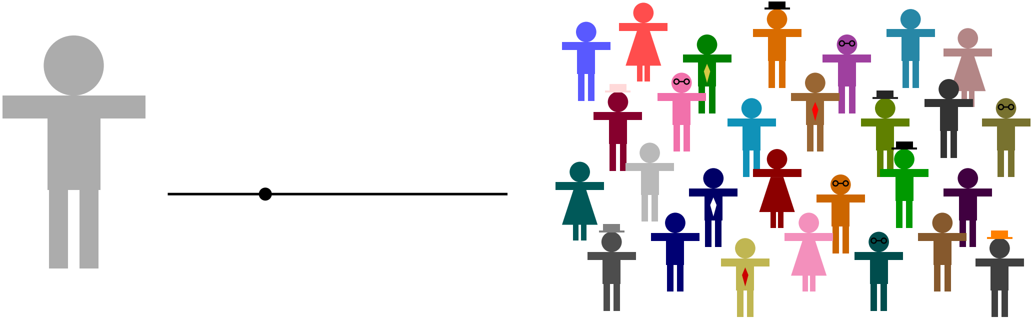
Sublevel set probabilities of the CATE function

Anders Munch
joint work with Thomas Gerds

Section of Biostatistics
University of Copenhagen

April 16, 2026

Reporting heterogeneity – a trade off



ATE CATE

Some notation

$Y^a \in \{0,1\}$ potential outcomes under different treatment levels $a \in \{0,1\}$
$W \in \mathbb{R}^d$ individual baseline characteristics
A actually given treatment
Y actually observed outcome

The CATE function (with respect to W) is

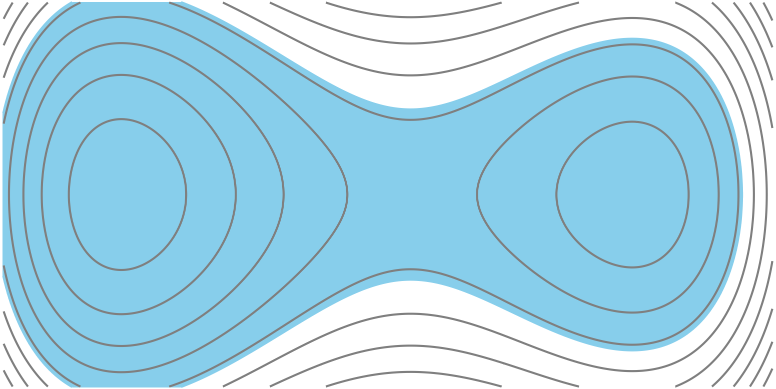
$$w \longmapsto \tau(w) = \mathbb{E}\big[Y^1 - Y^0 \mid W = w\big].$$

It can be identified under standard causal assumptions as

$$\tau(w) = \mathbb{E}[Y \mid A = 1, W = w] - \mathbb{E}[Y \mid A = 0, W = w].$$

Sublevel sets of the CATE function τ [†]

$$\mathcal{W}_\alpha = \{w \in W : \tau(w) \leq \alpha\}, \quad \alpha \in [-1, 1]$$



[†] Bonvini et al. [2023], Reeve et al. [2023], Tsybakov [1997]

Sublevel set size function

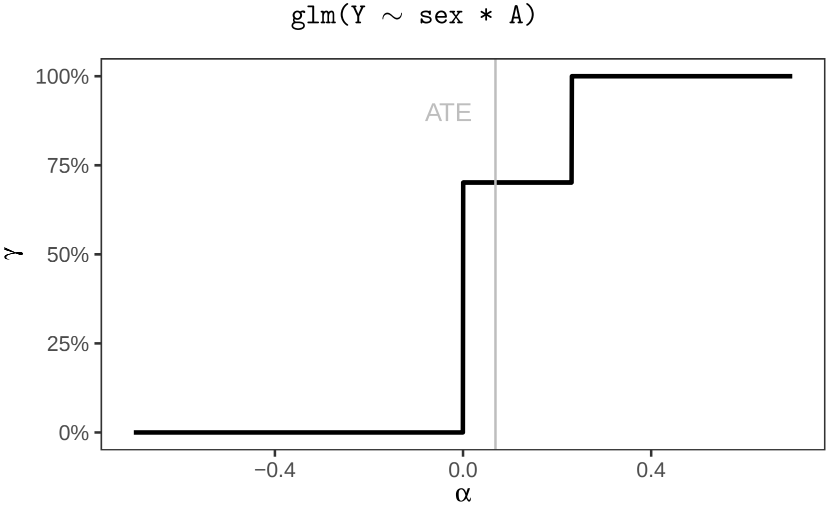
$$\gamma\colon [-1,1] \longrightarrow [0,1], \quad \gamma(\alpha) = P(\mathcal{W}_\alpha) = P(\tau(W) \le \alpha).$$

CDF of the random variable $\tau(W)$[†], i.e., **a univariate, nondecreasing function**.

The value $\gamma(\alpha)$ is the proportion of the population with an expected treatment effect (given available baseline information W) below or equal to α. E.g., $\gamma(0)$ is the proportion of the population for which the treatment is not expected to be beneficial, given available baseline information.

[†] Closely related to "the sorted effects method" [Chernozhukov et al., 2018b]

Illustration of the function γ



Estimation

Estimation of γ is "non-standard"

One possible estimator is the plug-in estimator

$$\hat{\gamma}_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\hat{\tau}_n(W_i) \leq \alpha\},$$

based on some CATE learner $\hat{\tau}_n$. Not easy to analyze its asymptotic behaviour when data-adaptive methods are used.

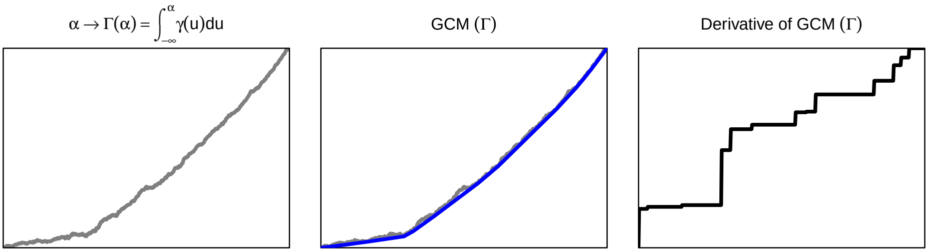
Efficient influence function based methods like targeted/debiased machine learning address this.[†]

Complicated by the fact that the parameter $\gamma(\alpha)$ is not pathwise differentiable.

[†] Hines et al. [2022], Chernozhukov et al. [2018a], van der Laan and Rose [2011]

Generalized Grenander-type estimator†

The primitive of a nondecreasing function is convex. Suggests the estimation strategy:

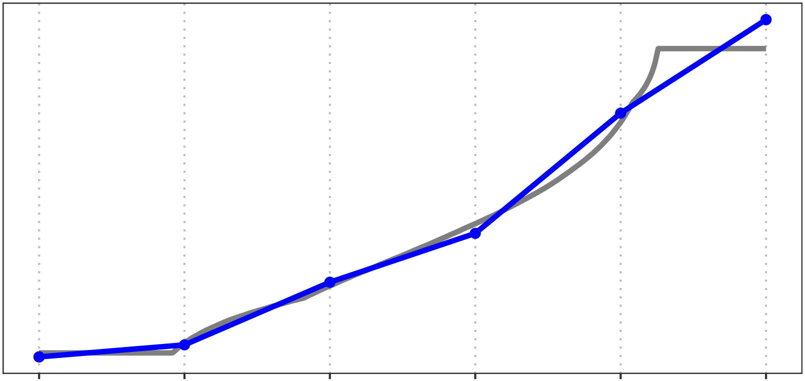


The parameter $\Gamma(\alpha)$ ***is*** pathwise differentiable at values α at which γ is continuous.

$\longrightarrow$ Standard toolbox available for estimating Γ.

† Westling and Carone [2020], van der Vaart and van der Laan [2006], Groeneboom [1983], Grenander [1956]

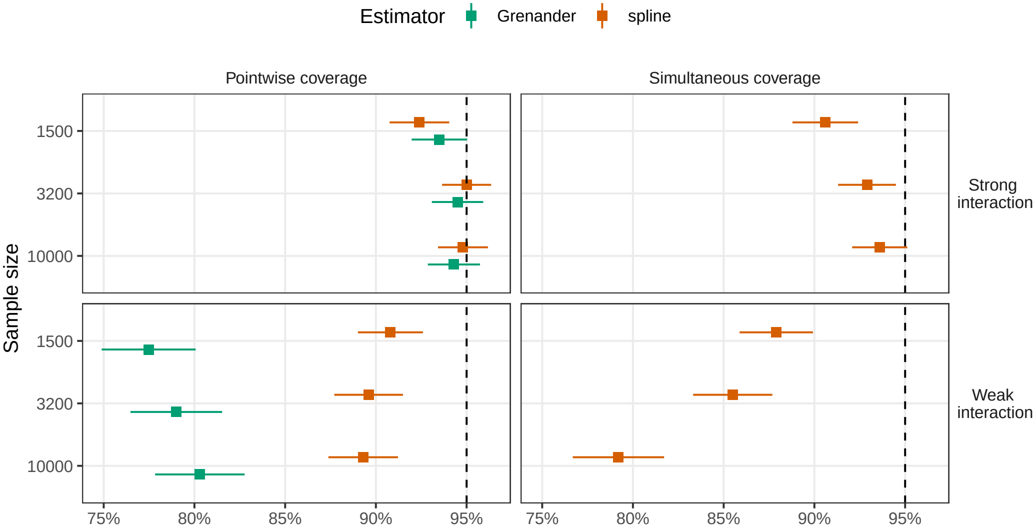
Spline approximation of γ



This **curve** is determined by coefficients that can be written as linear functionals of Γ.

$\longrightarrow$ Standard toolbox available again.

Results from some numerical experiments



Recap and outlook

Univariate monotone function for visualizing and summarizing treatment heterogeneity.

"Non-standard" estimation problem

- Monotone function estimation techniques
- Approximate spline estimand

A spline approximation of a monotone function needs not be monotone itself.

Our suggested estimators work under a continuity assumption.

Estimating the sublevel set size function γ still seems to be a statistically challenging problem in a nonparametric model.

References

Matteo Bonvini, Edward H Kennedy, and Luke J Keele. Minimax optimal subgroup identification. *arXiv preprint arXiv:2306.17464*, 2023.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 2018a.

Victor Chernozhukov, Iván Fernández-Val, and Ye Luo. The sorted effects method: Discovering heterogeneous effects beyond their averages. *Econometrica*, 2018b.

Ulf Grenander. On the theory of mortality measurement: part ii. *Scandinavian Actuarial Journal*, 1956.

P. Groeneboom. Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II*, 1983.

Oliver Hines, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 2022.

Henry WJ Reeve, Timothy I Cannings, and Richard J Samworth. Optimal subgroup selection. *The Annals of Statistics*, 2023.

Alexandre B Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*, 1997.

Mark J van der Laan and Sherri Rose. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media, 2011.

Aad W van der Vaart and Mark J van der Laan. Estimating a survival distribution with current status data and high-dimensional covariates. *The International Journal of Biostatistics*, 2006.

Ted Westling and Marco Carone. A unified study of nonparametric inference for monotone functions. *Annals of statistics*, 2020.

Nonparametric estimation of sublevel-set probabilities of the CATE function

Investigating and summarizing treatment heterogeneity

Anders Munch and Thomas Gerds

available on arXiv soon