LOO_MRP Simulation Design document

Swen Kuh

13/09/2021

This document will describe the simulations we plan to conduct and the expected results from using `loo` and `weighted_loo` (`wtd_loo`) to do model comparisons.

Description of variables

Say we want to explore having variables that are **strongly/weakly predictive of the outcome** and **strongly/weakly predictive of the response**. We will have 4 variables X_1, X_2, X_3 and X_4, and an outcome Y. Lets suppose:

Variable type	Weakly predictive of the (svy) response	Strongly predictive of the (svy) response
Weakly predictive of the outcome	X_1	X_3
Strongly predictive of the outcome	X_2	X_4

All the X's were sampled unevenly from the population, which gives us strong indication for incorporating inclusion probability in our models.

We expect $2^4 - 1 = 15$ different models for the combination of these variables. If we ran these models (in `R` syntax):

1) $Y \sim X_1 + X_2 + X_3 + X_4$
2) $Y \sim X_1 + X_2 + X_3$
3) $Y \sim X_1 + X_2 + X_4$
4) $Y \sim X_1 + X_3 + X_4$
5) $Y \sim X_2 + X_3 + X_4$
6) $Y \sim X_1 + X_2$
7) $Y \sim X_1 + X_3$
8) $Y \sim X_1 + X_4$
9) $Y \sim X_2 + X_3$
10) $Y \sim X_2 + X_4$
11) $Y \sim X_3 + X_4$
12) $Y \sim X_1$
13) $Y \sim X_2$

14) $Y \sim X_3$

15) $Y \sim X_4$

16) $Y \sim X_1$

17) $Y \sim X_2$

18) $Y \sim X_3$

19) $Y \sim X_4$

20) $Y \sim X_1 + X_2$

21) $Y \sim X_1 + X_3$

22) $Y \sim X_1 + X_4$

23) $Y \sim X_2 + X_3$

24) $Y \sim X_2 + X_4$

25) $Y \sim X_3 + X_4$

26) $Y \sim X_1 + X_2 + X_3$

27) $Y \sim X_1 + X_2 + X_4$

28) $Y \sim X_1 + X_3 + X_4$

29) $Y \sim X_2 + X_3 + X_4$

30) $Y \sim X_1 + X_2 + X_3 + X_4$

What should we expect?

If we only look at models #1-4, the expected performance is that using `loo` as comparison, it should pick out X_2 and X_4 as they are **strongly predictive of the outcome**. But if we use the `wtd_loo` as comparisons, it should pick out X_4 as we now take into account the survey design and weights accordingly, and X_4 is **strongly predictive of the survey response**.