Estimation of the Latent Distribution in Cure Survival Models using a Flexible Cox Model

Vincent Bremhorst  Philippe Lambert

Université catholique de Louvain (UCL)
Institut de Statistique, Biostatistique et sciences Actuarielles (ISBA)

1 February 2013
Outline of the talk

Introduction

The promotion time model

Estimation

Bayesian analysis

Simulation Study
Outline of the talk

Introduction

The promotion time model

Estimation

Bayesian analysis

Simulation Study
Introduction to cure models

- Common hypothesis in classical survival models:
  - Any observed subject will experience the monitored event if the follow up is sufficiently long.

- Realistic assumption?
  - How to deal with subjects who will never experience the event of interest?

⇒ **Cure models** incorporate the unknown proportion of immune subjects in survival models.
Introduction to cure models

- They are two well known families of cure models:
  - The mixture models (Berkson and Gage 1952; Farewell 1982 and 1986; Sy and Taylor 2000, for example);
  - The promotion time models (Tsodikov 1998; Chen Ibrahim and Sinha 1999; Zeng, Yin and Ibrahim (2006); Liu and Shen (2009), for example).

- Some authors have also defined unified approaches (Yin and Ibrahim 2005; Cooner and al. 2007, for example).
Outline of the talk

Introduction

The promotion time model

Estimation

Bayesian analysis

Simulation Study
The promotion time model

The promotion time model has a biological interpretation:

- Each subject is exposed to $N \sim \text{Pois}(\theta)$ carcogenic cells.
- Let $Y$ be the incubation time of one cell ($Y$ is often called latent event time).

Assumptions:

- The cancer mass of each cell is detected independently from each other.
- Only one cell needs to be detected for subject to fail.

$\downarrow$

- $Y_1, \ldots, Y_N$ are independent with a common distribution $F(t)$.
- The observed failure time is defined as $T = \min_i(Y_i)$.
- If $N = 0$, the subject is not exposed and is considered as cured.
The promotion time model

- We enter the covariates through:
  - The mean parameter of the number $N$ of carcogenic cells:
    \[
    \theta(x) = \exp(\beta_0 + x'\beta)
    \]
  - The distribution of the latent event times $F(t|z)$ using a Cox PH model:
    \[
    h(t|z) = h_0(t) \exp(z'\gamma) \quad ; \quad S(t|z) = S_0(t)^{\exp(z'\gamma)}
    \]
- Unconditional (on $N$) population survival function:
  \[
  S^*(t|x, z) = \exp[-\theta(x)F(t|z)]
  = \exp\left[-\exp(\beta_0 + x'\beta) \left(1 - S_0(t)^{\exp(z'\gamma)}\right)\right]
  \] (1)
The promotion time model: Identification issue

- Usual assumptions:
  1. The vector $z$ of covariates does not include an intercept to ensure the identifiability of the Cox PH model.
  2. The baseline distribution function $F_0(t) = 1 - S_0(t)$ is a proper distribution function: $\lim_{t \to \infty} F_0(t) = 1$

- Under i) and ii) we can show that:
  
  A) If the follow up of the study is sufficiently long then model (1) is identifiable.
  
  B) If the follow up of the study is not sufficiently long then model (1) is not identifiable unless vectors $x$ and $z$ do not share the same components.
The promotion time model: Remarks

- $S^*(t|x, z)$ is an improper survival function:

$$
\lim_{t \to \infty} S^*(t|x, z) = \exp \left[ - \exp \left( \beta_0 + x' \beta \right) \right] > 0
$$

- Remember that if the subject is not exposed to any carcogenic cell (if $N = 0$) then (s)he is considered as cured.

$$
P(N = 0) = \exp \left[ - \theta(x) \right] = \exp \left[ - \exp \left( \beta_0 + x' \beta \right) \right] = \lim_{t \to \infty} S^*(t|x, z)
$$
Outline of the talk

Introduction

The promotion time model

Estimation

Bayesian analysis

Simulation Study
We estimate the latent distribution $F(t|z)$ of (1) using a linear combination of cubic B-splines on $\log(h_0(t))$:

$$
\hat{h}_0(t) = \exp \left( \sum_{k=1}^{K} b_k(t) \phi_k \right) = \exp (B_t \phi)
$$

where $(b_k(\cdot), k = 1, \ldots, K)$ denote the cubic B-splines basis associated to a predefined number of equidistant knots defined on $[0, t_{Rcens}]$, where $t_{Rcens}$ is the upper bound of the follow up.
Knowing the relation between the survival and the hazard functions, we obtain the estimation of the baseline survival function:

\[ \hat{S}_0(t) = \exp \left( - \int_0^t \exp(B_u \phi) \, du \right) \]

The integral in this expression has no analytic form and needs to be evaluated numerically. We use the rectangle method.
Estimation : Rectangle method

- Partition of $[0, t_{Rcens}]$ into $J$ small intervals of equal width: $J_j = [\tau_{j-1}, \tau_j]$ where $0 = \tau_0 < \tau_1 < \ldots < \tau_J = t_{Rcens}$.

- Define
  - The middle of the $j^{th}$ interval: $u_j = \frac{\tau_j - \tau_{j-1}}{2}$;
  - The length of the $j^{th}$ interval: $\delta_j = \tau_j - \tau_{j-1}$

The baseline survival function can be approximated by:

$$\hat{S}_0(t) \approx \exp \left( - \sum_{j=1}^{j(t)} \left[ \exp \left( B_{u_j} \phi \right) \right] \delta_j \right)$$

where $j(t)$ corresponds to the interval containing $t$. 
Regularisation

- **Roughness penalty**: We choose a large number (say $K$) of B-splines and countrebalance the flexibility by introducing a penalty on finite differences of adjacent B-spline parameters:

$$\tau \sum_k (\Delta^r \phi_k)^2 = \tau \phi' D' D \phi,$$

where $\tau$ is the penalty parameter. (Eilers and Marx 1996)

- For example, when we use a third order penalty, the matrix $D$ is given by:

$$D = \begin{bmatrix}
1 & -3 & 3 & -1 & 0 & \ldots & 0 \\
0 & 1 & -3 & 3 & -1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & 1 & -3 & 3 & -1
\end{bmatrix}$$
Graphical illustrations

- Number of knots: 12
- $b_k(t) > 0$ for only 3 values of $k$

- blue curves $\equiv B_t \hat{\phi}$
- red curve $\equiv$ Estimation of $\log(h_0(t))$
Likelihood

- Contribution of the $i^{th}$ subject to the likelihood:

$$L(D_i, \Phi) = \hat{S}^*(t_i|x_i, z_i)(\hat{h}^*(t_i|x_i, z_i))^\nu_i$$

where

- $D_i = (t_i, \nu_i, x_i, z_i)$;
- $\nu_i$ is the event indicator;
- $\Phi$ is the set of all parameters specific to the model;
- $\hat{S}^*(t|x, z)$ is the survival function defined in (1) where the baseline survival function $S_0(t)$ was substituted by $\hat{S}_0(t)$;
- $\hat{h}^*(t|x, z)$ is the corresponding hazard function.
Outline of the talk

Introduction

The promotion time model

Estimation

Bayesian analysis

Simulation Study
Bayesian model

- In a Bayesian setting, the roughness penalty is translated into a prior distribution for $\phi$:

$$
\pi(\phi|\tau) \propto \tau^{\frac{\rho(P)}{2}} \exp \left( -\frac{\tau}{2} \phi'P\phi \right)
$$

where $P = D'D$

- We use a common non-informative prior for $\tau$:

$$
\pi(\tau) \propto G(a(= 1), b(= 10^{-4}))
$$

- We use an improper uniform prior for all the regression parameters:

$$
\pi(\beta_i) \propto 1 \quad \forall i \in I_{\beta}
$$

$$
\pi(\gamma_j) \propto 1 \quad \forall i \in I_{\gamma}
$$
Posterior distribution

- Using the Bayes’ theorem, the joint posterior distribution is given by:

\[ \pi(\phi, \tau, \beta_0, \beta, \gamma|D, I) \propto L(D, \Phi)\pi(\phi|\tau)\pi(\tau) \]

- Only the conditional posterior distribution for the penalty parameter \( \tau \) comes from a well known family:

\[ \pi(\tau|\phi, D, I) \propto G \left( a + \frac{\rho(P)}{2}, b + \frac{\phi'P\phi}{2} \right) \]
MCMC algorithm

- Gibbs step for $\tau$;
- Adaptive univariate Metropolis step for $\phi$, $\beta_0$, $\beta$ and $\gamma$ (Haario et al. 2001);
- Experience shows that the mixing of the posterior chain of the spline parameters can be very slow.

→ Applying the adaptive Metropolis step on a reparametrized posterior distribution can improve the mixing of the chain;
→ We estimate the correlation structure of the spline parameters using the link between survival data and the Poisson GLM;
→ We reparametrize the posterior distribution of the spline parameters using the obtained estimation.
Outline of the talk

Introduction

The promotion time model

Estimation

Bayesian analysis

Simulation Study
Simulation Study

Set up

- The baseline hazard function $h_0(t)$ is related to a Weibull distribution with mean 10.8 and standard deviation 5.64.

- 2 covariates are taken into account:
  
  \[ W_1 \sim N(0, 1) \quad \text{and} \quad W_2 \sim B(1, 0.5) \]

- Both covariates enter through the parameter $\theta$ and through the Cox PH model: $\mathbf{x} = (W_1, W_2) = \mathbf{z}$

- Sample size: $n = 500$

- Number of replications: $S = 200$

- Splines are defined on 12 knots and a 3rd order penalty is used.
### Investigated Scenarios

- In each scenario, a sufficient follow up is assumed.

<table>
<thead>
<tr>
<th>Cured (%)</th>
<th>Cured censored (%)</th>
<th>Susceptible censored (%)</th>
<th>Global censoring (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>20%</td>
<td>47%</td>
<td>8%</td>
<td>18%</td>
</tr>
<tr>
<td>40%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>40%</td>
<td>46%</td>
<td>7%</td>
<td>25%</td>
</tr>
</tbody>
</table>
Results of the simulations: Regression parameters

- 20% of cured individuals

<table>
<thead>
<tr>
<th>Global censoring (%)</th>
<th>Estimation</th>
<th>$L_{95%}$</th>
<th>$U_{95%}$</th>
<th>ESE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>$\beta_0 = 0.65$</td>
<td>0.624</td>
<td>0.396</td>
<td>0.888</td>
<td>0.129</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = 1.2$</td>
<td>1.146</td>
<td>0.932</td>
<td>1.394</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = 0.5$</td>
<td>0.521</td>
<td>0.227</td>
<td>0.809</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1 = -1$</td>
<td>-0.910</td>
<td>-1.169</td>
<td>-0.670</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2 = 2.5$</td>
<td>2.386</td>
<td>1.974</td>
<td>2.826</td>
<td>0.212</td>
</tr>
<tr>
<td>18%</td>
<td>$\beta_0 = 0.65$</td>
<td>0.625</td>
<td>0.369</td>
<td>0.933</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = 1.2$</td>
<td>1.170</td>
<td>0.932</td>
<td>1.437</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = 0.5$</td>
<td>0.528</td>
<td>0.199</td>
<td>0.847</td>
<td>0.163</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1 = -1$</td>
<td>-0.941</td>
<td>-1.204</td>
<td>-0.679</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2 = 2.5$</td>
<td>2.411</td>
<td>1.971</td>
<td>2.884</td>
<td>0.237</td>
</tr>
</tbody>
</table>
Results of the simulations: Regression parameters

- 40% of cured individuals

<table>
<thead>
<tr>
<th>Global censoring (%)</th>
<th>Estimation</th>
<th>$L_{95%}$</th>
<th>$U_{95%}$</th>
<th>ESE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>$\beta_0 = 0.4$</td>
<td>0.337</td>
<td>0.112</td>
<td>0.619</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = 1.75$</td>
<td>1.620</td>
<td>1.384</td>
<td>1.889</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = -0.75$</td>
<td>-0.668</td>
<td>-1.015</td>
<td>-0.352</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1 = -1$</td>
<td>-0.816</td>
<td>-1.082</td>
<td>-0.557</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2 = 2.5$</td>
<td>2.350</td>
<td>1.964</td>
<td>2.769</td>
<td>0.202</td>
</tr>
<tr>
<td>25%</td>
<td>$\beta_0 = 0.4$</td>
<td>0.355</td>
<td>0.081</td>
<td>0.687</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>$\beta_1 = 1.75$</td>
<td>1.635</td>
<td>1.365</td>
<td>1.929</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td>$\beta_2 = -0.75$</td>
<td>-0.659</td>
<td>-1.079</td>
<td>-0.316</td>
<td>0.208</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1 = -1$</td>
<td>-0.837</td>
<td>-1.134</td>
<td>-0.552</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2 = 2.5$</td>
<td>2.369</td>
<td>1.942</td>
<td>2.856</td>
<td>0.226</td>
</tr>
</tbody>
</table>
Conclusions of the simulations

- Regression parameters:
  - The point estimates are quite similar;
  - The width of the confidence interval for each parameter increases when we introduce right censoring;
  - For each parameter, the RMSE is quite similar whatever the considered scenario.
Conclusion of the simulations

- Baseline survival function:

**Figure:** 20% of cured subjects. Left: Without right censoring; Right: With right censoring
Conclusion of the simulations

- Baseline survival function:

**Figure:** 40% of cured subjects. Left: Without right censoring; Right: With right censoring.
Further work

- Simulations using a bimodal distribution for the baseline distribution.
- Applying model (1) on a real data set.
- Can we find a covariates structure for $F(t|z)$ such that model (1) is identifiable without a sufficient follow up when $x$ and $z$ share some components?
- Generalization to interval censored data.
- Generalization to hierarchical data.
Some references


