果MARS法のアルゴリズムの略図



共有基底条件付き平均回帰の諸型:因果MARS法 (Powers, et al., 2018)

因果MARS法では、後退ステップワイズ法および交差確認法(cross-validation)によるモデル選択を行うが、Powers et al. (2018)は、代替として、Baggingによるアンサンブルを行うことを推奨している.

数値検証のモデル(シミュレート回数:100回)

 $x_p \sim N(0,1), \quad z \sim Ber(a), \ a = \max[\min\{\sin(\pi x_1 x_2), 0.9\}, 0.1] \quad e \sim N(0,1)$

$$y = \mu(\boldsymbol{x}) + (z - 0.5)\Delta(\boldsymbol{x}) + e$$

 $\mu(\mathbf{x}) = \sin(\pi x_1 x_2) + 0.3x_3 + x_4 + 1, \quad \Delta(\mathbf{x}) = x_1 + \log\{1 + \exp(x_2)\} + 0.2$



シミュレーションによる評価(Powers et al., 2018より引用) Algorithm

1: 共変量の生成: $X_{pn} \sim N(0,1)$, 治療群の生成: $Z_n \sim \text{Bernoulli}(\pi(x))$

2: 主効果(main effect)・局所治療効果(local treatment effect)による応答の生成

- main effect function $\mu(x)$ local treatment effect function $\tau(x)$

 く 関数は表を参照
- Generate: $N(\mu(X_n) + (Z_n 1/2)\tau(X_n), \sigma_Y^2)$ < 各群の平均アウトカムは $\mu_1(x) = \mu(x) + \tau(x)/2, \ \mu_1(x) = \mu(x) \tau(x)/2$ く分散 σ_Y^2 (ノイズ)は表を参照

Functions

$$\begin{split} f_1(x) &= 0, \qquad f_2(x) = 5 \cdot I(x_1 > 1) - 5, \qquad f_3(x) = 2x_1 - 4, \\ f_4(x) &= x_2 x_4 x_6 + 2x_2 x_4 (1 - x_6) + 3x_2 (1 - x_4) x_6 + 4x_2 (1 - x_4) (1 - x_6) \\ &+ 5(1 - x_2) x_4 x_6 + 6(1 - x_2) x_4 (1 - x_6) + 7(1 - x_4) (1 - x_4) x_6 \\ &+ 8(1 - x_2) (1 - x_4) (1 - x_6), \\ f_5(x) &= x_1 + x_3 + x_5 + x_7 + x_8 + x_9 - 2, \\ f_6(x) &= 4I(x_1 > 1)I(x_3 > 0) + 4I(x_5 > 1)I(x_7 > 0) + 2x_8 x_9, \\ f_7(x) &= \frac{1}{2} \Big(x_1^2 + x_2 + x_3^2 + x_4 + x_5^2 + x_6 + x_7^2 + x_8 + x_9^2 - 11 \Big), \\ f_8(x) &= \frac{1}{\sqrt{2}} (f_4(x) + f_5(x)) \end{split}$$

Simulation Senario

	Scenarios							
	1, 9	2, 10	3, 11	4, 12	5, 13	6, 14	7, 15	8, 16
п	200	200	300	300	400	400	1000	1000
р	400	400	300	300	200	200	100	100
$\mu(x)$	$f_8(x)$	$f_5(x)$	$f_4(x)$	$f_7(x)$	$f_3(x)$	$f_1(x)$	$f_2(x)$	$f_6(x)$
$\tau(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$	$f_5(x)$	$f_6(x)$	$f_7(x)$	$f_8(x)$
$\sigma_{\scriptscriptstyle Y}^2$	1	1/4	1	1/4	1	1	4	4

シミュレーションの考察

RCTを想定した場合

- Scenario 4, 5, 7, 8ではBoosted因果MARSが最良の結果を示した(Scenario 4においてもPTO Forestと同程度)
- Scenario 2, 3, 4, 6ではPTO Forestが最良の結果を示した.

Null: Null model, TO: Transformed outcome forest, DB: Different-basis forest CF: Causal forest, PTO0: PTO forest, CB0: 因果Boosting, BCM0: Boosted因果MARS

Conorio	-	公成計田	順位			
Senario	土刈禾	<u> </u>	1位	2位	3位	
1	線形	NULL	CB0	NULL	CF	
2	線形	指示関数	РТО0	CB0	BCM0	
3	非線形	線形	РТО0	ТО	BCM0/CF	
4	非線形	非線形	PTO0/BCM0	-	CB0	
5	線形	線形	BCM0	CB0	ΡΤΟΟ	
6	NULL	指示+交互作用	ΡΤΟΟ	CB0	BCM0	
7	指示関数	非線形	BCM0	PTO0	ТО	
8	非線形	指示+交互作用	BCM0	PTO0 /CB0	-	

シミュレーションの考察

■ 観察研究を想定した場合

・ 因果Boosting(CB1)が8個の数値検証のうち、6個において最良の結果を示した。

Null: Null model, CF: Causal forest, PTO0: PTO forest, CB0: 因果Boosting(PS調整なし), CB1: 因果Boosting(PS調整あり), BCM0: Boosted因果MARS(PS調整なし), Boosted BCM1(PS調整あり)

Senario	之 が田	公康六十日	順位				
	土刈未	<u> </u>	1位	2位	3位		
9	線形	NULL	CB1	BCM1	CF		
10	線形	指示関数	CB1	BCM1	CB0		
11	非線形	線形	CB1	BCM1	CF		
12	非線形	非線形	CB1	CB0	PTO/CF		
13	線形	線形	BCM0	BCM1	CB1		
14	NULL	指示+交互作用	PTO	CB1	CB0		
15	指示関数	非線形	CF/PTO	-	BCM1/BCM0		
16	非線形	指示+交互作用	CB1	CF/BCM1	-		

いずれのScenarioにおいても因果Forest(Causal forest)が惨憺たる結果になっているが、Generalized Forest法において、その性能が大幅に向上しているので、そちらとの比較になると、その限りではないと思われる。

共有基底条件付き平均回帰の諸型:因果RuleFit法 (Wan, et al., 2023) <u>RuleFit法(Friedman & Popescu, 2008)</u>



Friedman, J.H. & Popescu, B.E. (2008). Predictive learning via rule ensemble, Ann. Appl. Stat. 2(3), 916-954.

共有基底条件付き平均回帰の諸型:因果RuleFit法 (Wan, et al., 2023)

因果RuleFit法のモデル

$$f_{\text{CRFit}}(\boldsymbol{x}, z) = \left\{ \sum_{k=1}^{K} \alpha_k^{(1)} r_k(\boldsymbol{x}) + \sum_{p=1}^{P} \beta_p^{(1)} l(x_p) \right\} z + \left\{ \sum_{k=1}^{K} \alpha_k^{(0)} r_k(\boldsymbol{x}) + \sum_{p=1}^{P} \beta_p^{(0)} l(x_p) \right\} (1-z)$$

Step 1: ルール項の探索

ルール項の探索には、変換アウトカムを応答としたGradient Boosting Machine (Friedman, 2001)を用いる.

変換アウトカム:
$$ilde{y}_n = rac{y_n}{\hat{e}(oldsymbol{x}_n)} z_n - rac{y_n}{1-\hat{e}(oldsymbol{x}_n)}(1-z_n)$$

Step 2:回帰パラメータの推定

同一基底関数のもとでの回帰パラメータをグループとしたう えで、Group lasso (Yuan & Lin, 2007)を用いて回帰パラメー タを推定する.

これにより,基底関数自体が刈り込める.

$$\sum_{k=1}^{K} \alpha_k^{(1)} r_k(\boldsymbol{x}) + \sum_{p=1}^{P} \beta_p^{(1)} l(x_p)$$
$$\sum_{k=1}^{K} \alpha_k^{(0)} r_k(\boldsymbol{x}) + \sum_{p=1}^{P} \beta_p^{(0)} l(x_p)$$

Wan, K., Tanioka, K., Shimokawa, T. (2023) Rule ensemble method with adaptive group lasso for heterogeneous treatment effect estimation, Statist. Med. in press.

因果RuleFit法の利点:基本学習器の評価

	No of subjects / subgroup size		Average Treatment Effect	Average Treatment Effect
Rules	Experimental	Control	[95% CI]	[95% CI]
wtkg< 96.84 & cd40>=451.5 & age>=39.5	15 / 31	16 / 31		— 261.43 [135.47, 387.39]
cd40< 452.5 & race< 0.5	286 / 569	283 / 569		81.51 [63.59, 99.44]
race< 0.5 & gender< 0.5	28 / 60	32 / 60	•	176.33 [118.66, 234.00]
cd80>=1006 & wtkg>=80.16 & age< 33.5	19 / 37	18 / 37	•	172.62 [85.43, 259.81]
cd40< 376.5 & cd40>=366	21 / 35	14 / 35	•	211.83 [113.25, 310.42]
cd40>=344 & age< 43 & age>=39.5	16 / 37	21 / 37	•	193.95 [95.68, 292.21]
cd80< 720.5	138 / 280	142 / 280	→	97.33 [72.08, 122.59]
age< 26.5 & cd40>=308	27 / 64	37 / 64	•	149.60 [90.48, 208.72]
cd40>=211.5 & wtkg>=96.84	24 / 45	21 / 45	_	2.33 [-58.06, 62.73]
age>=37.5 & str2< 0.5 & age< 42.5	24 / 55	31 / 55	←	127.77 [52.63, 202.91]
drugs< 0.5 & str2< 0.5	156 / 300	144 / 300	↓	81.18 [53.75, 108.61]
str2< 0.5 & cd40< 384.5	112 / 219	107 / 219	↓	81.04 [50.54, 111.54]
Overal ATE	436 / 855	419 / 855		73.81 [58.49, 89.13]
]
			-1000100200300Favors ZDVFavors ZDV +monotherapycombination the	400 DID rapy

因果RuleFit法では、基本学習器(ルール or 修正線形項)に基づいて、「どのようなルールがHTEに影響を及ぼしているのか」を評価できるため、HTEの予測とサブグループの検討を同時に行うことができる.

Wan, K., Tanioka, K., Shimokawa, T. (2023) Rule ensemble method with adaptive group lasso for heterogeneous treatment effect estimation, Statist. Med. in press.



治療効果モデルに対する最近の展開:time-to-eventデータへの拡張

生存時間解析では、生存時間の推定に関心があるとは限らない(むしろ少ない). ▼ そのため、様々なHTEの定義があり、それらに対する治療効果モデルが提案されている.



- Hu et al. (2021)は、加速モデルのもとで拡張されたBART法(Henderson, 2020)、Random Survival Forest法(Ishiwaran et al., 2008)、 Deep NN(Katzman et al., 2018)をS-learnerのもとで、治療効果モデルに拡張している.
- Xu et al.は, Meta-learnerにおいて, S-/T-は生存率の差, X-/R-生存時間の差として, 生存時間に拡張する方法を提示している. なお, 前者ではRandom Survival Forest, 後者ではIPCW(Inverse Probability Censoring Weighting)を伴うRF法を用いている.
- Tabib & Larocque(2017)は、平均生存時間(KMで推定したうえで右側打ち切りは指数分布で補間)の差としてHTEを定義したうえで RF法によって治療効果モデルを構築している。
- Zhu & Gallego(2020)は、T-learnerの枠組みでSuper Learner(van der Laan et al., 2007)を当てはめている。
- Cui(2022)は、一般化Random Forest法の損失関数にRMSTを用いるとともに、IPCWによる重み付けを伴う形式で生存時間データ へ拡張している。

治療効果モデル(treatment effect model)とサブグループ同定法(subgroup identification)

- 個人的な意見 -



サブグループ同定法(subgroup identification)の臨床応用例

[臨床応用例(1)]

「重度の制御不能な喘息患者に対するTralokinumabの有効性・安全性に関する無作為化比較第III相試験」に対するレスポンダー探索

Gotthlow et al.(2019)は, STRATOS1およびSTRATOS2試験における5つのバイオマーカー(Periostin(喘息と関連がある遺伝子), DPP-4(消化管ホルモンの不活性を行う酵素), Eosinophilis(好酸球, アレルギー反応の制御を行う白血球の一種), FeNO(呼気中ー酸化窒素濃度), IgE(免疫グロブリンE)について, TralokinumabのレスポンダーをSIDES法を用いて検討している.

Lipkovich I, Dmitrienko A, Denne J, EnasG. Subgroup identification based on differential effect search (SIDES): A recursive partitioning method for establishing response to treatment in patient subpopulations. Statistics in Medicine 2011, 30:2601-2621.

Gottlow, M. et al: Application of structured statistical analyses to identify a biomarker predictive of enhanced tralokinumab efficacy in phase III clinical trials for severe, uncontrolled asthma, BMC Pulmonary Medicine, 19(129), 2019.

[臨床応用例(2)]

「小細胞肺癌に対するNavitoclaxの単群第II相試験」のデータを用いたバイオマーカーの最適カットオフ値の探索 Navitoclaxは、小細胞肺癌に対するBcl-2およびBcl-xLを選択的に阻害する分子標的薬である.この試験では、Navitoclax投与によ る予後に影響を及ぼすバイオマーカーを評価することにある.評価対象となるバイオマーカーは、CYFRA21-1、NSE、CTCであ る.このとき、BATTing法を利用して、これらの最適なカットオフ値(応答は奏効割合)を探索している.

Devanarayan, V., Cummins, D.J., Tanzer, L.R.,, et al.: Application of GAM and tree models for assessing the role of drug resistance proteins in leukemia chemotherapy. Proceedings of the American Statistical Association Joint Statistical Meetings, 1999.

Rudin, C.M. et al.: Phase II study of single-agent Navitoclax (ABT-263) and biomarker correlates in patients with relapsed small cell lunge cancer, Clinical Cancer Research, 18(11), 3163-3169, 2012.

治療効果モデルの活用事例

Zhu, J. Gallego, B. (2020). Targeted estimation of heterogeneous treatment effect in observational survival analysis, Journal o Biomedical Informatics, 103474.

Zhu & Gallego (2020)は、サブグループ同定のためにT-learnerにより推定されたHTEに対して、BART法を当てはめることで変数重要度(VI)を推定している。



結局はサブグループ同定が目的であり、治療効果モデルは、共変量のスクリーニング に応用されるに過ぎない。

事例検討:膵体尾部猟奇疾患に対する切除術に対する無作為化比較第Ⅲ相試験

Yamada et al. (2021)は、膵体尾部領域疾患に対する切除術として、膵静脈剥離個別切除(Active: 個別 切除群)と膵静脈同地切断(Control:同時切除群)の多機関共同無作為化比較第III相試験を実施している.

アウトカム:出血量

trial, JAMA Surgery, 156(5), 418-428.

共変量:年齡,性別,BMI, PS(Performance Status), 糖尿病の有無、ステロイド服用の有無、抗凝固薬 服薬の有無、心血管障害の有無、放射線療法の有 無、化学療法の有無、膵切除予定部位の厚み、主 Amount of bleeding 膵管経, 膵切除ラインの位置, 血液・生化学検査 (WBC, Hb, PLT, T-Bil, Cre, Alb, Ami, HbA1c)

ここでは、因果Boosting法での結果について解釈する.



3000

2500

2000

500

1000

出血量の比較

8

0

0

8

8



(lower risk group)では、同時切断群(Control)のほうが出血量が多い。

Estimated heterogeneous treatment effect

pts





治療効果モデルに対する適応上の留意点:再現性の問題

Raghavan et al.(2022)は、2型糖尿病に対する血糖コントロールの有効性に関する2つの無作為化比較試験(ACCORD study, VADT Study)の主要エンドポイント(全死因死亡率)に対して因果Forest法を適用した。

因果Forest法の2試験に対する変数重要度



2試験のKendallの順位相関係数は0.717であり、決して高い数値ではないことから、同じ試験デザインであっても、HTEを再現することは困難であることを指摘している。

(限定的な)解決策として, Raghavan et al.(2022)は, 複数の試験 データを併合することで,より再現性の高い結果をもたらすことが できると結論付けている.

Raghavan, S. et al. (2022) Generalizability of heterogeneous treatment effects based on causal forests applied to two randomized clinical trials of intensive glycemic control, Ann. Epidemiol., 65, 101-108.

おわりに

- ・本講演では、異質性治療効果(HTE)に対する統計モデルとして、治療効果モデルを紹介した。
- ・治療効果モデルは、統計学・機械学習において多数の提案がなされているが、ここでは、
 (1) 樹木に基づくモデル

(2) Meta-learner

(3) 変換アウトカム/修正共変量法

(4) 共有基底条件付き平均回帰

に分けて紹介した.

- 治療効果モデルの多くがアンサンブル学習法を用いることからモデルがブラックボックス化される.このことは、治療効果モデルを臨床応用するうえにおいて、多大な欠点となり得ることを指摘した.
- •HTEを用いる一つの適応場面が探索的なサブグループ解析であることから、Virtual TWINSの流儀に倣い、事後的にCART樹木を用いて解釈する方法を紹介した。
- •結局のところ,HTEが推定できる統計モデルを今後,どのように活用するのかについては、今後の課題である.



Thank you for your kind attention

toshibow2000@gmail.com

shimokaw@wakayama-med.ac.jp

•

Supplementary

付録1:多変量適応型回帰スプライン法

多変量適応的回帰スプライン法 (MARS; Multivariate Adaptive Regression Splines)

Friedman (1991)

- MARS法では、ステップ関数による基底関数を打ち切りベキ乗基底関数に変更することで、CART法の問題を解決している。
- MARS法では、データの層別と回帰分析を同時に遂行することが可能であり、かつ非線形構造や交互作用も捉えることができる。
- MARS法のモデル構築アルゴリズムは、前進ステップワイズ過程、後退ステップワイズ過程、モデル系列からの 最適モデルの選択によって構成される.



Friedman, J.H.: Multivariate adaptive regression splines, The Annals of Statistics, 19(1), 1-67, 1991.



前進ステップワイズ法



予め設定された基底関数の最大数 M_{max} ,および交互作用の最大次数 J_{max} まで、過度に大きなモデルを 生成する.

このとき、以下の点に留意する:

(1) CARTと異なり、分割(交互作用項の追加)を行った場合でも、親ふし(追加前の基底関数)を削除しない.



(2) 基底関数のなかに同一の変数を含めない. (3) $[x_{p(j,m)} - c_{p(j,m)}]_+, [c_{p(j,m)} - x_{p(j,m)}]_+$ を各ステップで一度に追加する.

パラメータ推定(最小2乗法)

$$\hat{\beta} = \arg \min_{\beta} \sum_{n=1}^{N} (\hat{f}_{MARS}^{(M,J_m)}(x) - y)^2$$
 基底関数の追加 → 残差平方和が最小になるように基底関数を追加する.

前進ステップワイズ法のイメージ





[Step.0]
$$\hat{y} = \hat{\beta}_0$$

[Step.1]
$$\hat{y} = [Step.0] + \hat{\beta}_1 [-(x_{p(1,1)} - c_{p(1,1)})]_+ + \hat{\beta}_2 [+(x_{p(1,1)} - c_{p(1,1)})]_+$$

[Step.2]
$$\hat{y} = [\text{Step.1}] + \hat{\beta}_3 [+(x_{p(1,1)} - c_{p(1,1)})]_+ [-(x_{p(2,2)} - c_{p(2,2)})]_+ + \hat{\beta}_4 [+(x_{p(1,1)} - c_{p(1,1)})]_+ [+(x_{p(2,2)} - c_{p(2,2)})]_+$$

[Step.3]
$$\hat{y} = [\text{Step.2}] + \hat{\beta}_5 [-(x_{p(1,3)} - c_{p(1,1)})]_+ + \hat{\beta}_6 [+(x_{p(1,3)} - c_{p(1,1)})]_+$$

[Step.4]
$$\hat{y} = [\text{Step.3}] + \hat{\beta}_7 [+(x_{p(1,1)} - c_{p(1,1)})]_+ [-(x_{p(2,2)} - c_{p(2,2)})]_+ [-(x_{p(3,4)} - c_{p(3,4)})]_+ + \hat{\beta}_8 [+(x_{p(1,1)} - c_{p(1,1)})]_+ [-(x_{p(2,2)} - c_{p(2,2)})]_+ [+(x_{p(3,4)} - c_{p(3,4)})]_+$$

後退ステップワイズ法



前進ステップワイズ法で得られた、大きなモデルを後退すてっぷワイズ法により、切片項まで逐次に削除する. これにより、最大モデルから切片のみ(Nullモデル)までのモデル系列が得られる.

- (1) CARTの刈り込み過程と異なり、親ふし(基底関数)が削除されても、子ふしは削除しない.
- (2) 前進ステップワイズ法では、2個の打ち切りベキ乗基底関数を追加したが、後退ステップワイズ法では、1 個のみが削除される.



後退ステップワイズ法のイメージ





最適モデルの選定



予測確度に焦点を当てたモデル選択 (OriginalのMARSにおいて採用) 一般化交差確認法(GCV: Generalized Cross Validation)の修正版(GCV基準)



情報量規準に基づくモデル選択 (Survival MARS (LeBlanc & Crowley, 1999)において採用)
 赤池の情報量規準(AIC) 等 (重回帰のモデル選択と同様)
 AIC(M) = -2l(M) + 2αm
 M: モデルの番号, l(M): M番目のモデルの対数尤度
 m: 基底関数の個数
 α: 罰則項(通常はα=3)

■ その他のアプローチ

- ・縮小推定の利用(後退ステップワイズ法以降の代替に縮小推定を使用): Taylan & Weber (2018)
- ・Bagging-MARS法 (前進ステップワイズ法のみでBaggingによるアンサンブル)

LeBlanc, M. and Crowley, J.: Adaptive regression splines in the Cox model, Biometrics, 55, 204-213, 1999. Taylor, P. and Weber, G.W.: CG-Lasso Estimator for Multivariate Adaptive Regression Spline, Mathematical Methods in Engineering, 24, 121-136, 2018.

GCV基準による最適モデルの選定のイメージ



Supplementary

付録2:アンサンブル樹木

樹木に基づくアンサンブル学習法



樹木に基づくアンサンブル学習法の基本的アルゴリズム(Friedman & Popescu, 2008)

Input:学習標本:
$$\{y_i x_i\}_1^N$$

アンサンブル回数: B
初期モデル: $F_0(x) = \arg \min_{\alpha} \sum_{i=1}^N L(y_i, \alpha)$
For $b=1$ to B do
 $p_b = \arg \min_{p} \sum_{i \in S_b(\eta)} L\{y_i, F_{b-1}(x_i) + f(x_i, p)\}$
 $f_b(x) = f(x_i, p_b)$
 $F_b(x) = F_{b-1} + v \cdot f_b(x)$
End
Output: $ensumble = \{f_b(x)\}_1^B$

s_b(η) :η≦Nの部分標本 (MART:非復元抽出, Bagging/RandomForest:復元抽出)

p_b:基本学習器(樹木)のパラメータ(RandomForestでは、樹木の分 岐選択変数集合をランダム化に選択、刈り込みなし)

v:結合パラメータ(MART:任意, Bagging/RandomForest:0)

umble
MART
$$:F_{bst}(\mathbf{x}) = \sum_{b=1}^{B} \omega_b f_b(\mathbf{x})$$

Bagging/RandomForest $: F_{bts}(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} f_b(\mathbf{x})$



ens

Friedman, J.H.and Popescu, B.E.: Predictive learning via rule ensemble, Annals of Applied Statistics, 2(3), 916-954, 2008.

Boosting接近法に基づくアンサンブル学習法



反復更新によりモデルを精緻化:多重化法型回帰樹木法(Friedman, 2001)などが含まれる

勾配Boosting樹木のアルゴリズム



勾配Boosting樹木の例示

インドの小児の栄養失調に関する人口保険調査データ(Fenske et al., 2009) 応答:栄養不良スコア, 説明変数:小児のBMI,小児の月年齢,母親のBMI



縮小パラメータv=1.00 (縮小推定なし)の場合、アンサンブル回数の増加とともに損失関数が減少している(過 学習を起こしている)。

縮小パラメータv=0.01の場合、早期に樹木が上昇傾向を示している。つまり、縮小パラメータを小さくしすぎると、殆どアンサンブルが終了しないままになる。

Fenske, N., Kneib, T. and Hothorn, T. : Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression, Technical Report 052, University of Munich.
確率勾配Boosting樹木のアルゴリズム

Friedman(1999)は、部分標本を用いることで、勾配Boostingに比べて計算負荷が小さく、かつ、予測確度に優れた モデルを構築できることを指摘している。この方法は、確率勾配Boosting樹木と呼ばれる。



数値例での結果

Boston住宅価格(Harrinson and Rubinfeld(1978), MASSパッケージ)データにおいて、テスト標本(N=200)を ランダムに抽出し(非復元抽出)、残りのデータを学習標本に用いて、1,000回シミュレーションしたときの結果



Bootstrap接近法に基づくアンサンブル学習法



Bootstrap標本を用いて、個別に樹木を構成したうえで統合する. Bagging(Breiman, 1996)およびRandom Forest(Breiman, 2001)がこれに該当する.

Bagging法のアルゴリズム (回帰の場合)

Require ブートストラップ回数*B* For b=1 to *B* do (a) ブートストラップ標本を生成($L_{boot}^{(b)} = \{y_{u(n)}, x_{u(n)}\}_{n=1}^{N}, b=1,...B$) (b) ブートストラップ標本 $L_{boot}^{(b)}$ にCART樹木を当てはめる($T_{b}(x)$)

End For Ensemble: $\hat{F}_{bag}(x) = \frac{1}{B} \sum_{b=1}^{b} T_b(x)$

分類の場合には, B個の樹木における多数決ルールに基づいてクラス予測値が決定される.

単一樹木より、なぜブートストラップ標本を用いた複数樹木のほうが予測確度に優れるのか? いま、データ集合 $\mathcal{L} = \{y_n, x_n\}$ によるCART樹木をT(x)、 \mathcal{L} が従う確率分布をP、 \mathcal{L} の経験分布を P_N とする. また、 \mathcal{L} の ブートストラップ標本を $\mathcal{L}_{boot}^{(b)}$ 、ブートストラップ分布を \widehat{P}_N とするとき、分布収束 $\sqrt{N}(\widehat{P}_N - P_N) \rightarrow \sqrt{N}(P_N - P)$ より、+分 大きな $N(N \rightarrow \infty)$ に対して、 $1\sum_{k=1}^{b} T(x) = \widehat{P}[T(x + C^{(k)})] \rightarrow \mathbb{P}[T(x + C)]$ (1)

$$\frac{1}{B}\sum_{b=1}^{\infty}T_b(x) = \hat{E}\left[T(x_n; \mathcal{L}_{boot}^{(b)})\right] \to E\left[T(x_n; \mathcal{L})\right]$$
(1)

である。また、 $\mathbf{E}_{\mathcal{L}}\left[T(x_n;\mathcal{L})^2\right] \ge \mathbf{E}_{\mathcal{L}}\left[T(x_n;\mathcal{L})\right]^2$ であることから、

$$\mathbf{E}_{\mathcal{L}}\left[\left\{y_{n}-T(x_{n};\mathcal{L})\right\}^{2}\right] \geq \left\{y_{n}-\mathbf{E}_{\mathcal{L}}\left[T(x_{n};\mathcal{L})\right]\right\}^{2}$$
(2)

となる.式(1),式(2)より、ブートストラップ法を用いて多数の樹木を生成し、 $\hat{E}\left[T(x_n; \mathcal{L}^{(b)}_{boot})\right]$ で予測すれば、単一樹木での 残差平方和に比べて小さくなることがわかる.詳しい証明はBreiman(1996)を参照.

数値例での結果

Boston住宅価格(Harrinson and Rubinfeld(1978), MASSパッケージ)データにおいて、テスト標本(N=200) をランダムに抽出し(非復元抽出)、残りのデータを学習標本に用いて、1,000回シミュレーションしたときの CART樹木とBagging樹木の結果



Harrison, D. and Rubinfeld, D.L. : Hedonic prices and the demand for clean air. J. Environ. Economics and Management 5, 81–102, 1978.

RandomForest法

Bagging法とRandomForest法の違い

Bagging法

RandomForest法



RandomForestでは,

- ・各樹木(基本学習器)の生成は、ブートストラップ標本が用いられる (Baggingと同じ).
- ・CARTの分岐過程のみが行われる(過剰適合するほど大きな樹木を構築する)
- 各樹木の分岐選択では、P 個の変数のなかからランダムに選択された $\tilde{P}(\tilde{P} < P)$ が用いられる.

RandomForest法のアルゴリズム (回帰の場合)



なお、上記のアルゴリズムは、Breiman(2001)では、RandomForest-RIと呼ばれている。その他に、RandomForest-RCという方法がある。この方法では、 \tilde{P} 個の変数の線形結合に基づいてカットオフ値が選定される。

 $ilde{P}$ の選定において,Hastie et al.(2014)は,回帰問題の場合には, $ilde{P}=\sqrt{P}$,分類問題の場合には $ilde{P}=P/3$ を推奨している.

過剰適合するほど樹木を成長させること、並びに分岐変数選択にランダム性を導入することで予測確度が向上することの 証明は、Breiman(2001)、杉本他(2005)を参照されたい.

> 杉本知之・下川敏雄・後藤昌司:樹木構造接近法と最新の発展,計算機統計学,123-164,2005. Breiman, L.: Random Forests, Machine Learning, 45(1), 5-32, 2001.

数値例での結果

Boston住宅価格(Harrinson and Rubinfeld(1978), MASSパッケージ)データにおいて、テスト標本 (N=200)をランダムに抽出し(非復元抽出)、残りのデータを学習標本に用いて、1,000回シミュレーション したときのCART樹木とBagging, RandomForestの結果



樹木に対する縮小推定

アンサンブル・モデル:
$$F(\mathbf{x}) = \sum_{b=1}^{B} T_b(\mathbf{x}) = \sum_{b=1}^{B} \omega \sum_{t \in \tilde{T}^{(b)}} \beta_t \cdot I(\mathbf{x} \in t)$$

b番目の樹木の終結ふしの集合 $\tilde{T}^{(b)}$

Boosting樹木: $\omega = v$, $\beta_t = Z = Z = V$, $\beta_t = Z = Z = V$, $\beta_t = Z = Z = V$, $\beta_t = \overline{y}(t)$ (CART法と同様に推定された終結ふしの予測値)

青色の部分のみに注目すると、アンサンブル・モデルは、B個の変数の線形モデルと見做すことができる。

そのため、アンサンブル学習を行った後に、アンサンブル樹木を再評価し、不要な樹木を削除する(あるいは各樹木に重み付けを行う)ことが考えられる.

事後重み付けを伴うアンサンブル・モデル:
$$F(\mathbf{x}) = \sum_{b=1}^{B} \beta_{b} T_{b}(\mathbf{x})$$

最も直截的なアプローチは、lassoなどの縮小推定の方法を用いてパラメータ β_b を推定することである。すなわち、 $\{\hat{\beta}_b\} = \operatorname{argmin}_{\beta_b} \left\{ \sum_{n=1}^N \left(y_n - \sum_{b=1}^B \beta_b T_b(x_n) \right)^2 + \lambda \sum_{b=1}^B |\beta_b| \right\}$ である。

前進ステイジワイズ線形回帰(forward stagewise linear regression)

一方で, Hasite et al.(2007)は, 前進ステイジワイズ線形回帰が, 係数軌道が単調に増加するlasso(monotone lasso)と一致することを示している.

Initialize
$$\hat{a}_{b} = 0$$
 ($b = 1, 2, ..., B$)
For $m = 1$ to M do
(a) $\{\beta^{n}, b^{i}\} = \operatorname{argmin}_{\beta, b} \sum_{n=1}^{N} \left(y_{n} - \left(\sum_{k=1}^{B} \hat{a}_{k}T_{k}(x_{n})\right) - \beta T_{b}(x_{n})\right)^{2}$
(b) $\hat{a}_{b} \leftarrow \hat{a}_{b} + \epsilon \cdot \operatorname{sign}(\beta^{i})$
End For
Output: $\hat{F}_{FS}(x) = \sum_{b=1}^{b} \hat{a}_{b}T_{b}(x)$
 \hat{D} Row of $F = S(x) = \sum_{b=1}^{b} \hat{a}_{b}T_{b}(x)$
 \hat{D} Row of $F = S(x) = S($

 $t = \sum_{k} |\alpha_{k}|$

Iteration

Hastie, T. et al. (2009) The Element of Statistical Learning (2nd edition), Springer [杉山将他 (2014). 統計的機械学習の基礎,共立出版].

前テる

軌

過

Ø

縮小推定の有無による予測確度の違い

テスト標本(N=200)をランダムに抽出し(非復元抽出),残りのデータを学習標本に用いて,200回シミュレーションしたときの結果



Supplementary

付録3: Okasa(2022)のシミュレーション結果

		\overline{RMSE}				B	AS			C.	\overline{SD}			\overline{JB}				
		500	2000	8000) 3200	0 500	2000	8000	32000	500) 2000	8000	32000	<u> </u>	500	2000	8000	32000
	S	0.008	0.009	0.013	3 0.01	8 0.005	6 0.006	0.010	0.014	4 0.007	7 0.008	0.012	0.01	3 211	02.232	3656.421	285.128	12.451
	S-W	0.037	0.038	0.049	9 0.05	9 0.023	0.025	0.036	0.04'	7 0.033	3 0.032	0.040	0.04'	7 204	10.604	4973.704	364.501	12.494
	Т	0.225	0.168	0.128	8 0.10	1 0.180	0.135	0.103	0.082	0.200	6 0.149	0.109	0.08	3	2.071	2.280	2.000	1.912
	X-F	0.160	0.111	0.080	0.05	9 0.128	8 0.089	0.065	0.048	8 0.142	2 0.095	0.067	0.04	8	1.689	2.442	2.068	2.002
`	X-S	0.186	0.127	0.09	1 0.06	7 0.149	0.103	0.074	0.05!	5 0.162	2 0.106	6 0.073	0.05	3	1.916	2.152	2.180	2.125
)	X-C	0.152	0.106	0.07	5 0.05	4 0.123	0.087	0.062	0.04!	5 0.120	0.075	0.052	0.03	6	1.293	2.393	2.023	1.982
	DR-F	0.209	0.146	0.10!	5 0.07	9 0.167	0.117	0.084	0.064	4 0.196	6 0.135	0.095	0.07	O	4.677	18.496	17.015	7.978
	DR-S	0.343	0.241	0.170	0.12	2 0.272	0.191	0.135	0.09'	7 0.33	5 0.235	0.166	0.11	9	5.548	15.737	30.661	41.289
	DR-C	0.207	0.147	0.103	3 0.07	4 0.166	0.118	0.082	0.059	0.19^{-1}	1 0.137	0.097	0.07	0	2.606	3.632	9.329	13.029
	R- F	0.261	0.192	0.14	5 0.11	4 0.208	0.153	0.116	0.092	0.249	9 0.180	0.132	0.10	2	5.911	23.800	26.134	14.008
	R-S	0.334	0.243	0.18	1 0.13	7 0.265	5 0.193	0.144	0.109	9 0.32	1 0.232	0.168	0.12	4	3.192	5.140	15.179	15.980
	R-C	0.208	0.156	0.118	8 0.09	2 0.166	6 0.125	0.096	0.075	5 0.18	6 0.135	0.098	0.07	2	2.117	2.586	3.209	3.779
				RM	SE			BI	AS			\overline{SI}	5			\overline{J}	\overline{B}	
			500	2000	8000	32000	500	2000	8000	32000	500	2000	8000	32000	Ę	500 2000	8000	32000
		S	0.527	0.442	0.374	0.326	0.522	0.434	0.366	0.317	0.055	0.068	0.066	0.064	711.5	571 17.960	4.186	2.537
		S-W	0.463	0.357	0.303	0.265	0.431	0.328	0.280	0.246	0.177	0.151	0.120	0.099	268.3	394 2.900	2.520	2.207
		Т	0.434	0.358	0.303	0.265	0.392	0.328	0.280	0.246	0.204	0.154	0.120	0.099	2.2	206 2.464	2.466	2.250
		X-F	0.432	0.377	0.331	0.296	0.407	0.361	0.318	0.285	0.143	0.103	0.083	0.072	1.9	915 2.167	1.936	1.906
)		X-S	0.460	0.411	0.362	0.321	0.432	0.393	0.349	0.310	0.160	0.110	0.085	0.073	2.0	048 2.046	2.156	1.957
/		X-C	0.443	0.400	0.356	0.317	0.424	0.389	0.347	0.309	0.119	0.082	0.068	0.062	1.4	17 2.139	1.955	1.900
		DR-F	0.439	0.366	0.312	0.276	0.399	0.338	0.291	0.259	0.197	0.144	0.113	0.095	3.3	392 2.919	2.104	1.936
		DR-S	0.534	0.439	0.355	0.297	0.461	0.388	0.318	0.270	0.328	0.236	0.173	0.136	5.1	134 8.959	4.011	2.371
		DR-C	0.451	0.388	0.322	0.276	0.413	0.361	0.302	0.259	0.193	0.146	0.114	0.095	2.4	498 2.525	2.158	1.980
		R-F	0.458	0.373	0.307	0.266	0.404	0.333	0.277	0.241	0.251	0.188	0.146	0.122	4.2	201 4.206	2.437	2.021
		R-S	0.529	0.439	0.356	0.298	0.458	0.389	0.319	0.269	0.318	0.236	0.178	0.140	2.9	989 3.550	4.630	3.400
		DC	0.440	0 288	0 299	0.274	0 412	0.261	0.201	0.256	0 1 8 7	0 145	0 116	0.007	9.	105 2.280	9 107	1.040

Model 1 (Null, 等標本サイズ)

Model 2 (無相関, 等標本サイズ)

	\overline{RMSE}				BI	\overline{AS}			\overline{S}	\overline{D}			\overline{JB}					
		500	2000	8000	32000	500	2000	8000	32000	500	2000	8000	32000	500	2000	8000	32000	
	S	0.645	0.475	0.359	0.279	0.638	0.468	0.352	0.272	0.099	0.084	0.072	0.062	2.888	2.667	2.111	1.981	
	S-W	0.246	0.191	0.146	0.111	0.197	0.154	0.119	0.091	0.233	0.163	0.121	0.090	4.611	2.281	2.385	1.993	
	Т	0.244	0.191	0.146	0.111	0.195	0.154	0.119	0.091	0.227	0.164	0.121	0.090	2.806	2.271	2.243	1.964	
Model 3	X-F	0.180	0.123	0.090	0.068	0.144	0.098	0.072	0.054	0.175	0.118	0.085	0.061	3.663	2.552	4.441	2.820	
(一定,	X-S	0.226	0.149	0.110	0.078	0.180	0.119	0.087	0.062	0.219	0.143	0.104	0.072	2.367	2.678	3.058	3.192	
不等標本サイズ) Х-С	0.159	0.102	0.073	0.054	0.127	0.081	0.059	0.043	0.150	0.092	0.064	0.046	6.541	1.969	2.263	1.994	
	DR-F	0.287	0.202	0.146	0.110	0.222	0.158	0.116	0.089	0.279	0.188	0.129	0.093	3060.536	812.294	244.016	38.545	
	DR-S	0.649	0.502	0.334	0.218	0.475	0.365	0.250	0.168	0.645	0.498	0.329	0.213	1276.433	1496.545	795.711	258.858	
	DR-C	0.364	0.290	0.197	0.131	0.282	0.223	0.153	0.104	0.359	0.283	0.189	0.124	112.249	149.500	126.268	43.274	
	R-F	0.441	0.366	0.293	0.243	0.348	0.287	0.232	0.195	0.435	0.354	0.273	0.215	14.590	27.616	19.045	8.171	
	R-S	0.573	0.461	0.366	0.285	0.453	0.363	0.288	0.227	0.570	0.454	0.353	0.264	7.887	12.474	18.638	11.149	
	R-C	0.295	0.262	0.220	0.184	0.235	0.208	0.176	0.150	0.289	0.249	0.199	0.152	2.822	3.570	4.324	3.162	

		\overline{RMSE}				$\overline{ BIAS }$				\overline{S}	\overline{D}			\overline{JB}			
		500	2000	8000	32000	500	2000	8000	32000	500	2000	8000	32000	500	2000	8000	32000
	S	0.834	0.616	0.472	0.370	0.825	0.606	0.462	0.361	0.105	0.090	0.078	0.069	1.935	2.120	2.075	1.951
	S-W	0.443	0.336	0.258	0.206	0.390	0.300	0.233	0.187	0.229	0.162	0.120	0.093	2.575	2.330	2.124	1.957
	Т	0.443	0.335	0.258	0.206	0.390	0.300	0.233	0.187	0.229	0.163	0.120	0.093	2.552	2.278	2.130	1.938
	X-F	0.428	0.329	0.247	0.191	0.394	0.308	0.233	0.180	0.171	0.114	0.083	0.064	3.731	2.312	2.210	1.978
	X-S	0.501	0.399	0.307	0.232	0.456	0.375	0.291	0.220	0.213	0.136	0.097	0.073	6.602	2.762	2.298	2.113
ズ)	X-C	0.477	0.385	0.300	0.227	0.453	0.374	0.292	0.220	0.148	0.091	0.068	0.055	5.617	2.026	2.023	1.898
	DR-F	0.510	0.369	0.275	0.214	0.454	0.334	0.251	0.196	0.249	0.165	0.117	0.088	116.949	156.252	41.933	5.158
	DR-S	0.728	0.537	0.339	0.230	0.591	0.445	0.279	0.190	0.549	0.377	0.247	0.171	497.136	530.045	407.510	97.233
	DR-C	0.565	0.435	0.269	0.182	0.493	0.388	0.236	0.159	0.308	0.215	0.145	0.104	51.595	50.726	42.424	15.770
	R-F	0.537	0.426	0.349	0.293	0.454	0.364	0.303	0.258	0.337	0.250	0.192	0.151	8.764	13.754	7.349	2.839
	R-S	0.653	0.521	0.412	0.338	0.539	0.438	0.352	0.294	0.460	0.332	0.245	0.184	7.025	6.592	7.512	5.029
	R-C	0.524	0.440	0.361	0.303	0.469	0.402	0.333	0.282	0.246	0.185	0.142	0.113	2.700	2.900	2.732	2.318

Model 4 (ステップ関数, 不等標本サイス

Model 5 (線形関数, 不等標本サイズ)

	\overline{RMSE}				$\overline{ BIAS }$				\overline{S}	\overline{D}			\overline{JB}			
	500	2000	8000	32000	500	2000	8000	32000	500	2000	8000	32000	500	2000	8000	32000
S	0.823	0.606	0.461	0.358	0.817	0.599	0.454	0.351	0.101	0.087	0.075	0.066	1.796	2.150	2.057	1.986
S-W	0.305	0.244	0.196	0.164	0.255	0.209	0.170	0.145	0.222	0.159	0.117	0.089	2.457	2.189	2.054	1.957
Т	0.305	0.244	0.196	0.164	0.255	0.209	0.171	0.145	0.222	0.159	0.117	0.089	2.497	2.173	2.026	1.936
X-F	0.237	0.178	0.137	0.109	0.200	0.154	0.120	0.097	0.164	0.110	0.078	0.058	3.329	2.228	2.102	2.022
X-S	0.276	0.210	0.163	0.126	0.230	0.181	0.143	0.112	0.202	0.130	0.092	0.067	6.639	2.811	2.296	2.421
X-C	0.231	0.182	0.144	0.114	0.200	0.165	0.132	0.105	0.139	0.086	0.061	0.046	4.474	2.014	2.004	2.037
DR-F	0.314	0.248	0.203	0.166	0.258	0.212	0.179	0.148	0.237	0.159	0.112	0.084	123.780	364.063	249.515	26.116
DR-S	0.556	0.413	0.298	0.215	0.428	0.322	0.239	0.176	0.515	0.362	0.242	0.167	453.484	685.910	651.725	174.087
DR-C	0.354	0.280	0.217	0.162	0.287	0.232	0.185	0.140	0.289	0.205	0.140	0.099	50.509	61.849	72.770	22.771
R-F	0.392	0.312	0.254	0.223	0.314	0.253	0.211	0.190	0.339	0.253	0.187	0.146	12.888	28.931	17.700	4.568
R-S	0.500	0.388	0.305	0.248	0.398	0.311	0.248	0.206	0.457	0.336	0.246	0.179	9.107	10.173	16.684	12.581
R-C	0.311	0.262	0.222	0.194	0.256	0.219	0.191	0.172	0.241	0.185	0.139	0.104	2.925	3.617	3.874	3.072

Supplementary

付録4: Powers et al.(2018)のシミュレーション結果

シミュレーションの結果: π(x)=1/2 (RCTの場合) 1/2

Null: Null model, TO: Transformed outcome forest, DB: different-bassis forest, CF: Causal forest PTO0: PTO forest, CB0: 因果Boosting, BCM0(因果Boosting)











シミュレーションの結果: π(x)=1/2 (RCTの場合) 2/2

Null: Null model, TO: Transformed outcome forest, DB: different-bassis forest, CF: Causal forest PTO0: PTO forest, CB0: 因果Boosting, BCM0: 因果MARS





シミュレーションの結果: $\pi(x) = \exp(\mu(x) - \tau(x)/2)/(1 + \mu(x) - \tau(x)/2)$ (観察研究の場合) 1/2

Null: Null model, CF: Causal forest, PTO0: PTO forest, CB0: 因果Boosting(PS調整なし), CB1: 因果Boosting(PS調整あり), BCM0: 因果MARS(PS調整なし), BCM1(PS調整あり)



シミュレーションの結果: $\pi(x) = \exp(\mu(x) - \tau(x)/2)/(1 + \mu(x) - \tau(x)/2)$ (観察研究の場合) 2/2

Null: Null model, CF: Causal forest, PTO0: PTO forest, CB0: 因果Boosting(PS調整なし), CB1: 因果Boosting(PS調整あり), BCM0: 因果MARS(PS調整なし), BCM1(PS調整あり)



Supplementary

付録5:Virtual TWINS法

Virtual Twins法 Foster et al.(2011)



無作為化比較試験では、新治療あるいは既存治療のいずれかに割り付けられて、治療が行われる、



つまり、共変量X_iをもつi番目の被験者の治療効果(新治療と既存治療のアウトカムの差)は、いずれかのデータがないため、評価できない。

	新治療が評価	既存(標準)治療が評 価	
新治療群	山底に栖ち加ってモ	= $=$ $=$ $=$ $=$ $=$ $=$ $=$ $=$ $=$	
既存(標準)治療群			$\Delta(\boldsymbol{x}) = f(\boldsymbol{x}, z = 1) - f(\boldsymbol{x}, z = 2)$

さらに、Virtual Twins法では、サブグループ抽出のために $\Delta(x)$ を応答にしたもとでCART法を用いている.

Foster, J.C., Taylor, J.M.G., Ruberg, S.J.: Subgroup identification from randomized clinical trial data, Statistics in Medicine, 30:2867–2880, 2011.

Virtual Twins法

いま、N人の被験者に対して、治療指標 z_n (1:薬A、2:薬B)に対して説明変数 x_n 、応答 y_n が与えられているとする.

[STEP.1(Global Modeling過程)]:治療指標をダミー変数として加える RandomForest(RF)による被験者毎の治療効果の推定

	新治療が評価	既存(標準)治療が 評価			
新治療群	$PondomForost = \mathbb{E} \begin{bmatrix} \hat{f} & (\mathbf{r} & z) \end{bmatrix}$				
既存(標準)治療群		$\mathcal{I}_{\mathrm{RF}}(\boldsymbol{x},\boldsymbol{z})$			

Foster et al.(2011)では、RandomForest法を 用いているが、実際には、交互作用を含む非 線形モデルであれば、何でもよい.

STEP.1によって推定されたモデルを用いて治療効果を推定する: $\hat{\Delta}_n(\mathbf{x}) = \hat{f}_{RF}(\mathbf{x}, z=1) - \hat{f}_{RF}(\mathbf{x}, z=2)$

2値データの場合には、モデルにより推定される帰属確率を計算: $\hat{\Delta}_n(\mathbf{x}) = \hat{P}_{1n} - \hat{P}_{2n}$, $\hat{P}_{kn} = \widehat{\Pr}(y_n = 1 | \mathbf{x}_n, z_n = k)$

[STEP.2(Subgroup Identification過程)]:治療群のダミー変数は除外する 推定治療効果に対するCART樹木を当てはめ(アプローチによって2種類が存在)

VT(R): $\hat{\Delta}_n$ を応答とする場合(回帰樹木の利用).

■ VT(C): $\hat{\Delta}_n$ にしきい値を設定する場合(分類樹木の利用).

Virtual twinsの例示

肺血症に関する無作為化比較試験を想定したシミュレート・データ(aVirtualTwinパッケージのテストデータ)

• Outcome: 生存の有無

・共変量:THERAPY(1:新治療,0:既存治療),最初の治療がfailureしてからの期間(TIMFIRST),年齢(AGE),ベースラインでの血小板数(BILLPLAT),ベースラインSOFAスコア(心血管,血液,肝腎,呼吸器機能のスコア;blSOFA),ベースラインでのクレアチン(BILLCREAT),ベースラインでの臓器不全の個数(ORGANUM),輸血前のAPACHE IIスコア(PRACHE),ベースラインでのGlasgowスケール(BLGCS),ベースラインでの血清IL-6濃度(BLIL6),ベースラインでの活動性スコア(BLADL),ベースラインでの局所ビリルビン(BLLILBILI)



新治療(active)が既存治療(Control)に比べて良好なサブグループ

Supplementary

付録6:Rのパッケージ



	パッケージ	場所	Python
因果樹木法	causalTree	github(susanathey/causalTree)	0
因果 Forest 法	grf	CRAN	O (github)
因果 Boosting	causalLearning	github(saberpowers/causalLearning)	
PTO Forest	causalLearning	github(saberpowers/causalLearning)	
因果MARS	causalLearning	github(saberpowers/causalLearning)	



Meta-learnerは, pythonのライブラリ(causalml, EconML)に多数収録



S-learner, T-learner, X-learner,

R-learner, DR-learner

S-learner, T-learner, X-learner, DA-learner, DR-learner

	パッケージ	場所	Python
R-learner	rlearner	github(xnie/rlearner)	O (causalml)
DR-learner	drlearner	github (christophergandrud/drlearner)	O (causalml) (EconML)

Rでは、包括的なmeta-learnerのパッケージが存在しないようである.

ただし、自作してもさほど大変ではない.