

Constantine Frangakis\*

# A Falsifiability Characterization of Double Robustness Through Logical Operators

<https://doi.org/10.1515/jci-2018-0016>

Received June 4, 2018; revised October 25, 2018; accepted November 9, 2018

**Abstract:** We address the characterization of problems in which a consistent estimator exists in a union of two models, also termed as a doubly robust estimator. Such estimators are important in missing information, including causal inference problems. Existing characterizations, based on the semiparametric theory of projections, have seen sufficient progress, but can still leave one's understanding less than satisfied as to when and especially why such estimation works. We explore here a different, explanatory characterization – an exegesis based on logical operators. We show that double robustness exists if and only if we can produce consistent estimators for each contributing model based on an “AND” estimator, i. e., an estimator whose consistency generally needs both models to be correct. We show how this characterization explains double robustness through falsifiability.

**Keywords:** double robustness, logical operators

## 1 Motivation

Consider the following problem that often motivates doubly robust estimation [1]. For patients with a particular disease, one of two treatments,  $z = 1$  or  $0$  can be given, in which case a potential outcome  $Y(z)$  can be observed [2]. For each patient  $i$ , physicians first measure a set of covariates  $X_i$ , and assign treatment  $Z_i$  such that the mechanism of assignment given  $X_i$  is ignorable [2], i. e. such that  $Z_i$  is independent of  $Y_i(z)$  given  $X_i$ . Finally, physicians measure outcome  $Y_i := Y_i(Z_i)$ . The target is to estimate the average outcomes  $E_0\{Y_i(z)\}$  if all patients had been assigned  $z = 1$  vs  $0$ . For simplicity, focus on estimating one of the two averages, say  $\tau_0 := E_0\{Y_i(z = 1)\}$  using consistent estimators.

The average  $\tau_0$  is identified from the observed data distribution as  $E_0\{E_0(Y_i | Z_i = 1, X_i)\}$ . If the covariate is discrete, substituting the empirical distribution provides a consistent, nonparametric MLE. If, however, the covariate space is continuous, an accurate nonparametric estimator is infeasible [3], and researchers often consider additional assumptions.

One usual assumption is that the regression  $y_0(X_i) := E_0(Y_i | Z_i = 1, X_i)$  is in a small enough model that renders  $y_0$  consistently estimable by at least one computable estimator, say  $\hat{y}(X_i)$ . Under general conditions, this implies that the limit of the substitution estimator  $\frac{1}{n} \sum \hat{y}(X_i)$  is the same as the limit of  $\frac{1}{n} \sum y_0(X_i)$ , which is the desired  $\tau_0$ . An alternative assumption has been that physicians provide enough information on the propensity score  $e_0(X_i) := \text{pr}_0(Z_i = 1 | X_i)$  [4] to restrict it in a model that renders it consistently estimable by at least one computable estimator, say  $\hat{e}(X_i)$ . Generally, this implies that the estimator  $\frac{1}{n} \sum Y_i Z_i / \hat{e}(X_i)$  has the same limit as that of  $\frac{1}{n} \sum Y_i Z_i / e_0(X_i)$ , which, again, is the true target  $\tau_0$ . Below we abbreviate the first assumption by  $y_0 = y$ , where  $y$  is the general form of the probability limit of  $\hat{y}$  even if the assumption is false. Similarly, we abbreviate the second assumption by  $e_0 = e$ , where  $e$  is the general form of the probability limit of  $\hat{e}$  even if the above assumption is false.

An interesting case is if one assumes that, for the true distribution of the data,  $y_0 = y$  or  $e_0 = e$  (or both). As a union set, this is more plausible than either assumption considered only by itself. For this assumption,

---

\*Corresponding author: Constantine Frangakis, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA, e-mail: cfrangak@jhsph.edu

it is well known that the estimator

$$\frac{1}{n} \sum \hat{y}(X_i) + Y_i Z_i / \hat{e}(X_i) - \hat{y}(X_i) Z_i / \hat{e}(X_i) \quad (1)$$

is consistent for  $\tau_0$ , and is therefore called a “doubly robust estimator” [1, 5].

Although the verification of this result is easy, standard *derivations* of the above estimator in the first place are relatively lengthy and involve semiparametric projection theory. One such derivation includes finding the form of all regular and asymptotically linear estimators of  $\tau_0$  under assumption  $y_0 = y$  only; and then doing the same under assumption  $e_0 = e$  only. Then the set of estimators each of which can be expressed in both forms [6] characterizes the doubly robust estimators of the problem.

If the estimator solves the nonparametric influence function evaluated at some working models, then an alternative way to find the robustness conditions of such estimator is to find the zeros of the second order remainder in the von Mises expansion. Specifically if the remainder has factors that are differences between “posited minus true” value of certain component, then correct specification of that component is a sufficient condition for robustness (Mark van der Laan, personal communication).

More generally, such characterization based on semiparametric theory, have seen sufficient progress [7, 3], but still leave one’s understanding less than satisfied as to when and especially why such estimation works. Our purpose here is to add to this understanding by presenting an exegesis – an alternative characterization based on logical operators that focuses on falsifiability.

## 2 Characterization emphasizing falsifiability

We argue that double robustness is critically related to how to produce an OR operator using an AND operator. By an operator here we mean the equivalence class that contains a possibly random function with a probability limit, and all other functions with the same limit. Denote “share same equivalence class” by  $\stackrel{\ell}{=}$ .

For the example with the ignorable assignment of the previous section, consider an operator  $\tau^{\text{AND}}(y, e)$  that takes as input the components  $(y, e)$  and has the “AND” property that:

$$\tau^{\text{AND}}(y, e) \stackrel{\ell}{=} \tau_0 \text{ (the true value) if } y = y_0 \text{ AND } e = e_0. \quad (2)$$

Then, the operator defined as

$$\tau^{\text{OR}}(y, e) \stackrel{\ell}{=} \tau^{\text{AND}}(y, e_0) + \tau^{\text{AND}}(y_0, e) - \tau^{\text{AND}}(y, e) \quad (3)$$

produces the true value  $\tau_0$  if  $y = y_0$  OR  $e = e_0$ .

This operator provides explanatory and predictive power on a number of important points regarding the structure of doubly robust estimators.

**(a) The logical operator reproduces easily well known results.** To see how the logical operation argument can easily reproduce this result in the ignorable assignment example, we can choose  $\tau^{\text{AND}}(y, e)$  to depend on both components  $(y, e)$  as the operator

$$\tau^{\text{AND}}(y, e) \stackrel{\ell}{=} \frac{1}{n} \sum y(X_i) Z_i / e(X_i) \text{ with pr. limit } E_0 \{y(X_i) e_0(X_i) / e(X_i)\} \quad (4)$$

satisfying (2). Then, we can obtain  $\tau^{\text{AND}}(y, e_0)$  as equivalent to (with same probability limit as) the operator  $\frac{1}{n} \sum y(X_i)$  (the latter has the same limit as  $\frac{1}{n} \sum y(X_i) Z_i / e_0(X_i)$ ); and we can obtain  $\tau^{\text{AND}}(y_0, e)$  as equivalent to the operator  $\frac{1}{n} \sum Y_i Z_i / e(X_i)$  (the latter has the same limit as  $\frac{1}{n} \sum y_0(X_i) Z_i / e(X_i)$ ). Then, the OR operator in (3) is constructed as

$$\frac{1}{n} \sum \frac{y(X_i)}{\tau^{\text{AND}}(y, e_0)} + \frac{Y_i Z_i / e(X_i)}{\tau^{\text{AND}}(y_0, e)} - \frac{y(X_i) Z_i / e(X_i)}{\tau^{\text{AND}}(y, e)}. \quad (5)$$

Operator	$\tau^{\text{AND}}(y, e)$	$\tau^{\text{AND}}(y, e_0)$	$\tau^{\text{AND}}(y_0, e)$	$\tau^{\text{OR}}(y, e) = \tau^{\text{AND}}(y, e_0) + \tau^{\text{AND}}(y_0, e) - \tau^{\text{AND}}(y, e)$
estimator form	$\frac{1}{n} \sum_i \frac{\hat{y}(X_i)Z_i}{\hat{e}(X_i)}$	$\frac{1}{n} \sum_i \hat{y}(X_i)$	$\frac{1}{n} \sum_i \frac{Y_i Z_i}{\hat{e}(X_i)}$	<b>as in (6)</b>
limit form	$E_0 \left\{ \frac{y(X_i)e_0(X_i)}{e(X_i)} \right\}$	$E_0 \left\{ \frac{y(X_i)e_0(X_i)}{e_0(X_i)} \right\}$	$E_0 \left\{ \frac{y_0(X_i)e_0(X_i)}{e(X_i)} \right\}$	$\tau_0^+ E_0 \left[ \frac{\{e_0(X_i) - e(X_i)\} \{y_0(X_i) - y(X_i)\}}{e(X_i)} \right]$
result if $e = e_0$	$E_0 \{y(X_i)\}$	$E_0 \{y(X_i)\}$	$\tau_0$	$\tau_0$
result if $y = y_0$	$E_0 \left\{ \frac{y_0(X_i)e_0(X_i)}{e(X_i)} \right\}$	$\tau_0$	$E_0 \left\{ \frac{y_0(X_i)e_0(X_i)}{e(X_i)} \right\}$	$\tau_0$

Figure 1: Generation of OR from AND operator, and falsifiability in the example of Section 1.

So, estimating the components  $(y, e)$  by estimators  $(\hat{y}, \hat{e})$  produces the estimator version of (5)

$$\tau^{\text{OR}}(\hat{y}, \hat{e}) := \frac{\ell}{n} \sum \frac{\hat{y}(X_i)}{\tau^{\text{AND}}(\hat{y}, e_0)} + \frac{Y_i Z_i / \hat{e}(X_i)}{\tau^{\text{AND}}(y_0, \hat{e})} - \frac{\hat{y}(X_i) Z_i / \hat{e}(X_i)}{\tau^{\text{AND}}(\hat{y}, \hat{e})}, \quad (6)$$

which, by (3), has the property that  $\tau^{\text{OR}}(\hat{y}, \hat{e}) \stackrel{\ell}{=} \tau_0$  if  $\hat{y} \stackrel{\ell}{=} y_0$  OR  $\hat{e} \stackrel{\ell}{=} e_0$ . In other words, we have constructed in this alternative way the estimator  $\tau^{\text{OR}}(\hat{y}, \hat{e})$ , which equals to (1) (see also Fig. 1), as a doubly robust estimator.

For the above example, there exist a number of other estimators that are doubly robust [7]. However, all such estimators are in the same equivalence class as above, and they share the same influence function. Also, although the above example is relatively simple, many enriched variations, including parametric vs single value for  $(y, e)$ ; multiple time points; other censored or missing data versions; or more complex outcomes, share the basic structure as above.

**(b) The three-term structure arises from the general logical relation between an OR and an AND operator.** We can generalize the above arguments as follows.

**Result.** If a consistently estimable operator  $\tau^{\text{AND}}(y, e)$  has its two terms,  $\tau^{\text{AND}}(y, e_0)$  and  $\tau^{\text{AND}}(y_0, e)$ , consistently estimable, then the resulting estimator of the operator  $\tau^{\text{OR}}(y, e)$  in (3) produces a doubly robust estimator. Also, if there exists an OR operator, then it has the structure (3).

The first claim is easy to check from above. For the second claim, note that if there exists an **OR** operator, say  $\tau^{\text{OR}}(y, e)$ , then it is also an **AND** operator, and, when used as such in the RHS of (3) it produces the OR operator  $2\tau_0 - \tau^{\text{OR}}(y, e)$ . If we now use this new OR operator again as an AND operator in the RHS of (3), we obtain back the original OR operator.

This result predicts that the three-term structure of (1) is not specific to the above example, but rather obeys the logical structure that any OR operator has as constructed from an AND operator (3).

**(c) The role of the AND operator in falsifiability.** The AND operator can be seen more practically as essential for falsifiability. To see this, focus in the case where indeed either  $e = e_0$  or  $y = y_0$ , as in the last two rows of Figure 1 for the example discussed above. Suppose the researcher, who does not know which of the rows is correct, observes that the estimator forms of  $\tau^{\text{AND}}(y, e_0)$  and  $\tau^{\text{AND}}(y_0, e)$  differ beyond statistical variability. The first conclusion is that one of the two limit values is correct and the other is incorrect. It is by construction, and not happenstance, that in either row, the operator  $\tau^{\text{AND}}(y, e)$  points to the *incorrect* value of  $\tau$ , and it is in this way that we can conclude that its alternative is the true value.

**(d) Connection with variation dependent components.** Double robustness is usually examined only when the components  $(y, e)$  are variation independent, meaning that the hypothesis, say,  $H_1 : y_0 = y$  does not a priori alter the possibility that the hypothesis, say  $H_2 : e_0 = e$ , is correct. It is possible that the researcher

may want to consider hypotheses on components that are not variation independent. Nevertheless, the same characterization based on the AND operator in the above result appears to be unaffected by and applicable to problems with variation dependent components.

Consider an extreme example where the estimand now is the true regression  $y_0(x)$ , and, as with simple hypothesis testing, assume that it is either equal to a specified function  $y^{(1)}(x)$ , or equal to a different specified function  $y^{(2)}(x)$ . As with double robustness, this assumption for  $y_0(x)$  is a union of two models, but where, now, the components involved in each of the two models are the same and equal to the estimand, i. e., there is complete variation dependence. As is well known, though, one can easily discern with high precision that the true regression will be the one with the smallest estimated mean squared error in a large enough sample. This argument is also the result of the operator  $y^{\text{AND}}(y^{(1)}, y^{(2)})$  that chooses the regression function with the *larger* (worse) of the two estimated mean squared errors,  $\text{mse}(y^{(1)})$  and  $\text{mse}(y^{(2)})$  (and chooses any of the two if they have equal mse). Then, note that (i)  $y^{\text{AND}}(y^{(1)}, y^{(2)})$  is indeed an AND operator in the sense of (2): it equals  $y_0$  if (but not necessarily only if)  $y_0 = y^{(1)}$  and  $y_0 = y^{(2)}$ ; (ii) the resulting OR operator

$$y^{\text{OR}}(y^{(1)}, y^{(2)}) := y^{\text{AND}}(y^{(1)}, y_0) + y^{\text{AND}}(y_0, y^{(2)}) - y^{\text{AND}}(y^{(1)}, y^{(2)}) \quad (7)$$

still produces the true regression  $y_0(x)$  if either  $y_0 = y^{(1)}$  or  $y_0 = y^{(2)}$ ; and (iii) the AND operator actually never produces the true value  $y_0$ , but this does not matter because its role is to show the false value and cancel it in (7).

**(e) Further comments.** The interpretation of why and how one of the two component models is misspecified is important in the discussion between the physician and the statistician. As [3] (p. 923) also note, the construction of a doubly robust estimator allows us to point to which of the two component estimators it disagrees with the most, and interpret this as evidence of which component is misspecified. Although useful, this observation is different from that discussed in (c) above, and is formed *after* construction and *in terms of* the doubly robust estimator. Therefore, such observation by itself does not quite answer why and how this evidence arises, as with the more elemental use of the components in the three terms of the OR operator.

The present characterization does not antagonize use of semiparametric methods, which can be very useful for deriving, when possible, AND operators with regularly estimable marginal components. On the other hand, there exist union models with consistent estimators that are irregular [8], but which still obey the three-term structure discussed above.

## References

1. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J Am Stat Assoc.* 1999;94:1096–120.
2. Rubin DB. Inference and missing data. *Biometrika.* 1976;63:581–92.
3. Robins JM, Rotnitzky A. Comments on “Inference for semiparametric models: some questions and an answer” by Bickel PJ and Jaimyoung K. *Stat Sin.* 2011;11:920–36.
4. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70:41–55.
5. Hahn J. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica.* 1998;66:315–31.
6. Tsiatis A, Davidian M. Comments on “Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data” by Kang JDY and Schafer JL. *Stat Sci.* 2007;22:569–73.
7. Kang JD, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci.* 2007;22:523–39.
8. Frangakis C, Rubin D. Rejoinder to discussions on addressing an idiosyncrasy in estimating survival curves using double sampling in the presence of self-selected right censoring. *Biometrics.* 2001;57:351–3.