 DOI: 10.1111/biom.12365

Rejoinder to Discussions on: Deductive Derivation and Turing-Computerization of Semiparametric Efficient Estimation

Constantine E. Frangakis,* Tianchen Qian, Zhenke Wu,*** and Iván Díaz******

Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland 21205, U.S.A.
**email:* cfrangak@jhsph.edu
***email:* tqian@jhsph.edu
****email:* zhwu@jhu.edu
*****email:* idiaz@jhu.edu

We thank the Editors and the Associate Editor for the opportunity to have this exchange. We also thank the discussants for welcoming this search to make semiparametric inference more deductive, and for all their very interesting and useful comments.

1. Luedtke, Carone, and van der Laan

Luedtke, Carone, and van der Laan have raised a number of interesting and useful points on the use of the Gateaux derivative, on extensions to a priori restricted models, and on criteria beyond solving an efficient influence function.

1.1. *Discreteness versus Continuity (on Comments 1 and 2)*

Luedtke, Carone, and van der Laan pointed out that in some problems, for example, when parts of the data are continuous, one should be extra careful about the use of the Gateaux derivative. They also offered an interesting and useful regularization method that can overcome some of these concerns. These concerns are useful to consider, and below we suggest that many of them can be overcome also at the stage of formulation, through discreteness.

By discreteness here, first, we mean the formulation in which all measurements in the problem are in principle assumed to be discrete and bounded, even though the possible levels may, of course, be more than the data points. Bounded discreteness is true for any known measurement device, and discreteness is even acceptable in current physical theories such as quantum mechanics. Under such formulation, the concerns raised by the discussants seem to be alleviated. For example, for the estimand, $\tau^*(F) := \int y(x)dG_0(x)$ proposed by the discussants, the perturbed estimand $\tau^*(F_{d',\epsilon})$ of expression (8) in the main article is $\int y_{d',\epsilon}(x)dG_0(x)$, with $y_{d',\epsilon}(x)$ as in the article. Taking the Gateaux derivative, then, gives simply the first of the two summands in the last highlighted expression of Section 2.2 except that $p_{d',\epsilon=0}(x)dx$ should now be replaced by $dG_0(x)$. The result then becomes

$$\frac{r'\{y' - y(x')\}}{e(x')}\frac{p_0(x')}{p(x')},$$

as in the discussants' expression. See also van der Laan and Rose (2011), Appendix for analogous examples of discretization.

Discreteness is also justified if the estimand is chosen by the substantive scientist (e.g., physician). For example, if a physician does choose the discussants' second estimand, $\int fdF$, then most likely the physician has in mind, in principle, a discrete version of that estimand. It would then be useful for the statistician to encourage the physician to express what they mean by that estimand if the data are discrete. This would be not only practically acceptable from a physical perspective, as noted above, but could further enhance clarity in the communication between the statistician and the physician.

Such use of discreteness does not mean that a further parsimonious model would not be useful for estimation, but it means that parsimony across data does not necessarily need continuity of the data themselves. A possible treatment of parsimonious, restricted models is discussed next.

1.2. *Extension to Restricted Semiparametric Models (on Comment 3)*

We agree with the discussants that an important next step is to extend deductive methods to a priori restricted models. One possible way to capitalize on the deductive method proposed for the unrestricted models in the main article, in order to proceed to a restricted model with finite dimensional parameter is as follows. First, determine the estimands τ^{env} in the unrestricted model that "envelope" the restrictions in the restricted model, that is,

$$\tau^{env} := \left\{ \begin{array}{l} \text{all estimands in the unrestricted} \\ \text{model that are tied together by } \beta \\ \text{in the restricted model} \end{array} \right\} \qquad (1)$$

Using the methods of the article, one can then determine the unrestricted EIFs, say ϕ, of the estimands τ^{env} in (1). These EIFs are expected to contain most of the information from the original data to estimate the parameter β in the restricted model. Since, in a large enough sample, the sum of ϕ is approximately normal, the restricted EIF for β can be

Table 1
Perturbation model after a discretization to the sample data

Data	Original model, F^*	Discrete model, F	Perturbed model, $(1-\epsilon)F + \epsilon < d' >$, when				
			$d' = D_1$	...	$d' = D_i$	...	$d' = D_n$
D_1	$F^*(D_1)$	$F_1 = \frac{F^*(D_1)}{\sum_k F^*(D_k)}$	$(1-\epsilon)F_1 + \epsilon$		$(1-\epsilon)F_1$		$(1-\epsilon)F_1$
.	.	.	.		.		.
D_i	$F^*(D_i)$	$F_i = \frac{F^*(D_i)}{\sum_k F^*(D_k)}$	$(1-\epsilon)F_i$		$(1-\epsilon)F_i + \epsilon$		$(1-\epsilon)F_i$
.	.	.	.		.		.
D_n	$F^*(D_n)$	$F_n = \frac{F^*(D_n)}{\sum_k F^*(D_k)}$	$(1-\epsilon)F_n$		$(1-\epsilon)F_n$		$(1-\epsilon)F_n + \epsilon$

obtainable from the normal likelihood of the EIFs treated as sufficient statistics, following

$$\frac{1}{\sqrt{n}} \sum_i \phi(D_i, \beta, F-\beta) \sim \text{Normal}\{0, V(\beta, F-\beta)\}, \tag{2}$$

where $\phi(D_i, \beta, F-\beta)$ means the same function as in the unrestricted problem but where now the restrictions are inserted. This essentially amounts to reducing the data of the unrestricted problem to only the data involved in the EIFs ϕ. This reduction can often lead also to the likelihood (2) having a relatively simpler dependence on the nuisance parameters $F-\beta$.

To demonstrate, consider the classic example to estimate the regression parameter β in $E(Y \mid X, \beta) = g(X, \beta)$ from a random sample of observations $(X_i, Y_i), i = 1, ..., n$. Supposing first that X_i takes $1, ..., K$ levels, the restricted model ties together the conditional means $\mu_k = E(Y_i \mid X_i = k)$, so $\tau^{env} = \{\mu_k : k = 1, ..., K\}$. For each μ_k, the EIF in the unrestricted model can be obtained deductively as $\phi_k(D_i, \mu_k, F-\mu_k) := 1(X_i = k)\{Y_i - \mu_k\}/p_k$, where $p_k = \text{pr}(X_i = k)$. Then the log-likelihood of (2) based on the EIFs is

$$-\frac{1}{2} \sum_k \log\left(\frac{\sigma_k^2}{p_k}\right) - \frac{n}{2} \sum_k \frac{\{\bar{\phi}_k(\beta)\}^2}{(\sigma_k^2/p_k)} \tag{3}$$

where $\bar{\phi}_k(\beta) := \frac{1}{n}\sum_i 1(X_i = k)\{Y_i - g(k, \beta)\}/p_k$ and $\sigma_k^2 = \text{var}(Y_i \mid X_i = k)$. From (3), and after taking p_k to be the empirical distribution, one obtains the score, S_β, for β as

$$\begin{aligned} S_\beta &= \sum_k n_k \frac{\bar{\phi}_k(\beta)}{\sigma_k^2} \frac{\partial g(k, \beta)}{\partial \beta} \quad \text{in a discretization-specific expression} \\ &= \sum_i \frac{\{Y_i - g(X_i, \beta)\}}{\sigma^2(X_i)} \frac{\partial g(X_i, \beta)}{\partial \beta} \quad \text{in a discretization-invariant expression.} \end{aligned} \tag{4}$$

Because the score for the nuisance parameters σ_k^2 from (3) is orthogonal to S_β in that likelihood, the score in (4) is also the efficient score. This produces the EIF in the restricted model to be proportional to (4). This agrees with the well known result (e.g., van der Vaart (2000) example 25.28) derived by standard methods. Here, therefore, the ability to derive deductively the EIF ϕ under working models for $\sigma^2(X_i)$, suggests the ability to also derive deductively the EIF for β in the restricted model.

1.3. *Practical Implementation of the Approach (on Comment 4)*

We agree with Luedtke, Carone, and van der Laan that solving the EIF with a working model need not be the only criterion for choosing an estimator. It is for this reason that in the main article, we suggest (criterion (9) in step 2 of the method) that, when faced with many possible solutions, one can focus on those resulting in estimators that have small empirical variance. An interesting idea arising from the discussants' comment 4 is that one may intentionally select to not solve exactly the EIF in order to focus even more on overall accuracy of the estimation. Such methods would also benefit from a deductive approach since they may be even less tractable analytically.

An additional issue that relates to practical implementation is the derivation of the estimand $\tau\{F_{(D_i,\epsilon)}\}$ at perturbed distributions $F_{(D_i,\epsilon)}$. This derivation too can be facilitated by discreteness. Suppose, for example, that an original (possibly continuous) working model F^* is discretized, as in the Table below, by assigning to each data point D_i, mass F_i proportional to the likelihood $F^*(D_i)$.

Suppose also that the estimand τ can be computed as a function $\tau\{[F_1, ..., F_n], [D_1, ..., D_n]\}$ for any discrete distribution $[F_1, ..., F_n]$ on points $[D_1, ..., D_n]$. Then, because the ϵ-contamination of the discrete distribution F to any point D_i is also a discrete distribution (see Table 1 above), the Gateaux derivative-based EIF $\phi(D_i, F)$ is derivable based on the function τ as

$$\begin{aligned} \phi(D_i, F) \;\simeq\; &(\tau\{[F_{1,i}, ..., F_{n,i}], [D_1, ..., D_n]\} \\ &- \tau\{[F_1, ..., F_n], [D_1, ..., D_n]\})/\epsilon \\ \text{where } F_{k,i} &:= (1-\epsilon)F_k + \epsilon \cdot 1(k = i), \end{aligned}$$

for appropriate ϵ. Such discretization may not always be possible or desirable (e.g., see next section), but it suggests there can be generalizable ways of deriving the perturbed estimands.

1.4. *Detecting Irregularities (on Comment 5)*

As the discussants say in their fifth comment - we have indeed focused on estimands for which an EIF exists but has unknown functional form (see Section 2.1 of original article). It is certainly of interest to supplement the paper's algorithms with an algorithm that can determine whether an EIF actually exists to begin with, and it is useful to consider how such lines of work might look like.

Consider again the discussants' example of the "exceptional law," in which the estimand is not pathwise differentiable and an EIF does not exist. A first observation would be that this estimand can be expected to be nondifferentiable because it is defined through an indicator function. That observation can be countered, however, as follows: "But the median in a discrete distribution also depends on indicator functions. So what is it that creates a complication for the exceptional law but not for the median, and can this complication be detected?"

The median of a discrete distribution is, theoretically, not differentiable either, since small perturbations on one direction may not change the median but perturbation on the other direction may change it. However, when the sample space is dense enough (e.g., as would be a discretization of a useful continuous model) and when the perturbation is not too small, then: (a) the structure of the estimand includes many indicator functions and numerical derivatives become similar across different paths; and (b) the structure in (a) can be empirically revealed in the variation of the numerically obtained $\phi(D_i)$ for different data points D_i. In contrast, for the estimand of the exceptional law, when the sample space is large, then: (a') the structure of the estimand still includes only one indicator function, and smoothening of the discontinuity does not seem to occur; but (b') it seems possible that the problematic structure in (a') may also be empirically revealable if we can observe that for many perturbations (e.g., small contaminations on different D_is) the numerical derivative is exactly zero - while for many other perturbations (e.g., on other D_is) the numerical derivative is not zero. This does not mean that all irregularities are detectable (such certainty is rarely met in science) but it indicates that certain worrisome types of irregularities are detectable.

2. Stephens

Stephens has raised a number of interesting and useful points on comparing the proposed approach to others, and on model checks and other criteria for choosing between competing deductive estimators.

2.1. *Deductive versus Nondeductive Estimators*

Stephens is correct that our focus on deduction has been the derivation of the EIF, and this is because we considered that derivation to be a most common challenge in semiparametric inference. As Luedtke, Carone, and van der Laan pointed out, this discussion opens up the question of how to deductively derive other challenging parts of the process, such as, for example, how to determine whether an EIF even exists when this is not as clear to begin with.

In relation to comparing our proposed method to the TMLE (Stephens' second paragraph in the comments), we note that, in contrast to the proposed method, the TMLE is not generally produced in a single solution step, unless the working model is of a particular generalized linear model class. Moreover, present formulations of the TMLE *require* knowledge of the EIF (e.g., to determine a "clever covariate"), whereas our method is designed specifically to deduce (as opposed to refer to) this knowledge directly.

In relation to the (generally) incompatible estimator (5), Stephens suggests that incompatibility was purposely introduced to induce double robustness and efficiency. We do not know the intention of incompatibility, but we observe that it is not necessary for either double robustness of local efficiency, since both, the (nondeductive) TMLE, and the deductive proposed estimator enjoy both such properties while also being compatible.

2.2. *Role of Model checks and Other Criteria*

We agree with Stephens that effort should be made in order for the working model to be appropriate. Such effort is needed for most other uses of a working model. For our purpose, if the working model produces more than one estimators that solve the EIF, then a reasonable approach is to select one that minimizes the empirical variance as suggested by criterion (9) of step 2 of the method. The empirical variances in Table 1 of the article are calculated by the jackknife and are expected to be close to the sandwich variances.

We thank again all discussants for a very useful exchange. We also thank Michael Rosenblum for helpful discussions.

References

van der Laan, M. J. and Rose, S. (2011). *Targeted Learning*. New York: Springer.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press.

Received June 2014. *Revised December* 2014.
Accepted January 2015.