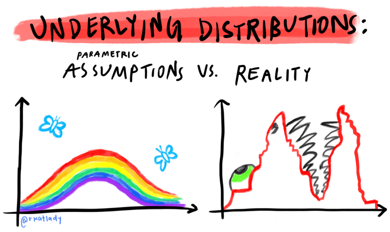
Targeted learning

Roadmap

- Underlying distribution

  $$O_1, \ldots, O_n \sim P_0$$

  - 

- Possible probability distributions
- Query

  $$\Psi : \mathcal{M}^n \to \mathbb{R}, P_0^n \to \Psi(P_0^n) \text{ (query)}$$

  - Example

    $$\Psi(P) = E_P(P(Y = 1|A = 1, W) - P(Y = 1|A = 0, W)))$$

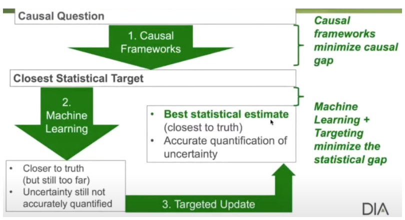
  - Example

    $$\Psi(P) = E(Y_1 - Y_0)$$

- Estimator

  $$f^* : \arg\min_{f \in \mathcal{M}^n} L(\hat{\Psi}_n \equiv f(O^n), \Psi(P_0^n))$$

  - CI
- Causal framework

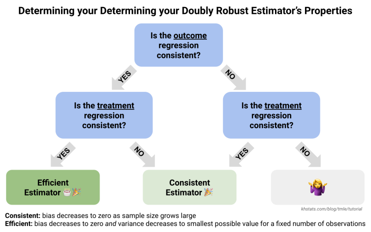
Advantages

- Smooth mapping

  Within our view of the target parameter as a statistical mapping evaluated at the law of the experiment, it is natural to inquire of properties this functional enjoys. For example, we may be interested in asking how the value of $\Psi(P)$ changes as we consider laws that *get nearer to* P in $\mathcal{M}$. If small deviations from P_0 result in large changes in $\Psi(P_0)$, then we might hypothesize that it will be difficult to produce stable estimators of ψ_0. Fortunately, this turns out not to be the case for the mapping Ψ, and so we say that Ψ is a *smooth* statistical mapping.

  - Pathwise differentiable

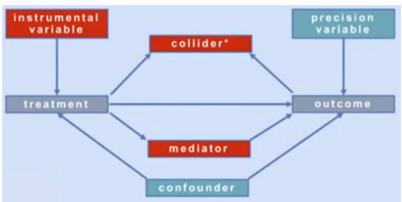
- Double robustness

- Local efficient
  - Efficiency: a measure of how variable (and biased) the estimator is from the "true" parameter
    - related to the estimator AND the sample-size AND the loss function

Practical issues

- Choosing variables

  - Possible confounders

    age, body mass index, gender, history of imprisonment, tobacco and alcohol use, diabetes mellitus, hepatitis C, HIV status, new versus prior treatment case, sputum smear status, cavitary disease

  - Possible precision variables

    number of effective drugs

- Number of CV folds
- Choosing algorithms
  - Recommendations

    "**Best algorithms**": random forests (`SL.ranger`), polymars (`SL.earth`), boosting (`SL.xgboost`), Your favorite GLM.

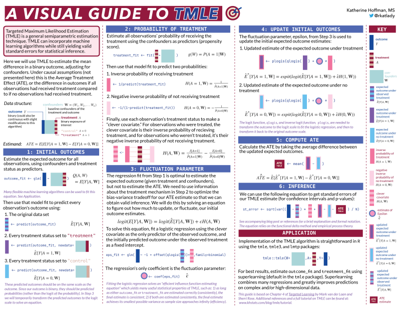
  - Some algorithms may not worth the tuning effort
- Multi-level outcomes
  - Two super learners
- Covariate interactions
  - Include diverse learners
- Reproductibility
  - Random seed
- Stablization issue

- Average through different seeds
- Average on TMLE scale

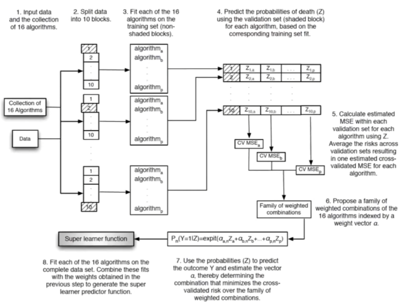
- **Semiparametric theories**

1. Some estimands allow for **asymptotically linear estimation**. This means that estimators can be represented as sample averages (plus a term that converges to zero).
2. The quantities being averaged for asymptotically linear estimators are called **influence functions**. An influence function is a function that quantifies how much influence each observation has on the estimator. For this reason, it is very **useful to characterize the variance of the estimator**. In parametric maximum likelihood estimation, the influence function is related the score function.
3. The **efficient influence function** (EIF) is the influence function that achieves the efficiency bound (think Cramer Rao Lower Bound from parametric maximum likelihood estimation) and **can be used to create efficient estimators.**
4. If we want to **construct an estimator that is efficient**, we can take advantage of the EIF to endow the estimator with useful asymptotic properties.

- **Poster**



- **Super learning**



- **Procedure**
  - Estimate the outcome

$$Q(A,W)=E(Y|A,W)$$

    - Ensemble learning
    - Example

TMLE Algorithm after Step 1

Y	A	Q_A	Q_0	Q_1
1	1	0.8461853	0.6770917	0.8461853
1	0	0.6986440	0.6986440	0.8589257
0	0	0.4932538	0.4932538	0.7188934
1	1	0.8213403	0.6363132	0.8213403
1	0	0.6266258	0.6266258	0.8151742
1	1	0.8578239	0.6966588	0.8578239

  - Estimate probability of treatment

$$g(W)=P(A=1|W)$$

    - Inverse probability of treatment

$$H(1,W)=\frac{1}{g(W)},H(0,W)=\frac{1}{1-g(W)}$$

- Estimate fluctuation parameter

  $$logit(\mathrm{E}[Y|A,\mathbf{W}]) = logit(\hat{\mathrm{E}}[Y|A,\mathbf{W}]) + \epsilon H(A,\mathbf{W})$$

  - Linear regression to solve ε

- Update the initial estimate of expected outcome

  $$\hat{\mathrm{E}}^*[Y|A,\mathbf{W}] = expit(logit(\hat{\mathrm{E}}[Y|A,\mathbf{W}]) + \hat{\epsilon}H(A,\mathbf{W}))$$

- Confidence intervals

  $$\hat{SE} = \sqrt{\frac{var(\hat{IF})}{N}}$$

  - Influence function

    $$\hat{IF} = (Y - \hat{E}^*[Y|A,\mathbf{W}])H(A,\mathbf{W}) + \hat{E}^*[Y|A=1,\mathbf{W}] - \hat{E}^*[Y|A=0,\mathbf{W}] - A\hat{T}E$$

    - The IF tells us how much each observation influences the final estimate

  - Asymptotic properties

    Note that a TMLE estimator is asymptotically normally distributed, so we could bootstrap the entire algorithm to get our standard errors instead.

Assumptions

- Informative censoring
- Assignment of treatment not random

  $$Q(A,W) = E(Y|A,W)$$