Residual Diagnostics for Censored Regression via Randomized Survival Probabilities

Longhai Li

Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon, SK, CANADA

Statistics Conference in Genomics, Pharmaceutical Science, and Health Data Science
16 August 2022
University of Victoria, Canada

Acknowledgements

- Thanks to my co-authors: Tingxuan Wu and Cindy Feng of this paper:

  Li, L., Wu, T., Feng, C., 2021. Model Diagnostics for Censored Regression via Randomized Survival Probabilities. <u>Statistics in Medicine</u>, 40:2, 1482-1497.

-

- Thanks to Prof. Xuekui Zhang for inviting me to this session.

Outline

Section 1

Introduction

Introduction

- Residual diagnostics is a standard tool for checking the overall goodness-of-fit (GOF) of normal regression models, graphically and numerically. It can also help identify the nature of the model mis-specification, eg, non-linear covariate effect.

- However, similar residual diagnostic tools are not completely available for censored regression.

- It is desired to have a tool like residual diagnostics for censored regression. In this paper, we propose to diagnose censored regression with a new residual called normalized randomized survival probabilities (NRSP). In particular, we show its effectiveness in detecting non-linear covariate effect.

Section 2

Review of Existing Model Checking Methods for Censored Regression

When there is no censoring

- Suppose the survival function of T_i^* based on a model is defined as $S_i(t_i^*) = P(T_i^* > t_i^*)$, where the subscript i indicates that the probability depends covariate x_i for the ith individual.
- The survival probabilities $S_i(T_i^*)$ are uniformly distributed when $S_i(\cdot)$ is the true survival function of T_i^*. $S_i(T_i^*)$ can be transformed to follow any desired distribution:
  - Cox-Snell (CS) residual is defined as

  $$r_i^c(T_i^*) = -\log(S_i(T_i^*)),$$

  where $-\log(\cdot)$ is the inverse survival function of $\exp(1)$. Therefore, CS residuals are exponentially distributed under the true model.
  - One can also define normally-distributed residuals:

  $$r_i^n(T_i^*) = \Phi^{-1}(S_i(T_i^*)),$$

  which we call by **normalized SPs**. Then we can apply a variety of residual diagnostic methods for normal regression to diagnose $S_i(\cdot)$.

When censoring occurs

- When T_i^* is right-censored, we only observe (T_i, d_i), where $T_i = \min(T_i^*, C_i), d_i = I(T_i^* < C_i)$.
- Since $T_i < T_i^*$, $S_i(T_i) > S_i(T_i^*)$. We call $S(T_i)$ **unmodified survival probabilities**.
- $S_i(T_i)$ is not uniformly distributed:
  - $S_i(T_i)$ is a left-censored observation of the uniform distribution.
  - The **unmodified CS residuals**, $r_i^c(T_i) = -\log(S_i(T_i))$ is a right-censored observation from $\exp(1)$ under the true model.
  - The **normalized unmodified SPs (NUSP)**, $r_i^n(T_i)$ is a left-censored observation from N(0,1) under the true model.

Overall GOF Checking for Censored Data

- The most widely used diagnostic tool is to apply KM methods to $\{(r_i^c(T_i), d_i)|i = 1, \ldots, n\}$ to get an estimate of the CHF of CS residuals. Under the true model, the CHF of CS residuals is expected to be close to the $45°$ straight line.
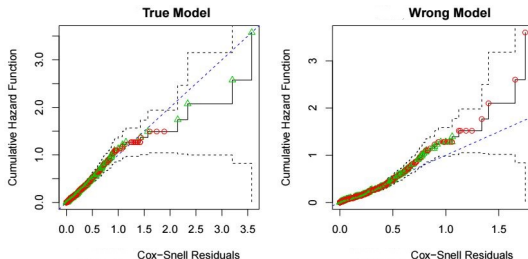
- Shapiro-Francia normality test has also been extended to censored data (available in R package EnvStats). We can apply this method to NUSP ($\Phi^{-1}(S_i(T_i))$), which we call **NUSP-CSF** method.

Overall GOF Checking for Censored Data (contd.)

- Checking the distribution of residuals is only the first-line model diagnostics. The GOF test results typically cannot reveal the nature of model mis-specification, for example non-linear effects, lack of independence, and heterogeneous variance.

- Therefore, it may not help improve our models.

Other Residuals for Censored Observations

- A commonly used method is to shrink the USPs of the censored failure times:

$$S_i'(T_i, d_i, \eta) = \eta\, S_i(T_i), \quad \text{if } T_i \text{ is censored, i.e., } d_i = 0, \qquad (1)$$

where $\eta \in (0, 1)$. Different choices of η exist based on different arguments, eg, $1/2, 1/e$, corresponding to adding $\log(2)$ or 1 to CS residuals. There are also other residuals like martigale and deviance residuals. The residuals for censored T_i are modified similarly with a **non-random** scheme.

- **Problem:** All these residuals lack a known reference distribution under censoring; this creates the obstacle in applying a variety of residual diagnostics tools, especially in quantifying the degree of model violation.

Section 3

Normalized Randomized Survival Probabilities
(Z-residuals)

Definition of Randomized Survival Probabilities

- The randomized survival probability (**RSP**) for T_i is defined as:

$$S_i^R(T_i, d_i, U_i) = \left\{ \begin{array}{ll} S_i(T_i), & \text{if } T_i \text{ is uncensored, i.e., } d_i = 1, \\ U_i\, S_i(T_i), & \text{if } T_i \text{ is censored, i.e., } d_i = 0, \end{array} \right.$$

$$(2)$$

  where U_i is a uniform random number on $(0, 1)$.

- We can prove that the randomized SP is uniformly distributed on $(0, 1)$ given x_i under the true model. Therefore, we can transform them into residuals with any desired distribution. We prefer to transforming them with the normal quantile:
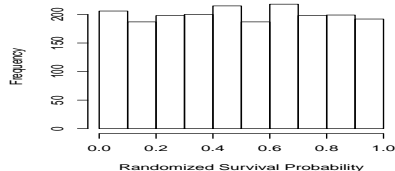
$$r_i^{\text{NRSP}}(T_i, d_i, U_i) = \Phi^{-1}(S_i^R(T_i, d_i, U_i)). \qquad (3)$$

- We call the residuals in (3) by normalized randomized SP (**NRSP**) residuals. Now we feel that Z-residual is a better name.
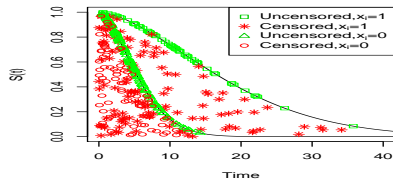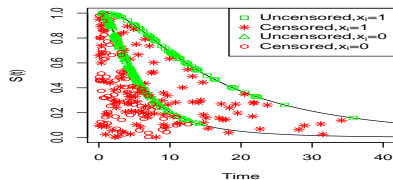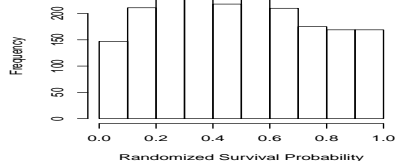
(a) RSPs, True Model

(b) Hist of RSPs, True Model

(c) RSPs, Wrong Model

(d) Hist of RSPs, Wrong Model

Figure 1: Illustration of the uniformity of RSPs.

Section 4

A Simulation Study

Data Generation

- The response variable is simulated from a Weibull AFT regression model with a non-linear link function: $\log(T_i^*) = 2 + 5\sin(2x_i) + \epsilon_i$. The covariate x_i was generated uniformly on $(0, 3\pi/2)$. The shape parameter of the Weibull distribution was set as 1.8. The censoring times C_i were generated from $\exp(\theta)$ with θ varied for obtaining different censoring rates.

- We considered fitting a Weibull AFT model assuming $\log(T_i^*) = \beta_0 + \beta_1 x_i + \epsilon_i$ as a wrong model, and fitting a Weibull AFT model assuming $\log(T_i^*) = \beta_0 + \beta_1\sin(2x_i) + \epsilon_i$ as the true model.
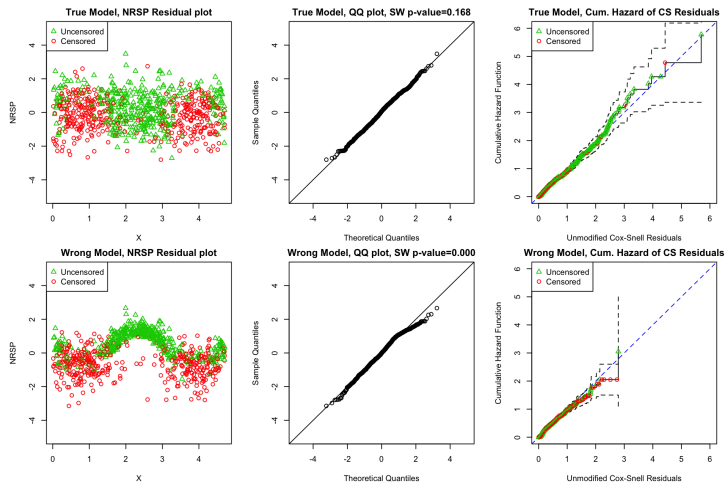
Graphical Model Checking

Figure 2: Performance of the NRSP residuals as a graphical tool for detecting non-linear effect in covariate. The dataset has a sample size n = 800 and a censoring rate $c \approx 50\%$.

Table 1: Comparison of the percentages of model rejections of various statistical tests. A model is rejected when the test p-value is smaller than 0.05.

		Under the true model							Under the wrong model						
n	$100c$	NRSP-SW	NRSP-SF	NUSP-CSF	NRSP-AOV	NRSP-KS	NMSP-SW	Dev-SW	NRSP-SW	NRSP-SF	NUSP-CSF	NRSP-AOV	NRSP-KS	NMSP-SW	Dev-SW
100	0	4.90	4.65	4.65	3.00	0.00	4.90	4.80	62.30	43.85	43.85	100.00	0.40	62.30	63.85
200	0	3.75	4.60	4.60	3.75	0.00	3.75	3.75	95.00	89.85	89.85	100.00	5.90	95.00	95.75
400	0	4.50	4.15	4.15	3.60	0.00	4.50	4.20	99.95	99.95	99.95	100.00	45.80	99.95	99.95
800	0	4.45	4.50	4.50	3.05	0.05	4.45	5.10	100.00	100.00	100.00	100.00	98.40	100.00	100.00
100	20	4.90	5.10	4.80	3.50	0.10	4.90	4.80	50.55	34.60	50.70	100.00	0.20	78.15	80.00
200	20	5.45	5.15	5.35	4.35	0.00	55.20	20.25	88.05	80.05	91.55	100.00	2.00	98.85	98.85
400	20	5.00	5.25	4.45	3.30	0.05	88.00	37.90	99.75	99.60	99.90	100.00	25.20	100.00	100.00
800	20	5.35	5.60	5.00	2.60	0.20	99.60	68.80	100.00	100.00	100.00	100.00	89.25	100.00	100.00
100	50	4.45	4.95	3.25	3.75	0.35	99.95	94.25	40.35	26.55	56.95	100.00	0.30	99.80	99.80
200	50	5.70	6.30	2.65	3.40	0.55	100.00	99.75	82.35	72.05	92.80	100.00	0.60	100.00	100.00
400	50	5.45	5.15	1.60	3.20	0.85	100.00	100.00	99.50	99.05	99.85	100.00	7.95	100.00	100.00
800	50	4.55	4.10	1.35	3.60	0.65	100.00	100.00	100.00	100.00	100.00	100.00	55.75	100.00	100.00
100	80	4.46	4.67	1.73	2.64	1.98	100.00	100.00	8.52	5.33	21.36	92.03	1.12	100.00	100.00
200	80	4.28	4.58	1.83	3.16	2.04	100.00	100.00	24.49	14.77	47.00	99.90	2.49	100.00	100.00
400	80	5.07	5.23	0.72	4.20	3.02	100.00	100.00	59.92	46.44	83.24	100.00	2.20	100.00	100.00
800	80	4.90	5.01	0.46	3.41	2.17	100.00	100.00	93.86	89.73	98.61	100.00	4.44	100.00	100.00

Notes:

- NRSP-AOV: a non-linearity test applied to NRSP residuals.
- NUSP-CSF: extended SF normality test for censored data (NUSP).

Section 5

A Real Data Example

A Breast Cancer Recurrent-free Dataset

- Source:
  Schumacher M, Bastert G, Bojar H, et al. Randomized 2 x 2 Trial
  Evaluating Hormonal Treatment and the Duration of Chemotherapy
  in Node-Positive Breast Cancer Patients. German Breast Cancer
  Study Group.. Journal of Clinical Oncology 1994; 12(10): 2086-2093.
  doi: 10.1200/JCO.1994.12.10.2086
- The sample size is 686. The censoring rate is 56.5%.
- The response variable of interest is the recurrence-free time, which is
  the time from entry to the study until a recurrence of cancer or death.
- We consider the following covariates: the tamoxifen treatment
  indicator, patient age, menopausal status, size and grade of the
  tumour, number of positive lymph nodes, progesterone and estrogen
  receptor status.
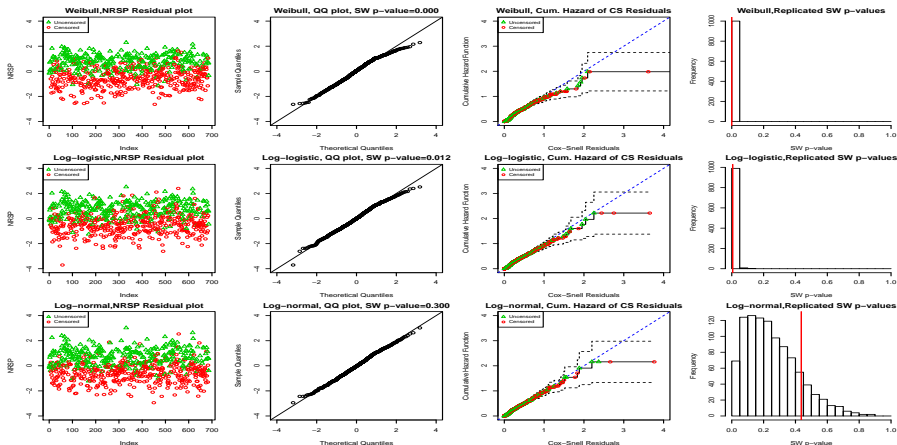
Graphical Checking and NRSP-based GOF Tests

Figure 3: NRSP residuals of the Weibull, log-logistic, and log-normal AFT models fitted to the breast cancer patients dataset. The last column presents the histograms of 1000 replicated NRSP-SW p-values of each model. The vertical red lines indicate p_{min} calculated with the 1000 replicated NRSP p-values.

A Upper-bound for Replicated NRSP-based Test p-values

- Suppose $p_1, \ldots, p_J$ are J replicated NRSP test p-values for a fitted model.
- A formula for bounding tail probabilities of order statistics of correlated samples gives the following inequality for the rth order statistics $p_{(r)}$:

$$P(p_{(r)} < t) \leq \min\left(1, t\frac{J}{r}\right). \tag{4}$$

- Based on (4), a p-value upper bound for observed (simulated) rth statistics $p_{(r)}^{\text{obs}}$ is given by $\min\left(1, p_{(r)}^{\text{obs}}\frac{J}{r}\right)$.
- To avoid the selection of r, we report the minimal upper bound for $r = 1, \ldots, J$, denoted by $p_{\min}$:

$$p_{\min} = \min_{r=1,\ldots,J} \min\left(1, p_{(r)}^{\text{obs}}\frac{J}{r}\right). \tag{5}$$

When a model has a small $p_{\min}$, it is highly suspected that the model can be improved for better fitting the dataset.

Results of Statistical Tests

Table 2: Model diagnostic test p-values or p_{min} for NRSP tests and AIC values of different models for the breast cancer data. The numbers in brackets for NRSP tests are the percentages of replicated NRSP test p-values being ≤ 0.05.

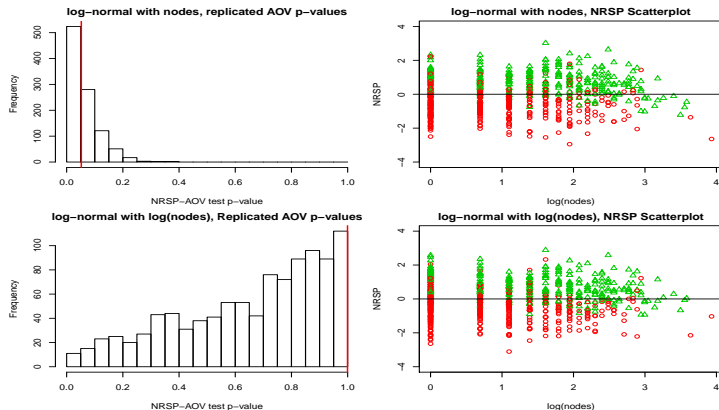| Model	Weibull	log-logistic	log-normal	log-normal with log(nodes)
AIC	5181	5153	5139	5121
NUSP-CSF p-values	1.69e-5	1.94e-3	0.133	0.172
NRSP-SW p_{min}	2.43e-05 (100.00)	6.01e-03 (99.00)	0.436 (6.90)	0.529 (5.70)
NRSP-SF p_{min}	9.50e-05 (100.00)	0.0152 (97.00)	0.485 (5.80)	0.637 (3.50)
NRSP-AOV p_{min}	0.020 (60.40)	0.007 (46.00)	**0.052 (52.40)**	0.999 (0.50)

Graphical Checking of Non-Linearity

Figure 4: NRSP non-linearity residual diagnosis for the breast cancer data. The red vertical lines in the histograms of replicated NRSP-AOV p-values show the $p_{\min}$ values.

Conclusion

- This paper has proposed using randomized survival probabilities (RSPs) to conduct model diagnostics for censored regression. NRSP residuals are approximately distributed with $N(0, 1)$ under the true model. With this unified reference distribution for NRSP residuals, we can conduct a wide variety of residual diagnostics for censored regression.

- Our simulation studies show that a non-linearity test with NRSP residuals has significantly higher power in detecting non-linearity than existing GOF tests. The real data analysis shows that the NRSP residual diagnostics successfully captures a subtle non-linear relationship in the dataset.

- R functions for computing NRSP residuals for `survreg` and `coxph` objects are also available on my website:
  `https://longhaisk.github.io/`

Future work

- We are working on extending the NRSP residual diagnostics to share frailty models (survival models with random effects).

- We expect that many other specific model mis-specification tests that target a particular model discrepancy have higher powers than GOF tests. For example, statistical tests for checking proportional hazard assumption in Cox regression seem to be demanded very often. We are working on developing such specific tests based on NRSP residuals.

- Cross-validatory NRSP residual (Z-residual). We are working on developing software packages for computing cross-validatory Z-residuals for shared frailty models and investigating the effect of using cross-validation or not.