Estimation and Inference for AFT Cure Model

Wenbin Lu

Department of Statistics
North Carolina State University

The 2011 Taipei International Statistical Symposium

Outline

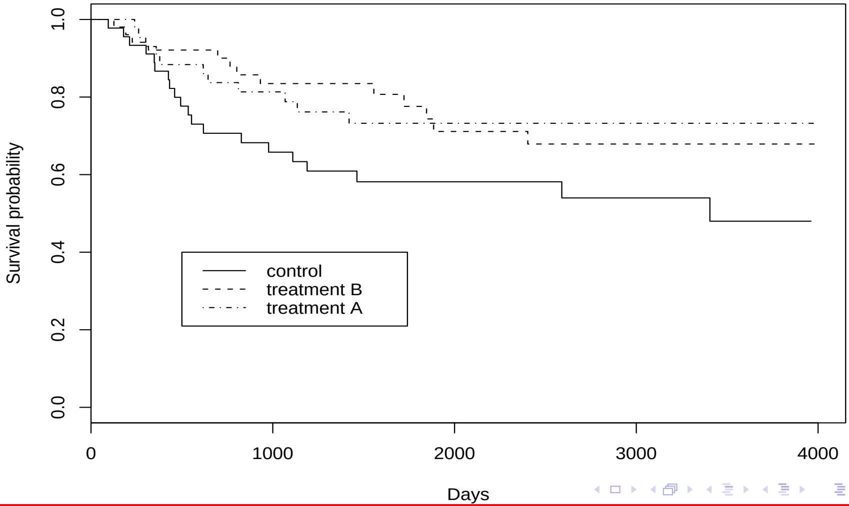
- Introduction and Motivating Example
- Review of Existing Methods for Cure Models
- Our Proposed Model and Method
  - AFT cure model
  - Kernel smoothing aided EM algorithm
  - Asymptotical properties
  - Variance estimation via numerical differentiation
- Numerical Examples
  - Simulation studies
  - Application to a breast cancer data
- Discussions

- Cure models are useful to handle situations in which a proportion of study subjects may never experience the event of interest.
- Some Applications:
  - in medical study: eg. time to relapse or death for breast cancer patients receiving adjuvant therapy (Farewell, 1986)
  - in criminological study: eg. time to return to prison after release for convicts (Partanen, 1969)
  - in consumers behavioral study: eg. time to buy the same type of product after previous purchase (Anscombe, 1961)
- See "Survival Analysis with Long-term Survivors" (Maller & Zhou, 1996) for a good reference.

Breast cancer data (Farewell, 1986)

- a clinical trial of adjuvant therapy for breast cancer involving 139 patients
- endpoint of interest: relapse free survival time (time to relapse or death) (68.3% censoring)
- maximum follow-up: more than 10 years
- risk factors: treatments, clinical stage I, pathological stage I, histological grade I disease and number of lymph nodes

Kaplan-Meier curves for three treatment groups



- Mixture model representation:

$$T = \eta T^* + (1 - \eta)\infty,$$

  $T^* < \infty$: the failure time of a susceptible subject; η: the cure status indicator.
    - parametric mixture cure models (Berkson & Gage, 1952; Farewell, 1982, 1986)
    - PH cure model (Kuk & Chen, 1992; Peng & Dear, 2000; Sy & Taylor, 2000)
    - linear transformation cure model (Lu & Ying, 2004)
- Bounded cumulative hazard representation
    - PH model extension (Tsodikov 1998, 2001; Tsodikov et al. 2003)
    - transformation model extension (Zeng et al., 2006)

- Accelerated failure time model for T^*:

$$\log T^* = \beta' Z + \epsilon,$$

where β is a p-dimensional vector of parameters and ϵ is the error term with a completely unspecified continuous density function.

- Logistic regression model for η:

$$P(\eta = 1|X) = \exp(\gamma' X)/\{1 + \exp(\gamma' X)\},$$

where γ is a q-dimensional vector of parameters and X includes intercept, e.g. $X = (1, Z')'$.

Some available methods:

- Modified M-estimation (Li & Taylor, 2002)
- Modified Gehan-rank estimation (Zhang & Peng, 2007)

Limitations:

- the theoretical properties of these estimates have not been studied
- they all use the bootstrap method to obtain the variance estimates since the likelihood based methods can't be applied
- they may not be semiparametric efficient

- Define $\tilde{T}_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \le C_i)$, where C_i is the censoring time. Assume T_i and C_i are conditionally independent given Z_i and X_i.
- Note that $\delta_i = 1$ implies $\eta_i = 1$; while when $\delta_i = 0$, η_i is unknown.
- Some notations: $\theta = (\beta', \gamma')'$, $\pi(a) = \exp(a)/\{1 + \exp(a)\}$, $R_i(\beta) = \log(\tilde{T}_i) - \beta' Z_i$, and f and S are, respectively, the density and survival functions of e^{ϵ}.

- The likelihood function for observed data

$$L_n^o(\theta, f) = \prod_{i=1}^{n} \Bigg(\left[\pi(\gamma' X_i) e^{-\beta' Z_i} f\{e^{R_i(\beta)}\} \right]^{\delta_i} \times \left[1 - \pi(\gamma' X_i) + \pi(\gamma' X_i) S\{e^{R_i(\beta)}\} \right]^{1-\delta_i} \Bigg).$$

- The complete data likelihood:

$$L_n^c(\theta, f) = \prod_{i=1}^{n} \Bigg\{ \left[\pi(\gamma' X_i) e^{-\beta' Z_i} f\{e^{R_i(\beta)}\} \right]^{\delta_i \eta_i} \times \left(\left\{ 1 - \pi(\gamma' X_i) \right\}^{1-\eta_i} \left[\pi(\gamma' X_i) S\{e^{R_i(\beta)}\} \right]^{\eta_i} \right)^{1-\delta_i} \Bigg\}.$$

Let $l_n^o(\theta, f) = (1/n)\log\{L_n^o(\theta, f)\} \equiv l_{n,1}^c(\gamma) + l_{n,2}^c(\beta, \lambda)$, where

$$l_{n,1}^c(\gamma) = \frac{1}{n}\sum_{i=1}^{n}[\eta_i \log\{\pi(\gamma' X_i)\} + (1-\eta_i)\log\{1-\pi(\gamma' X_i)\}],$$

$$l_{n,2}^c(\beta, \lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(-\delta_i \beta' Z_i + \delta_i \log[\lambda\{e^{R_i(\beta)}\}] - \eta_i \Lambda\{e^{R_i(\beta)}\}\right),$$

with λ and Λ being the hazard and cumulative hazard functions of e^{ϵ}, respectively.

Let $\mathcal{O}_i$ denote the observed data of the ith study subject and $\hat{\Omega}^{[k]} = (\hat{\beta}^{[k]}, \hat{\theta}^{[k]}, \hat{\lambda}^{[k]})$ be the parameter estimates at the kth iteration. Define

$$
\begin{aligned}
w_i^{[k]} &= P\left(\eta_i = 1 | \mathcal{O}_i, \hat{\Omega}^{[k]}\right) \\
&= \delta_i + (1-\delta_i) \frac{\pi(\hat{\gamma}^{[k]\prime} X_i) \hat{S}^{[k]}\{e^{R_i(\hat{\beta}^{[k]})}\}}{1 - \pi(\hat{\gamma}^{[k]\prime} X_i) + \pi(\hat{\gamma}^{[k]\prime} X_i) \hat{S}^{[k]}\{e^{R_i(\hat{\beta}^{[k]})}\}},
\end{aligned}
$$

and $\mathcal{O} = \{\mathcal{O}_i : i = 1, \cdots, n\}$.

We have

$$\tilde{l}^c_{n,1}(\gamma) \equiv E\{l^c_{n,1}(\gamma)|\mathcal{O}, \hat{\Omega}^{[k]}\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}[w_i^{[k]} \log\{\pi(\gamma' X_i)\} + (1 - w_i^{[k]}) \log\{1 - \pi(\gamma' X_i)\}],$$

$$\tilde{l}^c_{n,2}(\beta, \lambda) \equiv E\{l^c_{n,2}(\beta, \lambda)|\mathcal{O}, \hat{\Omega}^{[k]}\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(-\delta_i \beta' Z_i + \delta_i \log[\lambda\{e^{R_i(\beta)}\}] - w_i^{[k]}\Lambda\{e^{R_i(\beta)}\}\right).$$

- The maximization of $\tilde{l}^c_{n,1}(\gamma)$ can be easily done using the Newton-Ralphson method.
- The maximization of $\tilde{l}^c_{n,2}(\beta, \lambda)$ over β and λ is challenging. We adapted the kernel smoothing technique proposed by Zeng and Lin (2007) for standard AFT model.
- Specifically, consider a piece-wise constant hazard function

$$\lambda^*(x) = \sum_{j=1}^{J_n} c_j I(x \in [x_{j-1}, x_j)), \quad 0 \leq x < M,$$

where $[0, M]$ contains all $e^{R_i(\beta)}$'s and $0 \equiv x_0 < x_1 < \cdots x_{J_n} \equiv M$ is a partition.

- Replace $\lambda(x)$ by $\lambda^*(x)$ in $\tilde{l}^c_{n,2}(\beta,\lambda)$. For fixed β, maximizing $\tilde{l}^c_{n,2}(\beta,\lambda)$ with respect to c_j's, we have

$$\hat{c}_j = \frac{\sum_{i=1}^n \delta_i I(x_{j-1} \le e^{R_i(\beta)} < x_j)}{\sum_{i=1}^n w_i^{[k]}[\{e^{R_i(\beta)} - x_{j-1}\} I(x_{j-1} \le e^{R_i(\beta)} < x_j) + I(e^{R_i(\beta)} \ge x_j) M/J_n]}.$$

- It leads to the conditional profile log-likelihood function

$$\begin{aligned}
\tilde{l}^p_{n,2}(\beta) &= -\frac{1}{n}\sum_{i=1}^n \delta_i \beta' Z_i \\
&+ \sum_{j=1}^{J_n}\left\{\frac{1}{n}\sum_{i=1}^n \delta_i I(e^{R_i(\beta)} \in [x_{j-1}, x_j))\right\} \times \log\left\{\frac{J_n}{nM}\sum_{i=1}^n \delta_i I(e^{R_i(\beta)} \in [x_{j-1}, x_j))\right\} \\
&- \sum_{j=1}^{J_n}\left\{\frac{1}{n}\sum_{i=1}^n \delta_i I(e^{R_i(\beta)} \in [x_{j-1}, x_j))\right\} \times \\
&\log\left\{\frac{J_n}{nM}\sum_{i=1}^n w_i^{[k]}\{e^{R_i(\beta)} - x_{j-1}\} I(e^{R_i(\beta)} \in [x_{j-1}, x_j)) + \frac{1}{n}\sum_{i=1}^n w_i^{[k]} I(e^{R_i(\beta)} \ge x_j)\right\}.
\end{aligned}$$

- It can shown that as $n \to \infty$, $J_n \to \infty$ and $J_n/n \to 0$, $\tilde{l}^p_{n,2}(\beta)$ is asymptotically equivalent to the following kernel-smoothed conditional profile log-likelihood function

$$\tilde{l}^s_{n,2}(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\delta_i \log(\tilde{T}_i) + \frac{1}{n}\sum_{i=1}^{n}\delta_i \log\Big[\frac{1}{n}\sum_{j=1}^{n}\delta_j K_h\{R_j(\beta) - R_i(\beta)\}\Big]$$
$$-\frac{1}{n}\sum_{i=1}^{n}\delta_i \log\Big\{\frac{1}{n}\sum_{j=1}^{n} w_j^{[k]} \int_{-\infty}^{R_j(\beta)-R_i(\beta)} K_h(u)du\Big\},$$

where $K_h(x) = K(x/h)/h$ is a kernel function with bandwidth h obtained from a symmetric probability density function $K(x)$.

- Let $\hat{\gamma}^{[k+1]}$ and $\hat{\beta}^{[k+1]}$ denote the maximizers of $\tilde{l}^{c}_{n,1}(\gamma)$ and $\tilde{l}^{s}_{n,2}(\beta)$, respectively.
- Given $\hat{\beta}^{[k+1]}$ and $x > 0$, we can estimate $\lambda(x)$ by

$$\hat{\lambda}^{[k+1]}(x) = \frac{x^{-1}\sum_{j=1}^{n}\delta_j K_h\{R_j(\hat{\beta}^{[k+1]}) - \log x\}}{\sum_{j=1}^{n} w_j^{[k]}\int_{-\infty}^{R_j(\hat{\beta}^{[k+1]})-\log x} K_h(u)du}.$$

- Repeat E-step and M-step until convergence. Let $\hat{\theta} \equiv (\hat{\beta}', \hat{\gamma}')'$ and $\hat{\lambda}(x)$ denote the estimates at convergence obtained from the EM algorithm.

Let $\theta_0 = (\beta_0', \gamma_0')'$ denote the true value of θ and $\lambda_0(x)$ be the true value of $\lambda(x)$. Under some regularity conditions, we have

- Theorem 1. As $h \to 0$ and $nh^2 \to \infty$,

$$\sup_{x \in [0,\tau]} |\hat{\Lambda}(x) - \Lambda_0(x)| \to 0 \; a.s. \text{ and } ||\hat{\theta} - \theta_0|| \to 0 \; a.s.$$

- Theorem 2. As $nh^4 \to 0$ and $nh^2 \to \infty$, $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a mean-zero normal random vector with covariance matrix achieving the semiparametric efficiency bound of θ_0.

- Louis's formula (1982) for parametric EM algorithms is not really feasible here.
- We used an EM-aided numerical differentiation method for computing the empirical Fisher information matrix of the profile likelihood (Chen and Little, 1999).
- We used the optimal bandwidth proposed by Jones (1990) and Jones and Sheather (1991) for density estimation: $(8\sqrt{2}/3)^{1/5}\sigma_1 n^{-1/5}$ and $4^{1/3}\sigma_2 n^{-1/3}$, where σ_1 and σ_2 are the standard deviation of $R_i(\hat{\beta}^{[0]})$'s based on uncensored data and all the data, respectively.

- Z is generated from Bernoulli(0.5), $X = (1, Z)'$, and $\epsilon = a_0 + a_1 V$, where V is generated from three distributions: the extreme value distribution, the logistic distribution, and the standard normal distribution.
- We choose $\beta_0 = 1.0$ and $\gamma_0 = (0.5, -0.5)$ or $(1.0, -0.5)$, which give approximately 43.9% and 32.3% overall cure fractions, respectively.
- The censoring time C is generated from a uniform distribution on $[0, a_2]$, where a_0, a_1 and a_2 are chosen to obtain the desired censoring proportions.
- We set $n = 100$ and run 500 simulations for each setting.

ZP: Zhang and Peng (2007)'s estimator; RE: relative efficiency of ZP estimator vs. proposed estimator.

	Proposed				ZP		
Parameters	Bias	SD	SE	CP	Bias	SD	RE
	extreme value error						
β_0	0.017	0.220	0.220	0.938	0.015	0.197	1.247
γ_{01}	0.020	0.322	0.335	0.960	0.012	0.305	1.112
γ_{02}	0.026	0.466	0.479	0.942	0.030	0.452	1.064
	logistic error						
β_0	0.018	0.138	0.158	0.964	0.013	0.154	0.801
γ_{01}	0.012	0.314	0.322	0.960	0.016	0.316	0.988
γ_{02}	0.023	0.462	0.465	0.950	0.021	0.470	0.965
	normal error						
β_0	0.031	0.176	0.182	0.948	0.004	0.190	0.859
γ_{01}	0.028	0.321	0.317	0.958	0.009	0.321	0.999
γ_{02}	0.003	0.483	0.457	0.940	0.032	0.471	1.052

	Proposed				ZP		
	extreme value error						
Parameters	Bias	SD	SE	CP	Bias	SD	RE
	extreme value error						
β_0	0.018	0.178	0.202	0.962	0.005	0.179	0.992
γ_{01}	0.035	0.375	0.369	0.944	0.031	0.341	1.208
γ_{02}	0.014	0.542	0.518	0.938	-0.008	0.503	1.163
	logistic error						
β_0	0.015	0.127	0.142	0.970	0.007	0.143	0.790
γ_{01}	0.035	0.344	0.356	0.966	0.037	0.349	0.971
γ_{02}	-0.002	0.521	0.498	0.934	-0.008	0.536	0.946
	normal error						
β_0	0.012	0.145	0.162	0.962	0.017	0.160	0.820
γ_{01}	0.041	0.366	0.355	0.958	0.036	0.370	0.980
γ_{02}	0.040	0.534	0.502	0.940	0.040	0.529	1.018

- We fit the AFT cure model for the breast cancer data (Farewell, 1986).
- Four covariates are included: treatment A, treatment B, clinical stage I and Lymph nodes (>4).
- The results are given below:

	$\hat{\beta}$	SE	p-value	$\hat{\gamma}$	SE	p-value
intercept				0.210	0.491	0.669
treatment A	1.113	0.434	0.010	−0.067	0.133	0.612
treatment B	−0.107	0.421	0.800	−1.170	0.583	0.045
clinical stage I	1.350	0.562	0.016	−0.338	0.116	0.004
Lymph nodes	−0.552	0.485	0.255	1.152	0.763	0.131

Future works:

- develop some diagnostic tools, such as cumulative sums of martingale-based residuals for various types of cure models (e.g., PH cure vs. AFT cure).
- build in variable selection procedures for cure models

Reference:

- Wenbin Lu (2010), Efficient Estimation for Accelerated Failure Time Model with a Cure Fraction. Statistica Sinica, 20, 661-674.