§ 3.8 Propensity Score.

\subsection {What's Pps?}

o Propensity Score: probability of receiving treatment given covariates X.

在给定协变量X下接受treatment的概率

o 表示: subject i 的 propensity score.
$$\pi_i = P(A=1 | X_i)$$

例: o Covariate X = age.

年龄大的人更可能接受治疗 (treatment).

故 $\pi_i = P(A=1 | X_i) > \pi_j = P(A=1 | X_j)$, if $X_i > X_j$.

o 如果 Pps 的直观含义: 若 $\pi_i = 0.3$, 表示第 i 个病人接受treatment的权衡为0.3.

可能性为30%.

★ \subsection {Balancing score}

Motivation: 2 subjects have the same value of the propensity score.

but they possibly have different covariate values X.

$\pi = P(A=1 | X)$

o Pps is the function of X.

说明这两个subjects有相同概率被分到treatment group;

对于有相同propensity score的子群, 在两个treatment组中应该具有balance.

● Propensity score是一种 balance score.

balance score: If you condition on it, you'll have balance.

考点等式: $P(X = x | \pi(x) = p, A=1) = P(X = x | \pi(x) = p, A=0)$.

● Match on the propensity score ⟹ achieve balance.

(替代 Match on 所有的 covariates之间的 distance)

通过它们

o 在 Randomized trial中, Pps 是已知的. $P(A=1 | X) = P(A=1) = 0.5$.

分配A 与X无关

o 在 observational study中, Pps 未知的.

注意: 我们可以通过observed data: A、X ⟶ 估计 Pps.

所以当我们提起Pps时, 通常是估计的Pps.

\sub { Estimating Pps }

目的估计 $P(A=1 | X)$

方法: logistic 回归.

步骤: Step1: 拟合逻辑回归模型, A作为 Outcome, X协变量

Step2: 对每个subject: 得到 Predicted value of Pps.

§3.9. Pps matching

- Pps is a balancing score.
- Matching on the Pps _{should} achieve balance.

- Pps 是一个构量，每个 subject 都有1个 确定的 Pps.

- Match on Pps 的复杂度：只需要匹配了1个变量，matching 问题被 simplified.

对比 Match on covariates：需要匹配了一个变量集合.

\subsection{overlap}

- 比较 Pps 在 treated 组与 control 组 subjects 的 分布. (通过画图).

Example 1:
Fig. {example}

① 观察峰值
② π (treated) > π (control)

Overlap 是什么？
    Also 不管是 Pps 的哪部分数值，仍然有人接受处理或控制.

- 在 Fig. {ex} 的最右端，即使是 control 组，也总有人有很高的概率接受 treated.
- 在 Fig. {ex} 的最左端，即使是 treated 组，也总有人以很小的概率才能接受 treated.
∨ positivity assumption 成立.

Example 2: {反例} ⟹ Fig. {Example 2}

消除有着极端 Pps 的个体
⇓
trimming tails.

\sub {trimming tails} Fig. {Trimming tails}

在 lack of overlap 时^{使用}适用.

- 目的：去掉那些具有极端 Pps 值的 subjects.
- 极端 Pps 的含义：对于使用数据的人而言，接受某类处理的概率较高.

可行方法举例：

控制 : 去掉那些 Pps 值小于 "treated group 中最小 Pps" 的 subjects
组 subject

treated : 去掉那些 Pps 值大于 "control group 中最大 Pps" 的 subjects
组

作用：保障 positivity assumption 的合理性.

\sub {matching}

思想：计算 2 subjects Pps 的 distance.
    求 min distance.

方法：Greedy / optimal / 最近邻 matching.

又：计算每一个 treated subject 与每个 control 的 Pps.

取对数. $0 < Pps < 1$, 在0与1之间非常密集导致很多 Pps 都十分相似.

取 log 性质：① 单调性，保证 Pps 的 rank 不变.
    ② 无界性，拉大了 Pps 的分布，使其不再密集

⟹ 对 $\log(\pi)$ 作 Match，而不用 π.

\sub {Caliper}

目的：避免接受 bad matches.

设定：最大距离

我们能接受的 Pps 间的。

一般设定：　　0.2 × Sd ⟹ 如果对 balance 的要求更高 ⟹ fewer matched pairs, better matches.

$Sd = Sd(\log(\pi_{\tilde{c}}))$　可降到 0.1 Sd

所有的 person？

treated 和 control 合在一起？

流程：省略。 ==Fig {Caliper}.==

Small Caliper → less bias, more variance → treatment effect estimate will have more variability

trade-off in bias 和 variance 考虑得多了.

\sub {After matching}

同 Match based on distance of covariates.

Match 之后要检验 Match 是否达标. | Randomization test.

方法：是否达到 balance. | GEE. 等.