

The International Journal of Biostatistics

Volume 3, Issue 1

2007

Article 4

A Doubly Robust Censoring Unbiased Transformation

Daniel Rubin, *University of California, Berkeley*

Mark J. van der Laan, *Division of Biostatistics, School of
Public Health, University of California, Berkeley*

Recommended Citation:

Rubin, Daniel and van der Laan, Mark J. (2007) "A Doubly Robust Censoring Unbiased Transformation," *The International Journal of Biostatistics*: Vol. 3: Iss. 1, Article 4.

DOI: 10.2202/1557-4679.1052

A Doubly Robust Censoring Unbiased Transformation

Daniel Rubin and Mark J. van der Laan

Abstract

We consider random design nonparametric regression when the response variable is subject to right censoring. Following the work of Fan and Gijbels (1994), a common approach to this problem is to apply what has been termed a censoring unbiased transformation to the data to obtain surrogate responses, and then enter these surrogate responses with covariate data into standard smoothing algorithms. Existing censoring unbiased transformations generally depend on either the conditional survival function of the response of interest, or that of the censoring variable. We show that a mapping introduced in another statistical context is in fact a censoring unbiased transformation with a beneficial double robustness property, in that it can be used for nonparametric regression if either of these two conditional distributions are estimated accurately. Advantages of using this transformation for smoothing are illustrated in simulations and on the Stanford heart transplant data.

KEYWORDS: survival analysis, censoring unbiased transformations, nonparametric regression, imputation

1 Introduction

Random design nonparametric regression is a popular subject of study in the statistical literature, because informally the regression function provides the best prediction of a response given covariates. Nonparametric methods are often required because modern datasets are complicated enough so that any assumed parametric or semiparametric model would almost certainly be misspecified.

When the responses are subject to right censoring, additional complications arise in most smoothing problems. Let $X = (W, Y)$ denote the possibly unavailable covariate and response pair, and suppose our interest is in estimating the regression function $m(w) = E[Y|W = w]$, for the purpose of being able to predict response values at different vector-valued covariate levels. But instead of observing an i.i.d. sample $\{X_i\}_{i=1}^n$, assume that we only can observe each survival time Y_i up to a random censoring time C_i . Formally, consider observing an i.i.d. sample $\{O_i\}_{i=1}^n$, where

$$O = (W, \Delta = 1(Y \leq C), \tilde{Y} = Y \wedge C),$$

and let $\bar{F}(\cdot|W)$ and $\bar{G}(\cdot|W)$ denote the conditional survival functions of the desired response Y and censoring time C given the covariates W . For convenience, we will assume that the survival and censoring times are continuous random variables, although this is unnecessary.

Throughout this work we will also make the standard assumption that the response and censoring time are conditionally independent given the covariates W , or that

$$\{Y \perp C|W\}. \tag{1}$$

In fact, the regression function $m(W) = E[Y|W]$ is often unidentifiable from such observed data. Consider the case of a censoring time corresponding to a study endpoint, which the true response time Y may sometimes exceed. Nothing can be known about the survival time distribution beyond this endpoint, and hence the regression function will be unidentifiable. For regression to remain a worthwhile endeavor with right censored data, the response must often first be transformed. We can consider truncating the responses at some value τ and estimating the regression function $w \rightarrow E[Y \wedge \tau|W = w]$. The response is also often transformed to the log scale, and in this case we would consider the parameter of interest to be the function $w \rightarrow E[\log(Y)|W = w]$. To simplify notation, we will assume that such transformations have already been incorporated into the response and censoring times Y and C , and continue to let $m(w) = E[Y|W = w]$ denote the desired regression function.

A popular approach to prediction with right censored data is to replace the possibly unavailable responses $\{Y_i\}_{i=1}^n$ with surrogate values $\{Y^*(O_i)\}_{i=1}^n$ using an appropriate mapping $Y^*(\cdot)$ of the observed data, and then enter the imputed data $\{W_i, Y^*(O_i)\}_{i=1}^n$ into standard smoothing algorithms. The key requirement is that imputation mapping $Y^*(\cdot)$ is approximately what Fan and Gijbels (1996) term a “censoring unbiased transformation,” meaning that

$$E[Y^*(O)|W] = E[Y|W] = m(W). \quad (2)$$

The motivation behind such a requirement is that adaptive smoothing techniques would ideally still be able to estimate the regression function with imputed response data under (2), due to $Y^*(O)$ having the correct conditional mean structure.

Unfortunately, censoring unbiased transformations generally depend on nuisance parameters, and cannot be directly applied to the observed data. As will be discussed in the following section, existing censoring unbiased transformations that have been proposed for right censored data fall into two categories.

1. Transformations depending on the conditional survival function $\bar{F}(\cdot|W)$, or a functional of this conditional distribution. To apply such transformations, one would first have to construct a preliminary estimate for the conditional distribution of the response Y given covariates W .
2. Transformations depending on the censoring mechanism, or the function $\bar{G}(\cdot|W)$. Applying such transformations thus necessitates estimating this censoring mechanism, which is not directly related to the parameter of interest $m(W) = E[Y|W]$.

In this paper we propose using a *doubly robust* censoring unbiased transformation, to be given in section 2.3. While this mapping has been introduced in statistical contexts unrelated to nonparametric regression, we give the new result that it will have the correct conditional mean structure as in (2) if at least one of the two nuisance parameters $\bar{F}(\cdot|W)$ or $\bar{G}(\cdot|W)$ is correctly specified. Hence, the doubly robust mapping gives the data analyst two chances to form a valid censoring unbiased transformation, and this property can be utilized to form enhanced smoothing procedures.

We give an overview of existing censoring unbiased transformations and present the doubly robust transformation in section 2. Advantages of the doubly robust procedure are highlighted in simulations and with the Stanford heart transplant data in sections 3 and 4. Doubly robust censoring unbiased

transformations can be utilized for more general types of censored responses than arise in the right censored data structure. Rubin and van der Laan (2006) discuss the applicability of such imputation schemes to regression problems with missing responses, in causal inference problems, and with current status data, and give an abstract treatment of double robustness for more general censored data structures.

2 Censoring Unbiased Transformations

Our overview of censoring unbiased transformations in this section is partially adapted from the discussion of Fan and Gijbels (1996).

2.1 The Buckley-James Transformation

One of the earliest censoring unbiased transformations was the Buckley-James (1979) mapping, given by,

$$Y^*(O) = \Delta Y + (1 - \Delta)Q_{\bar{F}}(W, C), \quad (3)$$

for

$$Q_{\bar{F}}(w, y) = E[Y|W = w, Y > y] = \frac{1}{\bar{F}(y|W = w)} \int_y^\infty y dF(y|W = w). \quad (4)$$

This transformation is the best predictor of the original response, in that it minimizes $E|Y^*(O) - Y|^2$ among all censoring unbiased transformations $Y^*(\cdot)$, leading Fan and Gijbels to note that it can be regarded as the “best restoration.” The nuisance parameter required to evaluate (3) is the function $Q_{\bar{F}}(W, \cdot)$, which is in turn a functional of the conditional survival function $\bar{F}(\cdot|W)$ associated with the response Y . The original proposal by Buckley and James for estimation of this nuisance parameter depended on strong assumptions, such as the linearity of the true regression function $m(w) = E[Y|W = w]$. Fan and Gijbels (1994, 1996) considered more adaptive estimates of $Q_{\bar{F}}$, that corresponded to local average estimators and locally linear estimators.

2.2 Transformations Depending on the Censoring Mechanism

While the Buckley-James mapping given in (3) depends on $\bar{F}(\cdot|W)$, other censoring unbiased transformations have been proposed that instead depend

only the censoring mechanism. For example, Koul et al. (1981) considered the mapping

$$Y^*(O) = \frac{Y\Delta}{\bar{G}(Y|W)}, \quad (5)$$

which has also been termed the *inverse probability of censoring weighted* (IPCW) mapping. To evaluate this transformation, one would first have to estimate the conditional distribution of the censoring time given the covariates. It is frequently the case that the censoring time is completely independent of both the response time Y and covariates W , as might occur if censoring is caused by the end of a study. In this setting, the conditional survival function $\bar{G}(\cdot|W) = \bar{G}(\cdot)$ could be estimated efficiently with the Kaplan-Meier estimator

$$\hat{\bar{G}}(c) = \prod_{\{i: \tilde{Y}_i \leq c\}} \left(1 - \frac{1}{\#\{j: \tilde{Y}_j \geq \tilde{Y}_i\}} \right)^{1-\Delta_i}. \quad (6)$$

Zheng (1987) studied more general censoring unbiased transformations, which for nonnegative and continuous response and monitoring times took the form,

$$Y^*(O) = \int_0^{\tilde{Y}} \frac{1}{\bar{G}(c|W)} dc + \int_0^{\tilde{Y}} \frac{d(W, c)}{\bar{G}(c|W)} d\bar{G}(c|W) + (1 - \Delta)d(W, C).$$

Fan and Gijbels (1994) considered this mapping with

$$d(w, c) = \frac{\alpha c}{\bar{G}(c|W = w)},$$

for different choices of α . They noted that the IPCW transformation given in (5) corresponded to $\alpha = -1$, while a transformation given by Leurgans (1987) corresponded to $\alpha = 0$. Fan and Gijbels proposed to instead apply the mapping with the data-dependent choice of

$$\hat{\alpha} = \min_{\{i: \Delta_i=1\}} \frac{\int_0^{Y_i} \{\bar{G}(c|W_i)\}^{-1} dc - Y_i}{Y_i \{\bar{G}(Y_i|W_i)\}^{-1} - \int_0^{Y_i} \{\bar{G}(c|W_i)\}^{-1} dc},$$

after constructing an estimator of $\bar{G}(\cdot|W)$.

2.3 A Doubly Robust Censoring Unbiased Transformation

The censoring unbiased transformations considered in sections 2.1 and 2.2 respectively depend on the nuisance parameters $\bar{F}(\cdot|W)$ and $\bar{G}(\cdot|W)$. Simulations in the following section will show that a poor preliminary estimator for

$\bar{F}(\cdot|W)$ can degrade the performance of regression based on the Buckley-James transformation, while a poor preliminary estimator for $\bar{G}(\cdot|W)$ can degrade the performance of regression based on the transformations given in section 2.2.

In fact, it is possible to construct a censoring unbiased transformation $Y^*(O)$ that will have the correct conditional mean structure if *either* $\bar{F}(\cdot|W)$ or $\bar{G}(\cdot|W)$ is well approximated. This “doubly robust” transformation provides a clear advantage over the existing procedures described in sections 2.1 and 2.2, because with such a transformation one only needs to solve at least one of two function approximation problems. For $Q_{\bar{F}}(\cdot, \cdot)$ defined as in (4), we propose using the censoring unbiased transformation given by,

$$\begin{aligned} Y^*(O) &= Y_{\bar{F}, \bar{G}}^*(O) \\ &= \frac{Y\Delta}{\bar{G}(Y|W)} + \frac{Q_{\bar{F}}(W, C)(1 - \Delta)}{\bar{G}(C|W)} - \int_{-\infty}^{\bar{Y}} \frac{Q_{\bar{F}}(W, c)}{\bar{G}^2(c|W)} dG(c|W), \end{aligned} \quad (7)$$

possessing the double robustness property formalized in the following theorem. The theorem is proven in the appendix.

Theorem 1. *Suppose the conditional independence assumption (1) holds, that Y and C are continuous random variables, and that the conditional distribution of $\{C|W\}$ has a conditional density $g(\cdot|W)$. Assume that $Y \leq \tau < \infty$ for some τ and that $\bar{F}_1(\tau|W) = 0$ with probability one for some conditional survival function $\bar{F}_1(\cdot|W)$. Suppose further that $\bar{G}_1(\tau|W) \geq \epsilon > 0$ for some ϵ and conditional survival function $\bar{G}_1(\cdot|W)$, with corresponding conditional density $g_1(\cdot|W)$. Assume that $g_1(\cdot|W = w)$ is absolutely continuous with respect to $g(\cdot|W = w)$ for all w . We will use the convention that $Q_{\bar{F}_1}(w, y)$ is set to zero if $\bar{F}_1(y|W = w) = 0$. Then,*

$$E[Y_{\bar{F}_1, \bar{G}_1}^*(O)|W] = E[Y|W] \text{ if } \bar{F}(\cdot|W) = \bar{F}_1(\cdot|W) \text{ or } \bar{G}(\cdot|W) = \bar{G}_1(\cdot|W). \quad (8)$$

The statistical literature concerning double robustness is primarily related to estimation in semiparametric models, and is discussed in great detail in van der Laan and Robins (2003). In fact, the transformation (7) can be seen as a special case of the doubly robust mappings in chapter 3 of this work, used to construct estimating equations for regular parameters with censored data. Theorem 2.1 of van der Laan and Robins implies the weaker form of (8) that

$$E[Y_{\bar{F}_1, \bar{G}_1}^*(O)] = E[Y] \text{ if } \bar{F}(\cdot|W) = \bar{F}_1(\cdot|W) \text{ or } \bar{G}(\cdot|W) = \bar{G}_1(\cdot|W). \quad (9)$$

Later work by van der Laan and Dudoit (2003) used the property (9) for doubly robust model selection with censored data, and for constructing M -estimates of irregular parameters. The theoretical novelty in our work lies in

the result that the doubly robust mapping $Y_{\bar{F}_1, \bar{G}_1}^*(O)$ not only has the correct mean if one of $\bar{F}_1(\cdot|W)$ or $\bar{G}_1(\cdot|W)$ is correctly specified as in (9), but also the correct conditional mean given observed covariates as in (2). The benefit of this seemingly minor discovery is the implication that the doubly robust mapping can be used advantageously for nonparametric regression in censored data problems, with tangible benefits compared to the existing methodology.

3 Simulations

We assessed the quality of the doubly robust transformation through simulations, and compared its performance to that of the Buckley-James transformation (3) and the IPCW transformation (5).

Implementing the regression procedures based on these transformations required estimates of the function $Q(w, y) = E[Y|W = w, Y > y]$, the censoring mechanism $\bar{G}(\cdot|W)$, and choosing a smoothing procedure to use with the imputed data $\{W_i, Y^*(O_i)\}_{i=1}^n$. For the smoothing procedure, we used the **smooth.spline()** function in the R language, which fit a cubic smoothing spline to the imputed responses. In all simulations we fit the censoring mechanism through the Kaplan-Meier estimator (6), which has been the standard recommendation in the statistical literature related to censoring unbiased transformations.

We fit $Q(\cdot, \cdot)$ through a nearest neighbor estimate that was similar to that proposed by Fan and Gijbels (1994, 1996). We estimated $Q(w, y)$ by taking the mean of the k uncensored responses greater than y , whose covariate value W was closest to w . If less than k such responses were available, we took the average of these responses. If no such responses were available, we estimated $Q(w, y)$ by y itself. Like Fan and Gijbels, we chose the number of nearest neighbors k by implementing the Buckley-James transformation for each $k \leq \frac{n-1}{2}$ to form imputed data $\{W_i, Y_k^*(O_i)\}_{i=1}^n$, evaluated the squared error leave-one-out cross-validation criterion for the **smooth.spline()** regression fit to this data, and selected the k minimizing this criterion.

Our first set of simulations demonstrated that regression based on the doubly robust censoring unbiased transformation could indeed adapt to the shape of a regression curve, if given sufficient data. For univariate covariates W , errors ϵ , and censoring times C generated independently, we generated

$n = 200$ observations O through the following mechanism.

$$\begin{aligned}
 W &\sim U(0, 1) \\
 \epsilon &\sim 2(\text{Beta}(4, 4) - \frac{1}{2}) \\
 C &\sim \text{exponential}(\frac{1}{2}) - 1 \\
 Y &= m(W) + \epsilon \\
 O &= (W, \Delta = 1(Y \leq C), \tilde{Y} = Y \wedge C)
 \end{aligned} \tag{10}$$

We generated such data using four choices for the regression function m , corresponding to linear, quadratic, sigmoidal, and oscillating functions. For such data, the censoring times were indeed independent of the covariates and response times, so we expected the Kaplan-Meier estimator to be a good fit. We also expected no problems with the fit for Q , as nearest neighbor methods typically do not break down with univariate data. For the four choices of regression function $m(W)$, 52%, 44%, 51%, and 52% of the responses were censored. The results are displayed in Figure 1, and show the doubly robust procedure could accurately approximate these four smooth curves.

In a second set of simulations, we compared the doubly robust transformation with the Buckley-James and IPCW transformations. We generated $n = 200$ replicates of O according to the following mechanism, where the true regression function was simply the identity function $m(W) = W$.

$$\begin{aligned}
 W &\sim U(0, 1) \\
 \epsilon &\sim 2(\text{Beta}(4, 4) - \frac{1}{2}) \\
 \{C|W \leq \frac{1}{2}\} &= +\infty \text{ (meaning no censoring)} \\
 \{C|W > \frac{1}{2}\} &\sim \text{exponential}(1) - 1 \\
 Y &= W + \epsilon \\
 O &= (W, \Delta = 1(Y \leq C), \tilde{Y} = Y \wedge C).
 \end{aligned} \tag{11}$$

The censoring mechanism here depended on the covariates, as censoring never occurred if the covariate W did not exceed $\frac{1}{2}$. Hence, the assumptions for the Kaplan-Meier estimate of $\bar{G}(\cdot|W)$ were violated. In this simulation, 38% of the responses were censored. From the results in the top row of Figure 2, we see that the regressions using the Buckley-James and doubly robust transformations accurately fit the regression line, while the IPCW estimator behaved erratically.

In a final set of simulations, we considered covariates not only consisting of the univariate W , but also of a $\{0, 1\}$ random variable V . We generated $n = 400$ replicates as follows, where again the regression function $E[Y|W] = m(W) = W$ was simply the identity function.

$$\begin{aligned}
 W &\sim U(0, 1) \\
 V &\sim \text{Bernoulli}\left(\frac{1}{2}\right) \\
 \epsilon &\sim 2(\text{Beta}(4, 4) - \frac{1}{2}) \\
 \{C|V = 0\} &= +\infty \text{ (meaning no censoring)} \\
 \{C|V = 1\} &\sim \text{exponential}\left(\frac{1}{3}\right) - 2 \\
 Y &= W + 2\left(V - \frac{1}{2}\right) + \epsilon \\
 O &= (W, V, \Delta = 1(Y \leq C), \tilde{Y} = Y \wedge C)
 \end{aligned} \tag{12}$$

We considered correctly modeling the censoring mechanism, so that we set $\bar{G} = 1$ for all observations with $V = 0$ in the IPCW and doubly robust transformations, while using the Kaplan-Meier estimator for observations with $V = 1$. In practice, one would expect to notice with $n = 400$ data points if censoring never occurred at a certain level of a binary covariate, so such a fit might be fairly realistic. However, we did not correctly model the nuisance parameter $Q(w, y) = E[Y|W, Y > y]$, because we fit the function while ignoring the covariate V . We imagine that such an estimate could also be fairly common in practice, because if a univariate smoother was desired for a specific covariate, one might be reluctant to adjust for additional covariates. The problem with ignoring V in the fit of Q was that the conditional independence assumption $\{Y \perp C|W\}$ did not hold, but rather the conditional independence $\{Y \perp C|W, V\}$. This was due to the event $\{V = 1\}$ being associated with both large Y values and small censoring times. Under this censoring mechanism, 32% of the responses were censored. The results displayed in the bottom row of Figure 2 show that the IPCW and doubly robust mappings led to fairly accurate fits of the regression line, while the Buckley-James transformation led to a severe underestimate of this line.

Therefore, simulated data from the mechanisms in (11) and (12) show that a misspecified censoring mechanism $\bar{G}(\cdot|W)$ can degrade the performance of the IPCW transformation, while a misspecified $Q_{\bar{F}}(W, \cdot)$ can degrade the Buckley-James transformation. In colloquial jargon, the Buckley-James and IPCW transformations put all of their eggs in one basket. The doubly robust

transformation can be applied whenever either of the Buckley-James or IPCW function approximation problems has been solved, even if the data analyst is not sure which of $\bar{G}(\cdot|W)$ or $Q_{\bar{F}}(W, \cdot)$ has been well-approximated, and is in this sense a superior censoring unbiased transformation.

4 Stanford Heart Transplant Data

We applied censoring unbiased transformations to the Stanford heart transplant data, which has been studied by Miller and Halpern (1982), Doksum and Yandell (1982), and Fan and Gijbels (1994), and is somewhat of a benchmark dataset for right censored regression methods. In this study, patients receiving a heart transplant were followed until either death or a single study endpoint. Among other covariates, the age of each patient at transplantation was recorded, and there appears to have been medical interest in determining how heart transplantation risk was associated with age. For comparison with previous analyses, we considered the dataset to consist of only the 157 patients for which there was information on the tissue type, 55 of whom had censored survival times, and we used the $\log_{10}(\text{days})$ time scale for the survival and censoring times.

While Miller and Halpern and Doksum and Yandell considered linear and quadratic fits of $E[\log_{10}(\text{Days of Survival})|\text{Age}]$ based on various regression models, Fan and Gijbels attempted to fit this function through adaptive smoothing. After fitting the Buckley-James nuisance parameter $Q_{\bar{F}}(\text{Age}, \cdot)$ with a local averaging estimator, and then applying a local linear smoother to estimate the regression function from the transformed data, Fan and Gijbels concluded that their fitted curve

...reflects the fact that for earlier age, the log-survival time is nearly independent of age, but at later age it decreases linearly with aging.

From the smoothed curve, they suggested the relationship

$$E[\log_{10}(\text{Days of survival})|\text{Age in years}] = 2.74 - 0.078(\text{Age} - 48)_+, \quad (13)$$

which is shown in the top left panel of Figure 3. Commenting on the utility of smoothing methods for censored data, in comparison to the linear and quadratic fits that had been implemented previously for the Stanford heart data, Fan and Gijbels concluded about (13) that

...such a relation appears to be new. The result supports our intuition and moreover, gives a deeper insight into the heart transplantation risk at various ages. In comparison with previous studies by,

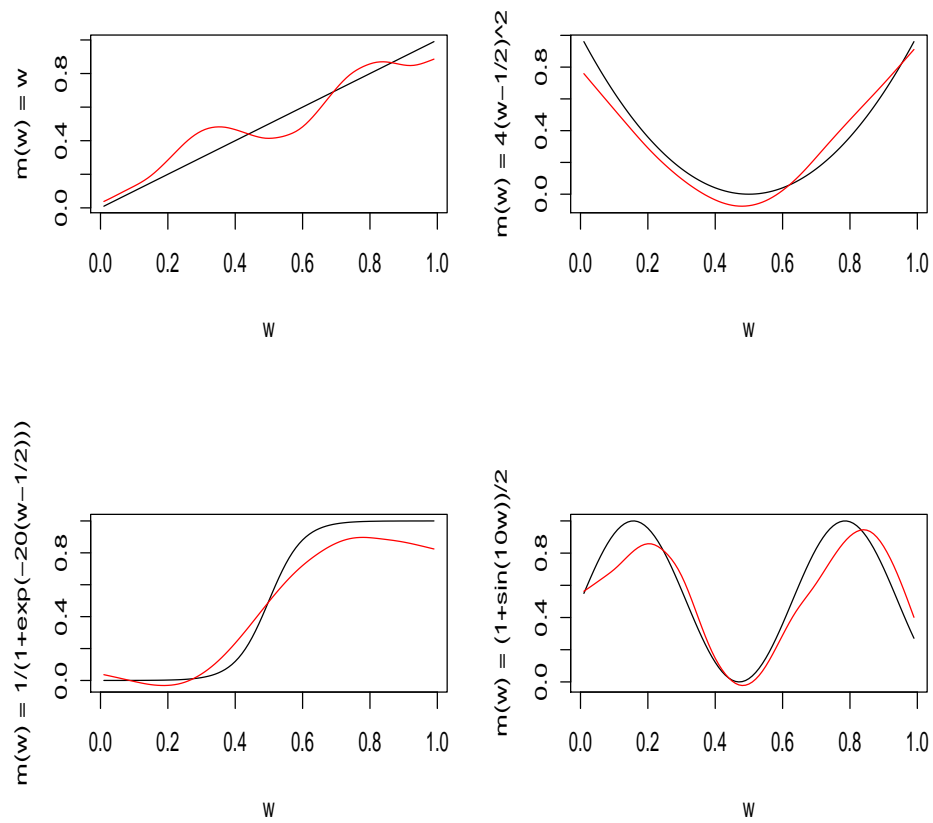


Figure 1: Doubly robust fits of four regression functions, for data generated as in (10). Black lines indicate the regression function, and red lines indicate the fit.

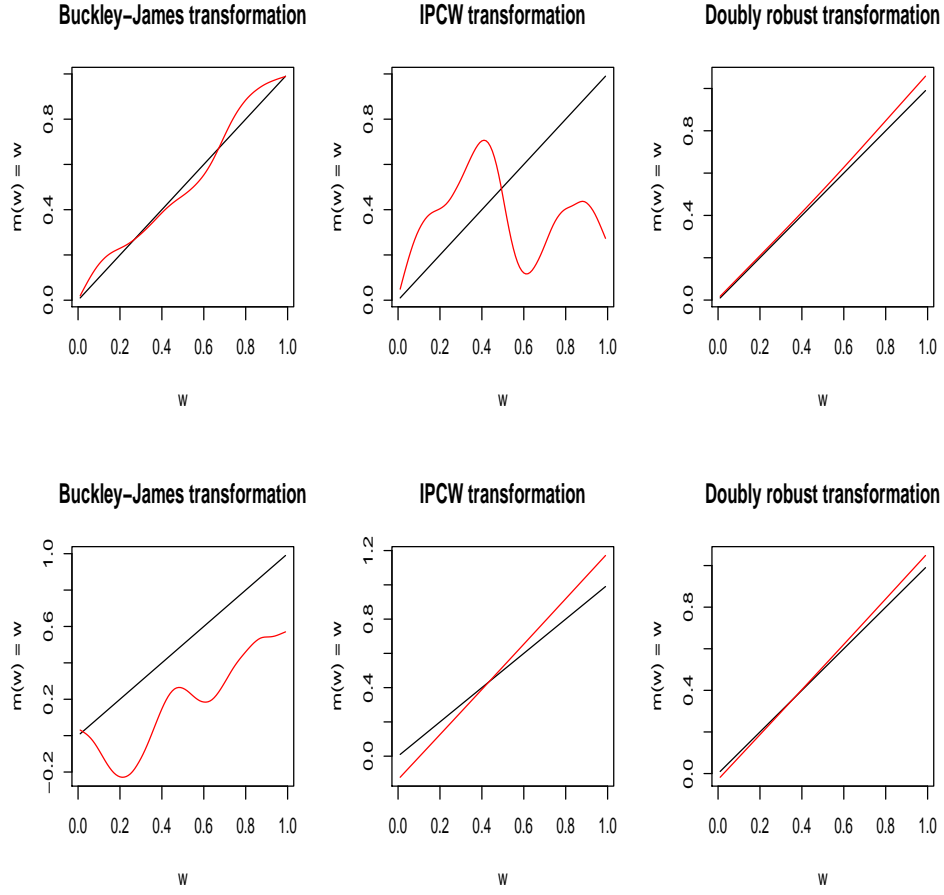


Figure 2: The first row gives the fits for data generated as in (11), and the second row for the data generated as in (12). Black lines indicate the regression function, and red lines indicate the fits.

for example, Miller and Halpern (1982) and Doksum and Yandell (1982), our analysis gives a more precise description of the data structure.

In analyzing this heart transplantation data, we considered slightly modifying the parameter of interest to the function

$$E[\log_{10}(\text{Days of survival}) \wedge \tau | \text{Age in years}], \quad (14)$$

for $\tau = 3.26$. As a practical matter, we simply truncated the 21 values of $\log_{10}(\tilde{Y})$ exceeding τ to τ , set the censoring indicator Δ to one for these observations, and then attempted to estimate the regression function as if the original observations had been this transformed data. Our motivation was the fact that a regression function with right censored data can only be estimated when the response time is sufficiently small so that given any covariate values, the response has a nontrivial chance of being uncensored. The Kaplan-Meier fit for the censoring time distribution suggested this did not hold, as the fit gave extremely small values of $\hat{G}(\tilde{Y}_i)$ for some observed data points, and zero for one data point. The level $\tau = 3.26$ in (14) corresponded to truncating survival at the five year mark, and the refit Kaplan-Meier curve gave values of $\hat{G}(\tilde{Y}_i)$ no smaller than 0.40 for all observed data points. Truncation as in (14) is a useful tactic in many applied regression problems with right censored data, because it allows us to handle identifiability problems, while retaining an interpretable parameter of interest.

We first estimated the regression function (14) with the Buckley-James transformation, estimating the nuisance parameter $Q_{\tilde{F}}(\text{Age}, \cdot)$ with the nearest neighbor method described in the previous section. The cross-validation method previously described selected $k = 6$ nearest neighbors to use for this nuisance parameter estimate. After obtaining the resulting imputed response values, we again used the **smooth.spline()** procedure to estimate the regression function. The resulting curve fit is displayed in the top right panel of Figure 3. Notice that the fit closely resembles the suggested relationship of Fan and Gijbels, in that the curve is roughly constant (very slightly increasing) until between the ages of 40 and 50, when it begins to decrease linearly. Our Buckley-James fit appears slightly smaller than the Fan and Gijbels piecewise linear function, possibly due to our truncation scheme.

We next applied the IPCW censoring unbiased transformation to the Stanford heart transplant data. We used the Kaplan-Meier estimator (6) to fit the censoring mechanism $\bar{G}(\cdot | \text{Age})$, which ignored the age values and fit a marginal survival function. Using the **smooth.spline()** once more with the transformed

responses, we obtained an estimate of the regression function (14). The resulting fit is presented in the bottom left panel of Figure 3, and appears very different from Fan and Gijbels' suggested relationship, or our regression fit based on the Buckley-James transformation. In fact, using the IPCW transformation would have led us to the counterintuitive conclusion that a patient's expected log survival time actually slightly increases with age.

Thus, two popular censoring unbiased transformations led to contradictory interpretations of how heart transplantation risk was associated with age. The two transformations respectively depended on accurate estimation of the conditional distributions of the survival and censoring times, given age. On the surface, it does not appear either of these function approximation problems should have been difficult to solve. Nearest neighbor methods are generally reliable in low dimensions, and we had no particular reason to distrust our estimate of $Q(\text{Age}, \cdot)$. Because censoring was caused by the end of the study, domain knowledge also suggested that the censoring time distribution did not depend on the age of the subject, and hence that the Kaplan-Meier estimator of the censoring mechanism should have been reliable. Indeed, one can verify that a Cox model for the censoring distribution does not show any significance for age.

Because the doubly robust mapping was immune to misspecification of one of $Q_{\bar{F}}(\text{Age}, \cdot)$ or $\bar{G}(\cdot|\text{Age})$, it served as a data analytic tool to resolve the inconsistencies stemming from the currently used censoring unbiased transformations. That is, if either the Buckley-James or IPCW fits were accurate, we would have expected the doubly robust fit to also be accurate. Again using the nearest neighbor fit for $Q_{\bar{F}}(\text{Age}, \cdot)$, the Kaplan-Meier estimator $\bar{G}(\cdot|\text{Age})$, and the **smooth.spline()** function with the transformed responses, we fit the doubly robust estimator to the heart transplant data. The resulting curve in the bottom right panel of Figure 3 in fact looks almost identical to the curve based on the Buckley-James transformation. This seems to support the conclusion that $Q_{\bar{F}}(\text{Age}, \cdot)$ and not $\bar{G}(\cdot|\text{Age})$ has been well approximated, and give further credence to the relationship between transplantation risk and age suggested by Fan and Gijbels.

Of course, both $Q_{\bar{F}}(\text{Age}, \cdot)$ and $\bar{G}(\cdot|\text{Age})$ could have been misspecified, and factors other than misspecification of the nuisance parameters in a censoring unbiased transformation can also contribute to inaccurate regression fits. Such factors might include violations of the i.i.d. assumption or the conditional independence assumption (1), or the regression function $m(\text{Age})$ being complex and difficult to estimate even with full and uncensored covariate and response data $\{\text{Age}_i, \text{Survival Time}_i\}_{i=1}^n$.

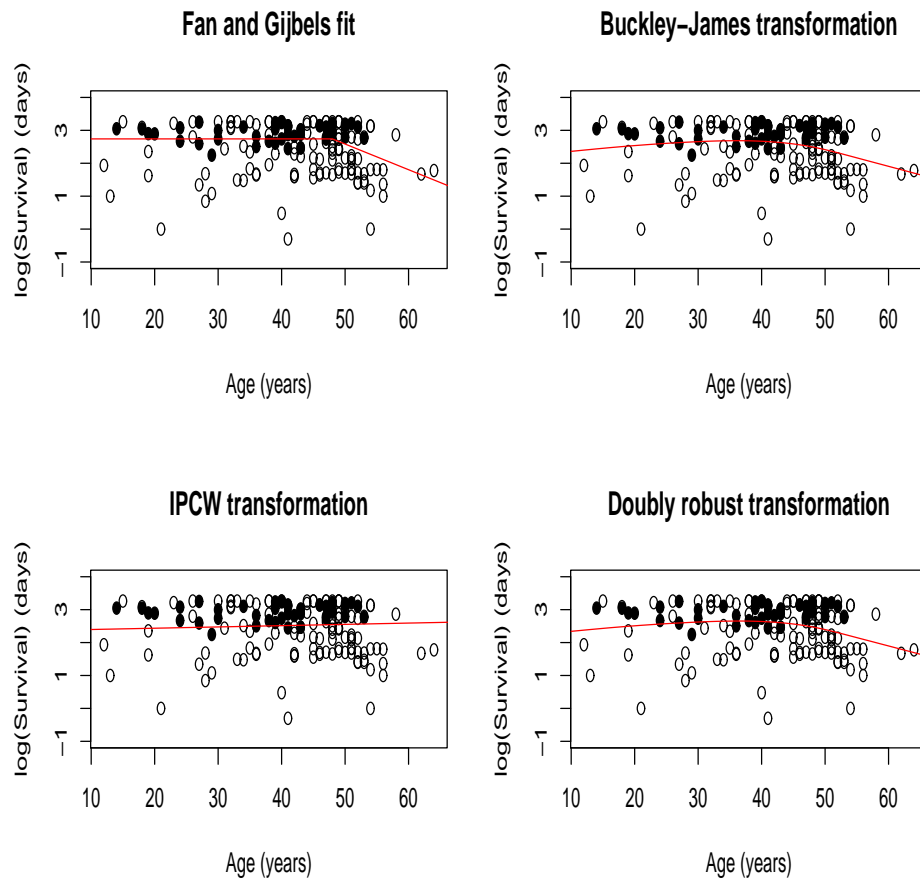


Figure 3: Smoothed Stanford heart transplant data using different censoring unbiased transformations. Open circles represent uncensored observations, while solid circles represent censored observations.

5 Discussion

While the censoring unbiased transformation introduced in this work has the advantage of double robustness, we do not mean to suggest that it should be favorable to existing transformations in all nonparametric regression problems. Like the IPCW transformation but unlike the Buckley-James transformation, the method depends on inverse weighting by the censoring mechanism $\bar{G}(\cdot|W)$, which can lead to outlier problems if inverse weighting by a small quantity. Hence, regression using this method typically depends on the cumbersome step of first applying an artificial truncation to the response variable and parameter of interest. Further, the doubly robust transformation can be more complicated to implement than traditionally used censoring unbiased transformations, because it necessitates two nuisance parameter estimates instead of one, and computing integral with respect to the censoring mechanism $dG(\cdot|W)$.

We expect future work to more carefully explore when our proposed imputation mapping should and should not be utilized. We have not derived consistency, rates of convergence, or the distribution theory for specific estimators based on our censoring unbiased transformation. As with previous transformations, performance will depend on the nuisance parameter estimates and full data algorithm into which we plug the imputed responses, and it remains to be seen how to best make these choices. More extensive simulations should help in this regard.

In spite of the remaining work to be done and our transformation's less desirable qualities, we feel that its double robustness is an appealing property, and provides a valuable addition to the right censored regression toolbox.

Appendix

Proof of Theorem 1. Recall that we use the convention that $Q(W, c) = \frac{1}{\bar{F}_1(c|W)} \int_c^\infty y dF_1(y|W)$ is zero if $\bar{F}_1(c|W) = 0$. This and the theorem assumptions ensure that the resulting conditional expectations below are well defined and finite. We will see in the following proof that the theorem assumptions are most likely excessive, particularly the upper bound τ on the response Y and the almost sure boundedness of $\bar{G}_1(\tau|W)$ away from zero. We write,

$$\begin{aligned} Y^*(O) &= Y_{\bar{F}, \bar{G}}^*(O) \\ &= \frac{Y\Delta}{\bar{G}(Y|W)} + \frac{Q_{\bar{F}}(W, C)(1 - \Delta)}{\bar{G}(C|W)} - \int_{-\infty}^{\tilde{Y}} \frac{Q_{\bar{F}}(W, c)}{\bar{G}^2(c|W)} dG(c|W) \\ &= T_1 + T_2 - T_3. \end{aligned}$$

First observe that,

$$\begin{aligned}
 E[T_1|W] &= E\left[\frac{Y\Delta}{\bar{G}_1(Y|W)}|W\right] \\
 &= E\left[E\left[\frac{Y\Delta}{\bar{G}_1(Y|W)}|W, Y\right]|W\right] \\
 &= E\left[\frac{Y}{\bar{G}_1(Y|W)}P(\Delta = 1|W, Y)|W\right] \\
 &= E\left[\frac{Y}{\bar{G}_1(Y|W)}\bar{G}(Y|W)|W\right] \\
 &= \int_{-\infty}^{\tau} y \frac{\bar{G}(y|W)}{\bar{G}_1(y|W)} dF(y|W).
 \end{aligned} \tag{15}$$

Next note that,

$$\begin{aligned}
 E[T_2|W] &= E\left[\frac{Q_1(W, C)(1 - \Delta)}{\bar{G}_1(C|W)}|W\right] \\
 &= E\left[E\left[\frac{Q_1(W, C)(1 - \Delta)}{\bar{G}_1(C|W)}|W, C\right]|W\right] \\
 &= E\left[\frac{Q_1(W, C)}{\bar{G}_1(C|W)}P(\Delta = 0|W, C)|W\right] \\
 &= E\left[\frac{Q_1(W, C)}{\bar{G}_1(C|W)}\bar{F}(C|W)|W\right] \\
 &= E\left[\frac{\bar{F}(C|W)}{\bar{F}_1(C|W)} \int_C^{\tau} y dF_1(y|W) \bar{G}_1^{-1}(C|W)|W\right] \\
 &= \int_{-\infty}^{\infty} \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \bar{G}_1^{-1}(c|W) \left\{ \int_c^{\tau} y dF_1(y|W) \right\} dG(c|W) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \bar{G}_1^{-1}(c|W) 1(c < y < \tau) y \right\} dF_1(y|W) dG(c|W) \\
 &= \int_{-\infty}^{\tau} y \left\{ \int_{-\infty}^y \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \bar{G}_1^{-1}(c|W) dG(c|W) \right\} dF_1(y|W).
 \end{aligned} \tag{16}$$

Finally, observe that,

$$\begin{aligned}
 E[T_3|W] &= E\left[\int_{-\infty}^{\min(Y, C)} \frac{Q_1(W, c)}{\bar{G}_1^2(c|W)} dG_1(c|W)|W\right] \\
 &= E\left[\int_{-\infty}^{\infty} 1(Y > c) 1(C > c) \frac{Q_1(W, c)}{\bar{G}_1^2(c|W)} dG_1(c|W)|W\right] \\
 &= \int_{-\infty}^{\infty} P(Y > c, C > c|W) \frac{Q_1(W, c)}{\bar{G}_1^2(c|W)} dG_1(c|W)
 \end{aligned}$$

$$\begin{aligned}
 &= \int_{-\infty}^{\infty} P(Y > c|W)P(C > c|W) \frac{Q_1(W, c)}{\bar{G}_1^2(c|W)} dG_1(c|W) \\
 &= \int_{-\infty}^{\infty} \bar{F}(c|W) \bar{G}(c|W) \frac{Q_1(W, c)}{\bar{G}_1^2(c|W)} dG_1(c|W) \\
 &= \int_{-\infty}^{\infty} \frac{\bar{G}(c|W)}{\bar{G}_1^2(c|W)} \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \int_c^{\tau} y dF_1(y|W) \} dG_1(c|W) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \frac{\bar{G}(c|W)}{\bar{G}_1^2(c|W)} \left\{ \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} 1(c < y < \tau) y \right\} dF_1(y|W) \right\} dG_1(c|W) \\
 &= \int_{-\infty}^{\tau} y \left\{ \int_{-\infty}^y \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \frac{\bar{G}(c|W)}{\bar{G}_1^2(c|W)} dG_1(c|W) \right\} dF_1(y|W) \\
 &= \int_{-\infty}^{\tau} y \int_{-\infty}^y \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \frac{\bar{G}(c|W)}{\bar{G}_1^2(c|W)} \frac{dG_1}{dG}(c|W) dG(c|W) dF_1(y|W). \quad (17)
 \end{aligned}$$

Further, note from elementary calculus that for $c < \tau$ (so the demoninator is nonzero),

$$\frac{d}{dc} \left\{ \frac{\bar{G}}{\bar{G}_1}(c|W) \right\} = - \left\{ \frac{1}{\bar{G}_1(c|W)} - \frac{\bar{G}(c|W)}{\bar{G}_1^2(c|W)} \frac{g_1(c|W)}{g(c|W)} \right\} g(c|W). \quad (18)$$

Thus, combining (15), (16), (17), and (18), we see that,

$$\begin{aligned}
 E[Y^*(O)|W] &= E[T_1 + T_2 - T_3|W] = E[T_1|W] + E[T_2|W] - E[T_3|W] \\
 &= \int_{-\infty}^{\tau} y \frac{\bar{G}(y|W)}{\bar{G}_1(y|W)} dF(y|W) \\
 &\quad + \int_{-\infty}^{\tau} y \left\{ \int_{-\infty}^y \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \bar{G}_1^{-1}(c|W) dG(c|W) \right\} dF_1(y|W) \\
 &\quad - \int_{-\infty}^{\tau} y \int_{-\infty}^y \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \frac{\bar{G}(c|W)}{\bar{G}_1^2(c|W)} \frac{dG_1 dG}{dG}(c|W) dF_1(y|W) \\
 &= \int_{-\infty}^{\tau} y \frac{\bar{G}(y|W)}{\bar{G}_1(y|W)} dF(y|W) \\
 &\quad + \int_{-\infty}^{\tau} y \left\{ \int_{-\infty}^y \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \left[\frac{1}{\bar{G}_1(c|W)} \right. \right. \\
 &\quad \left. \left. - \frac{\bar{G}(c|W)}{\bar{G}_1^2(c|W)} \frac{dG_1}{dG}(c|W) \right] dG(c|W) \right\} dF_1(y|W) \\
 &= \int_{-\infty}^{\tau} y \frac{\bar{G}(y|W)}{\bar{G}_1(y|W)} dF(y|W) \\
 &\quad - \int_{-\infty}^{\tau} y \left\{ \int_{-\infty}^y \frac{\bar{F}(c|W)}{\bar{F}_1(c|W)} \left[\frac{d}{dc} \frac{\bar{G}(c|W)}{\bar{G}_1(c|W)} \right] dc \right\} dF_1(y|W).
 \end{aligned}$$

If $G = G_1$, then $\frac{d}{dc} \frac{\bar{G}(c|W)}{G_1(c|W)} = 0$, so the second term in (19) vanishes, and we are left with,

$$E[Y^*(O)|W] = \int_{-\infty}^{\tau} y \frac{\bar{G}(y|W)}{G(y|W)} dF(y|W) = \int_{-\infty}^{\tau} y dF(y|W) = m(W).$$

If $F = F_1$, then (19) becomes,

$$\begin{aligned} E[Y^*(O)|W] &= \int_{-\infty}^{\tau} y \frac{\bar{G}(y|W)}{G_1(y|W)} dF(y|W) \\ &- \int_{-\infty}^{\infty} y \left\{ \int_{-\infty}^y \frac{d}{dc} \frac{\bar{G}(c|W)}{G_1(c|W)} dc \right\} dF(y|W) \\ &= \int_{-\infty}^{\tau} y \left\{ \frac{\bar{G}(y|W)}{G_1(y|W)} - \int_{-\infty}^y \frac{d}{dc} \frac{\bar{G}(c|W)}{G_1(c|W)} dc \right\} dF(y|W) \\ &= \int_{-\infty}^{\tau} y \left\{ \frac{\bar{G}(y|W)}{G_1(y|W)} - \left[\frac{\bar{G}(y|W)}{G_1(y|W)} - \frac{\bar{G}(-\infty|W)}{G_1(-\infty|W)} \right] \right\} dF(y|W) \\ &= \int_{-\infty}^{\tau} y \left\{ \frac{\bar{G}(y|W)}{G_1(y|W)} - \frac{\bar{G}(y|W)}{G_1(y|W)} + \frac{1}{1} \right\} dF(y|W) \\ &= \int_{-\infty}^{\tau} y dF(y|W) \\ &= m(W). \end{aligned}$$

This proves the desired result. \square

References

- Buckley, J. and James, I.R. (1979). Linear regression with censored data. *Biometrika*, **66**, 429-436.
- Doksum, K.A. and Yandell, B.S. (1982). Properties of regression estimates based on censored survival data. In *A Festschrift for Erich Lehmann* (P.J. Bickel, K.A. Doksum and J.L. Hodges, Jr., eds), 140-156. Wadsworth & Brooks/Cole, California.
- Fan, J. and Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *Journal of the American Statistical Association*, **89**, 560-570.

- Fan, J. and Gijbels, I. (1996). Local Polynomial Modeling and Its Applications. Chapman & Hall.
- Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Annals of Statistics*, **9**, 1276-1288.
- Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika*, **74**, 301-309.
- Miller, R.G. and Halpern, J. (1982). Regression with censored data. *Biometrika*, **69**, 521-539.
- Rubin, D.B. and van der Laan, M.J. (2006). Doubly Robust Censoring Unbiased Transformations. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 208. bepress.com/ucbbiostat/paper208.
- van der Laan, M.J. and Dudoit, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: finite sample oracle inequalities and examples. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 130. bepress.com/ucbbiostat/paper130
- van der Laan, M.J. and Robins, J.M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer-Verlag, New York.
- Zheng, Z. (1987). A class of estimators of the parameters in linear regression with censored data. *Acta Mathem. Applic. Sinica*, **3**, 231-241.