Cluster Membership provides Guaranteed Balancing Scores

Bob Obenchain

Risk / Benefit Statistics LLC

Yin = Dark = Evil = Risk

Yang = Light = Good = Benefit

In 1983, Rosenbaum & Rubin introduced the conditional independence theorem of propensity scoring and demonstrated that the unknown, true propensity score is the "most coarse" balancing score while the observed X-vector of covariate values is the "most detailed" balancing score. Here, we argue that membership in a X-space cluster of patients that is relatively small and compact provides a balancing score somewhere between the above extremes of coarse or detailed. In other words, it really is not necessary to estimate propensity scores and perform somewhat tedious checks for balance. Rather, local nonparametric estimates of propensity to be treated are provided by the observed treatment fractions within each cluster. Unlike LATE estimation where covariates are assumed to be instruments, we concentrate here on estimation of Local Treatment Differences (LTDs) within informative clusters.

The bootstrap calculations and graphical displays illustrated here are implemented in my R-package “USPS.”

“Local” Terminology:

- **Subgroups of Patients**
- **Subclasses…**
- **Strata…**
- **Clusters… (natural or forced)**

Near Neighbors

Fundamental PS Theorem

Joint distribution of $\boldsymbol{x}$ and t given $\mathbf{p}$:

$$\Pr(\boldsymbol{x}, t \mid \mathbf{p}) \equiv \Pr(\boldsymbol{x} \mid \mathbf{p}) \Pr(t \mid \boldsymbol{x}, \mathbf{p})$$
$$= \Pr(\boldsymbol{x} \mid \mathbf{p}) \Pr(t \mid \boldsymbol{x})$$
$$= \Pr(\boldsymbol{x} \mid \mathbf{p}) \text{ times } \mathbf{p} \text{ or } (\mathbf{1}-\mathbf{p})$$
$$= \Pr(\boldsymbol{x} \mid \mathbf{p}) \Pr(t \mid \mathbf{p})$$

...i.e $\boldsymbol{x}$ and t are conditionally independent given the propensity for new, $\mathbf{p} = \Pr(t = 1 \mid \boldsymbol{x})$.

This is a deceptively simple theorem in statistics / probability that requires only rather weak assumptions.

The first line above follows from the very definition of conditional probability.

The second line then follows from the fact that p is only a function of X: p = p(X).

The third line then follows because the final factor is the PS vector, with elements p and 1-p.

The fourth line then follows because the PS if a function of X only through the numerical value of p when there are 2 treatments.

I call this the "Fundamental Conditional Independence Theorem" of Propensity Scoring. I think it is misleading to refer to this as the "PS Balancing" Theorem because… NEXT SLIDE!!!

Conditioning on Propensity Scores implies both…

Unconfounding: local X-covariate distributions are the same for both treatments

and

Imbalance: Unequal local treatment fractions …unless Pr(*t* | p) = p = 1−p = 0.5

Because Pr(t|x) is unknown in most cases, not only does Pr(t|p) need to be estimated but also balance needs to be checked / verified.

What is LESS "coarse" than

$$\Pr(x, t \mid p) = \Pr(x \mid p)\,\Pr(t \mid p) \;?$$

Conditioning upon *Cluster Membership* is intuitively somewhere between the two PS extremes in the limit as individual clusters become numerous, small and compact …as long as *t* information is not used to form clusters

$$\Pr(x, t \mid C) \equiv \Pr(x \mid t, C)\,\Pr(t \mid C)$$
$$\approx \Pr(x \mid C)\,\Pr(t \mid C)$$

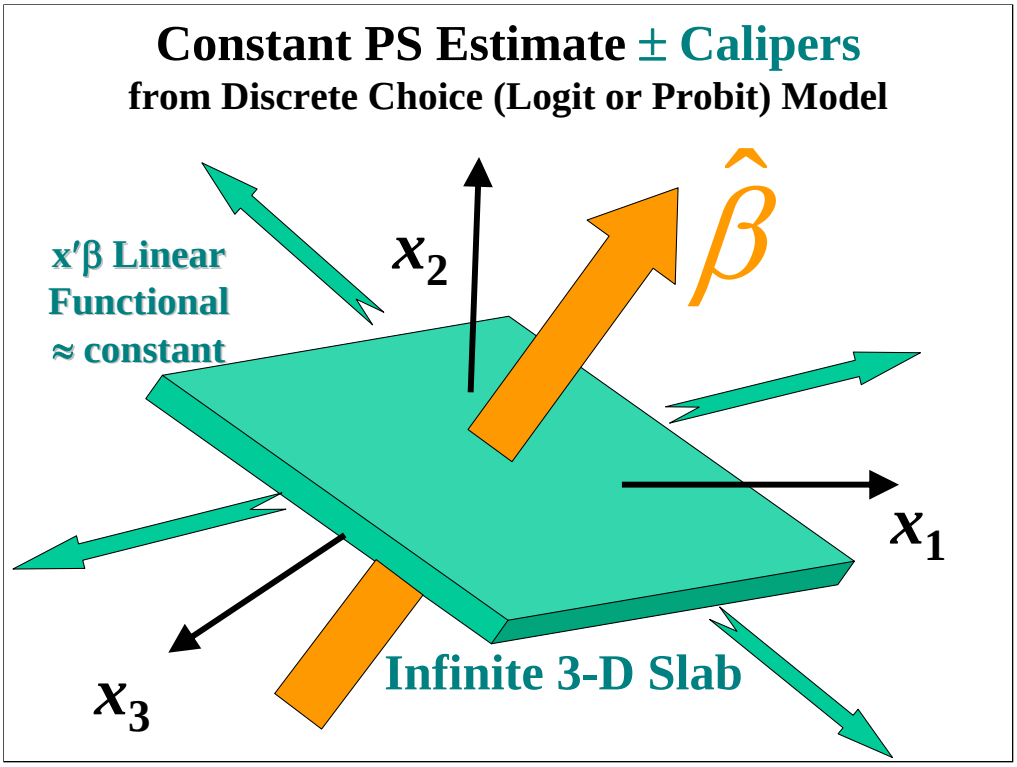
But LESS "detailed" than

$$\Pr(x, t) = \Pr(x)\,\Pr(t \mid x) \;?$$

Here, we propose using (hierarchical) clustering to form numerous and compact (complete linkage) patient sub-groups.

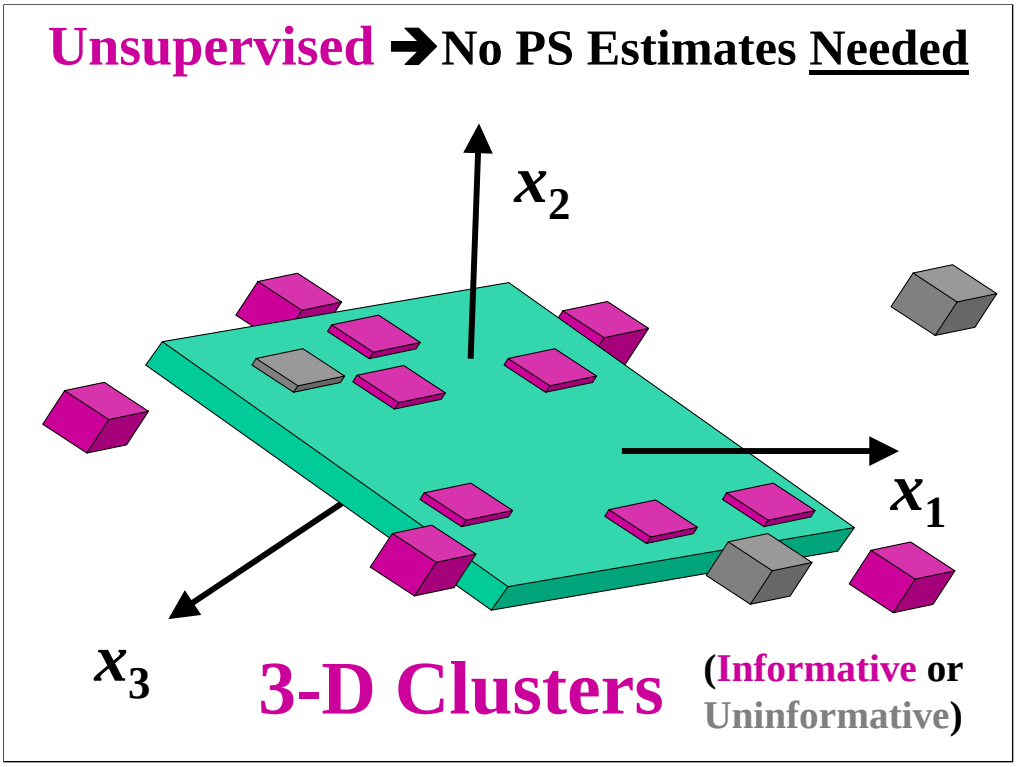
The middle approximation is very poor when clusters are large; otherwise, PSs could not be the MOST COARSE balancing scores.

Clusters will still vary in SIZE relative to both [1] number of patients within the cluster and [2] X-space volume of the cluster.



Slab extends to plus/minus infinity in all directions orthogonal to beta-hat (2 dimensional space here.) Note that the slab has finite depth = (PS plus/minus Calipers) but has infinite volume.

Patients within this X-space slab could certainly have very different x1, x2 and x3 coordinates. Thus no balance on x-factors is automatic.



A cluster is "Informative" when it contains at least one patient from each treatment group.

Local Treatment Differences (LTDs) in outcomes can then be computed.

Observed Treatment Fractions within Clusters are Local, non-parametric PS estimates.

Nested ANOVA

Source	Degrees-of-Freedom	Interpretation
Clusters (Subgroups)	K = Number of Clusters	Local Average Treatment Effects (LATEs) are **Cluster Means**
Treatment within Cluster	Number of "Informative" Clusters ≤ K	Local Treatment Differences (LTDs)
Error	≥ Number of Patients – 2K	Uncertainty

Although a NESTED model can be (technically) WRONG, it is <u>sufficiently versatile</u> to almost always be USEFUL as the number of "clusters" increases.

McClellan et al. (1994) and many economists have studied "instrumental variable" approaches. The key assumption is that observed X-covariates determine only treatment selection and do NOT influence outcome, Y, except through treatment choice. Cluster means are plotted vertically along a horizontal axis describing within-cluster fraction treated (propensity score.)

When X-covariates measure disease severity and/or patient frailty, they are usually predictive of both treatment selection (especially when expensive) and ultimate outcome. In this case, cluster means from a nested model are totally confounded and "K" degrees-of-freedom are immediately lost. But within cluster treatment differences are ALWAYS relevant and become more-and-more relevant as number of clusters increases and, thus, sizes of clusters decrease.

History of Local Control Methods for Human Studies

- **Epidemiology (case-control & cohort) studies**
- **Post-stratification and re-weighting in surveys**
- **Stratified, dynamic randomization to improve balance on predictors of outcome**
- **Matching and Sub-grouping using Propensity Scores**
- **Econometric Instrumental Variables (LATEs)**
- **Marginal Structural Models (IPW ∝ 1/PS)**
- **Unsupervised Propensity Scoring: Nested Treatment-within-Cluster ANOVA model …with LATE, LTD and Error sources of variation**

Why are "Human Studies" being singled out here? Primarily, because human subjects can refuse to participate in designed experiments, and some designs are unethical on human subjects.

Local ➔ make only the clearly more relevant comparisons.