Optimal Data Analysis
Vol. 6 (June 8, 2017), 43-46

2155-0182/10/$3.00

Computing Propensity Score Weights for CTA Models Involving Perfectly Predicted Endpoints

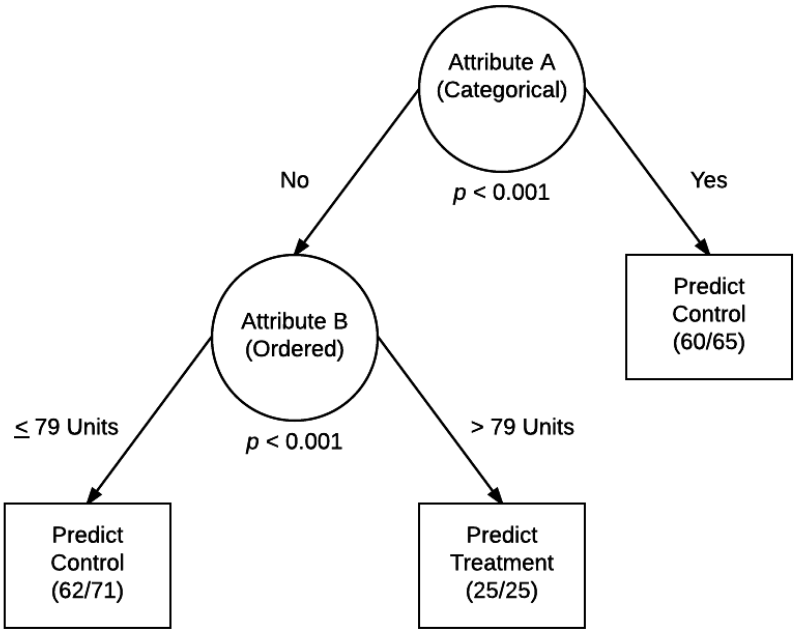
Paul R. Yarnold, Ph.D., and Ariel Linden, Dr.P.H.

Optimal Data Analysis, LLC Linden Consulting Group, LLC

The use of CTA[1,2] to construct propensity score weights[3] is complicated by division by zero in models having any perfectly predicted endpoints: omitting undefined propensity scores yields a degenerate solution. This note presents an algorithmic remedy to this situation.

The three-strata CTA propensity-score model illustrated in Figure 1 emerged in an applied analysis. As seen, all 25 of the observations scoring "No" on attribute A, and also scoring "> 79 units" on attribute B, were correctly predicted to be from the Treatment condition.

Figure 1: CTA Propensity Score Model with One Perfectly Predicted Endpoint



2155-0182/10/$3.00

For CTA models, a stratified propensity score weight is generated for each observation based on their actual treatment assignment and assigned model endpoint. Observations have the same propensity score weights if they have the same actual treatment assignment (either treated or non-treated) and are thus classified into the same endpoint. CTA model-based propensity score weights[3,4] are computed as:

$$\frac{n_s \times \Pr(Z = z)}{n_{z=z,s}}$$

where n_s is the total number of individuals in a given stratum s, $\Pr(Z = z)$ is the estimated probability of assignment to treatment group z (i.e., the proportion of individuals actually receiving treatment z in the sample), and $n_{z=z,s}$ is the total number of individuals in stratum s who were actually assigned to treatment z. This stratified weighting approach ensures weights conform exactly to the underlying geometry and findings of the CTA model, reduces bias due to imbalance in the covariates used to create the propensity score, and standardizes each treatment group to the target population.[3,4]

Table 1 presents the propensity score weight computed for each model endpoint and actual class assignment: as seen, this weight is undefined for control observations scoring "No" on attribute A, and > 79 units on attribute B.

Table 1: Empirical Finding Involving One Perfectly Predicted Model Endpoint

Attribute A	Attribute B	Class	Propensity Score Weight	Adjusted Propensity Score Weight
Yes	---	Control	0.7578*65/60=0.8210	0.7593*65/60=0.8226
Yes	---	Treatment	0.2422*65/5=3.1486	0.2407*65/5=3.1291
No	≤ 79 units	Control	0.7578*71/62=0.8678	0.7593*71/62=0.8695
No	≤ 79 units	Treatment	0.2422*71/9=1.9107	0.2407*71/9=1.8989
No	> 79 units	Control	0.7578*25/0=Undefined	0.7593*26/1=19.7418
No	> 79 units	Treatment	0.2422*25/25=0.2422	0.2407*26/25=0.2503

In Table 1 the sample mean propensity score weight obtained for the five defined values is: (60*0.8210 + 5*3.1486 + 62*0.8678 + 9*1.9107 + 25*0.2422) / (60+5+62+ 9+25) = 142.0579 / 161 = 0.8824. This mean weight value indicates how dropping observations with an undefined weight confounds propensity score weighting for the remaining sample. In this weighting scheme weights are standardized by the marginal probability of treatment, and thus should have a mean of 1.0 (not all weighting schemes are adjusted in this manner and as a result, some weights may be extremely large).

Accordingly, the identical method used to compute the odds of class membership in a staging table created for a CTA model with a perfectly homogeneous endpoint[1,5] is adopted here to define a propensity score weight for a perfectly homogeneous endpoint: *an undefined profile is modified by adding one misclassified observation*. For example, for the undefined weight in Table 1, instead of basing the propensity score weight computation on the observed empirical result of 25 of 25 correct classifications (and obtaining an undefined propensity score weight value), propensity score weight

2155-0182/10/$3.00

computation is instead based on the adjusted (hypothetical) result of 25 of 26 correct classifications. Note that $\Pr(Z = z)$ is modified to reflect the adjusted sample—which has one more observation than the actual sample.

The sample mean adjusted propensity score weight (Table 1) obtained for all six of the defined values is: (60*0.8226 + 5*3.1291 + 62* 0.8695 + 9*1.8989 + 1*19.7418 + 25* 0.2503) / (60+5+62+ 9+1+26) = 161.9999 / 162 = 0.9999. This value is within rounding error of the expected mean sample weight of 1.0.

Table 2 is an example of the extreme case in which the model achieves perfect classification and every model endpoint is perfectly homogeneous.[6] As seen, each undefined profile is modified by adding one misclassified observation, and $\Pr(Z = z)$ is modified to reflect the adjusted sample—which has three more observations than the actual sample.

Table 2: Hypothetical Finding Involving All Perfectly Predicted Model Endpoints

Attribute A	Attribute B	Class	Propensity Score Weight	Adjusted Propensity Score Weight
Yes	---	Control	0.7578*65/65=0.7578	0.4085*66/65=0.4148
Yes	---	Treatment	0.2422*65/0=Undefined	0.5915*66/1=39.0390
No	≤ 79 units	Control	0.7578*71/0=Undefined	0.4085*72/1=29.4120
No	≤ 79 units	Treatment	0.2422*71/71=0.2422	0.5915*72/71=0.5998
No	> 79 units	Control	0.7578*25/0=Undefined	0.4085*26/1=10.6210
No	> 79 units	Treatment	0.2422*25/25=0.2422	0.5915*26/25=0.6152

In Table 2 the mean propensity score weight for all three defined values (endpoints) is: (65* 0.7578 + 71*0.2422 + 25*0.2422) / (65+ 71+ 25) = 72.5082 / 161 = 0.4504.

The mean adjusted propensity score weight for all six model endpoints is: (65* 0.4148 + 1* 39.0390 + 1* 29.4120 + 71*0.5998 + 1* 10.6210 + 25* 0.6152) / (65+1+1+71+ 1+25) = 163.9998 / 164 = 1.0000.

In application it is important to test for covariate balance before and after adjusting the propensity score weights, to ensure the weights did in fact adjust for imbalances.[7] In this regard, a sequential sensitivity analysis[8] may be used to assess limits of generalizability of the findings obtained by applying the adjustment procedure. In Table 1, for example, in the first step of the sensitivity analysis the adjusted propensity score weight computation is based on a hypothetical result of 25 of 26 correct classifications. In the second step computation is based on a hypothetical result of 26 of 27 correct classifications, and so forth. This procedure may be continued until the applied finding either changes as regards quantitative and/or qualitative implications, or converges to a stable solution.

References

[1]Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

[2]Yarnold PR (2017). What is optimal data analysis? *Optimal Data Analysis*, *6*, 26-42.

[3]Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*. DOI: 10.1111/jep.12744

2155-0182/10/$3.00

[4]Linden A (2014). Combining propensity score-based stratification and weighting to improve causal inference in the evaluation of health care interventions. *Journal of Evaluation in Clinical Practice, 20,* 1065-1071.

[5]Soltysik RC, Yarnold PR (2010). Automated CTA software: Fundamental concepts and control commands. *Optimal Data Analysis, 1*, 144-160.

[6]Yarnold PR, Linden A (2016). Theoretical aspects of the D statistic. *Optimal Data Analysis*, *5*, 171-174.

[7]Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice, 22,* 848-854.

[8]Linden A, Yarnold PR (2017). Minimizing imbalances on patient characteristics between treatment groups in randomized trials using classification tree analysis. *Journal of Evaluation in Clinical Practice*. DOI: 10.1111/jep.12792

Author Notes

This statistical article is exempt from IRB review. No conflict of interest was reported.