Transformation model estimation of survival under dependent truncation and independent censoring

Journal Title
XX(X):1–2

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Sy Han Chiou[1], Matthew D. Austin[1], Jing Qian[2], Rebecca A. Betensky[1]

Abstract
The Supplementary Material contains additional simulation results in support of the main article.

Web Appendix A: Goodness-of-fit diagnostics

As stated in the main manuscript, the two ways the transformation model can fail even if there is convergence to a unique solution, a, are (1) the conditional Kendall's tau that is used to estimate the dependence model may not achieve true quasi-independence between $T'(a)$ and X and (2) the linear dependence model may be incorrect. To address the first potential source of failure and to detect potential residual dependence between $T'(a)$ and X, we recommend permutation based hypothesis tests that are powerful for nonmonotone dependence [1] and the incorporation of $T'(\hat{a})$ (or categorized versions of it) as covariates in a weighted Cox model that accounts for biased selection of uncensored observations. This procedure can be easily implemented using standard software such as the `coxph` function in the `R survival` package. In the absence of censoring, the conditional Pearson's product-moment correlation coefficient [2] can be used, as well.

To address the second source of failure for the transformation model, the linear dependence structure can be verified by examining the linear relationship between X and T under the transformation model assumption. If the linear transformation model holds, the regression coefficient for X in a model for $X - T$ should be close to $(1 + a)$, as required by

$$\begin{aligned} X - T =& -(1+a)E\{T'(a)\}+ \\ & (1+a)X - (1+a)[T'(a) - E\{T'(a)\}] \\ :=& \beta_0 + \beta_1 X + \epsilon. \end{aligned}$$

A more formal diagnostic based on the truncated regression model above is to consider a linear piecewise truncated regression model and test if the slopes in every segment are equal. Specifically, this can be constructed by allowing K changepoints at $b_k, k = 1, \ldots, K$, such that $0 = b_0 < b_1 < \ldots < b_K < b_{K+1} = \max\{X_i\}$. The piecewise truncated regression model on (X, T) is then expressed as $X - T = \beta_0 + \alpha_1 X_1 + \ldots + \alpha_{K+1} X_{K+1} + \epsilon$, where $X_j = \min\{\max(X - b_{k-1}, 0), b_k - b_{k-1}\}, j = 1, \ldots, K + 1$. We recommend selecting changepoints that are uniformly spaced across the range of X, after removing outliers using the Tukey rules on quantiles [3], and subject to the requirement that each segment contains at least 10 events.

Rejection of the global test of $H_0 : \alpha_1 = \ldots = \alpha_{K+1}$ is tantamount to rejection of the transformation model. When the global test is rejected, we apply separate transformation models to each segment to obtain a total of K transformation parameters, each of which is used to compute the latent independent truncation times. In the process, we continue to use the inverse weighting by the censoring distribution to adjust the selection bias due to the restriction of uncensored events. The revised transformation estimator can be obtained by a weighted Kaplan–Meier estimator using the aggregate data that combine $T'(a)$ and X from all segments.

Once non-linearity is detected, it is possible to consider alternative methods using fractional polynomials or monotone splines in place of applying the transformation model separately to each segment. These approaches, like the transformation model, also require estimation of unknown parameters. The estimation of transformation models for each segment has intuitive appeal, is computationally simple and does not require additional assumptions on the underlying structure that are hard to verity. A more in-depth comparison merits future research.

Web Appendix B: Simulation when copula assumptions hold

We compare the performance of the transformation estimator to that of the copula estimators when data are generated under the copula assumption. Specifically, we generated T and X from the conditional joint distribution of (T, Y):

$$\Pr(T \le t, X \ge x | T < X) \propto \phi_\alpha^{-1}[\phi_\alpha\{F_T(t)\} + \phi_\alpha\{S_X(x)\}],$$

where the generator function, $\phi_\alpha(v)$ was either the Frank copula, $\phi_\alpha(v) = \log\{(1 - \alpha^{-1}/1 - \alpha^{-v}\}, \alpha > 0$ or the Clayton copula, $\phi_\alpha(v) = (v^{-(\alpha-1)} - 1)/(\alpha - 1), \alpha \ge 0$. Under these models, the dependence on (T, X) is completely specified by the copula parameter, α. We selected $\alpha = 0.054$

[1] Harvard T.H. Chan School of Public Health, U.S.A.
[2] University of Massachusetts Amherst, U.S.A.

Corresponding author:
Sy Han Chiou, Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, U.S.A
Email: syhan.chiou@gmail.com

and 1.857, corresponding to $\tau = 0.3$ and -0.3 under the Frank and the Clayton copulas, respectively. We generated the failure time, X, from a Weibull distribution with shape parameter 2 and scale parameter 4. We denote this with Weibull$(2, 4)$. We generated the truncation time, T, from a Weibull$(2, p)$ distribution, with scale parameter p tuned to achieve an average truncation probability of $\Pr(T \leq Y) = 50\%$. We generated the censoring variable, C, from a log-normal$(\mu, 1)$ distribution, where μ was selected to obtain three levels of censoring: 20%, 40%, and 60%. For each of 1000 repetitions of our simulation, we retained $n = 200$ observations that satisfied $T < \min(X, C)$.

For each scenario considered, data generated under the Frank and Clayton copula models yield similar average values of $\Pr(X > X_{(n)})$, $\Pr(X > T_{(1)})$ and $\Pr(X > T'_{(1)}(\hat{a}))$. In particular, the averages of $\Pr(X > T_{(1)})$ and $\Pr(X > T'_{(1)}(\hat{a}))$ range from 0.996 to 1.000 under all configurations, so $S'_X(x)$ is expected to be very close to the unconditional survival curve, $S_X(x)$. On the other hand, the average values of $\Pr(X > X_{(n)})$ range from 0.001 to 0.021 under 20% and 40% censoring, but increase to 0.055 under 60% censoring. Therefore, we expect $S''_X(x)$ to be close to $S_X(x)$ and $S'_X(x)$ under 20% and 40% censoring, but not under heavier censoring. Figure 1 displays our proposed estimate, the Frank copula estimate, the Clayton copula estimate and the conventional Kaplan–Meier estimate under independent truncation, under the different scenarios that we consider. Under 20% and 40% censoring, the copula estimators perform poorly when dependence structure is misspecified, while the transformation estimator displays less departure from the target survival curve. Under 60% censoring, the estimators are in closer agreement with the corresponding target survival curve. This is because the dependent truncation is weakened by independent censoring. Overall, these observations indicate that the proposed adjusted transformation estimator exhibits reasonable performance even under a misspecified model. The numerical results summarized in Table 1 confirm these findings.

[Figure 1 about here.]

[Table 1 about here.]

REFERENCES

1. Chiou SH, Qian J, Mormino E et al. Permutation tests for general dependent truncation. *Computational Statistics & Data Analysis* 2018; 128: 308–324.
2. Chen CH, Tsai WY and Chao WH. The product-moment correlation coefficient and linear regression for truncated data. *Journal of the American Statistical Association* 1996; 91(435): 1181–1186.
3. Tukey JW. *Exploratory data analysis*. Reading, Mass., 1977.

Table 1. Summary statistics of the simulation data with $n = 200$. The dependence structure of (T, X) is specified by copula models. Bias is the average bias; ESE is the empirical standard error; MSE is the average mean squared error.

		Transformation				Frank's estimator			Clayton's estimator		
Censoring	$\hat{S}^*(\cdot)$	bias	ESE	ASE	MSE	bias	ESE	MSE	bias	ESE	MSE
						Under Frank's copula dependence structure					
20%	0.8	−0.056	0.090	0.088	0.011	−0.013	0.085	0.007	−0.089	0.079	0.014
	0.6	−0.084	0.086	0.087	0.015	−0.015	0.086	0.008	−0.119	0.070	0.019
	0.4	−0.074	0.068	0.067	0.010	−0.014	0.072	0.005	−0.104	0.053	0.014
	0.2	−0.038	0.042	0.042	0.003	−0.008	0.045	0.002	−0.057	0.031	0.004
40%	0.8	−0.036	0.078	0.075	0.007	−0.008	0.060	0.004	−0.060	0.064	0.008
	0.6	−0.047	0.076	0.078	0.008	−0.008	0.063	0.004	−0.077	0.061	0.010
	0.4	−0.034	0.062	0.064	0.005	−0.007	0.055	0.003	−0.062	0.049	0.006
	0.2	−0.012	0.044	0.045	0.002	−0.005	0.042	0.002	−0.033	0.036	0.002
60%	0.8	−0.008	0.050	0.051	0.003	−0.006	0.043	0.002	−0.032	0.045	0.003
	0.6	−0.003	0.057	0.057	0.003	−0.004	0.050	0.003	−0.032	0.050	0.004
	0.4	0.008	0.058	0.055	0.003	−0.002	0.056	0.003	−0.021	0.054	0.003
	0.2	0.019	0.056	0.055	0.004	−0.003	0.056	0.003	−0.012	0.053	0.003
						Under Clayton's copula dependence structure					
20%	0.8	−0.181	0.309	0.037	0.128	0.152	0.057	0.026	0.008	0.170	0.029
	0.6	−0.136	0.240	0.238	0.076	0.278	0.083	0.084	0.006	0.155	0.024
	0.4	−0.078	0.168	0.170	0.034	0.369	0.108	0.148	0.006	0.116	0.013
	0.2	−0.032	0.090	0.090	0.009	0.340	0.118	0.130	0.005	0.065	0.004
40%	0.8	−0.157	0.295	0.294	0.112	0.109	0.094	0.021	−0.004	0.148	0.022
	0.6	−0.107	0.231	0.233	0.065	0.205	0.145	0.063	−0.003	0.132	0.017
	0.4	−0.063	0.162	0.162	0.031	0.270	0.174	0.104	−0.001	0.097	0.009
	0.2	−0.026	0.087	0.081	0.007	0.235	0.161	0.081	−0.001	0.056	0.003
60%	0.8	−0.096	0.243	0.244	0.068	0.014	0.130	0.017	−0.012	0.110	0.012
	0.6	−0.065	0.190	0.190	0.040	0.064	0.171	0.033	−0.010	0.094	0.009
	0.4	−0.037	0.135	0.136	0.020	0.096	0.194	0.047	−0.009	0.075	0.006
	0.2	−0.007	0.082	0.084	0.007	0.084	0.164	0.034	−0.006	0.058	0.003

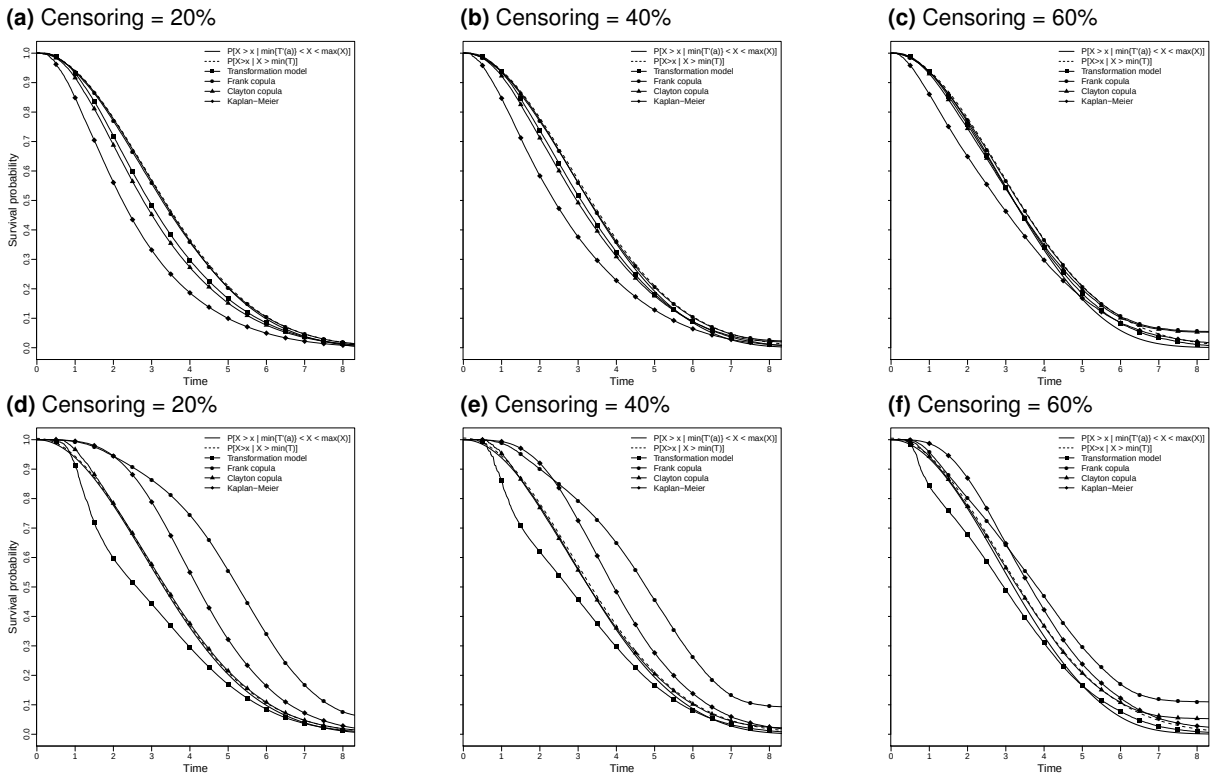

Figure 1. The average of the estimators of the distribution of X when the dependence structure of (T, X) is specified by Frank's copula in (a), (b), (c) and Clayton copula in (d), (e), (f). The results are based on 1000 data sets and a sample size of $n = 200$.