Cox-Over Reported Failure Model

Bumba Mukherjee

May 17, 2017

Abstract

In this paper, we implement Bayesian Inference on the new split population survival model, that explicitly models the misclassification probability of failure (vs. right censored) events. This includes two parametric survival models (Exponential and Weibull) and (possibly) Cox proportional hazards regression model.

Review on Parametric Zombie Survival Model

Likelihood function

Recall from Ben's "Parametric Zombie Survival Model" that the probability of misclassification (that is, subset of non-censured failure outcomes that are being misclassified) is

$$\alpha = \Pr(C_i = 1 | \widetilde{C}_i = 0). \tag{1}$$

The unconditional density is thus given by the combination of an observation's misclassification probability and its probability of experiencing an actual failure conditional on not being misclassified,

$$\alpha_i + (1 - \alpha_i) * f(t_i) \tag{2}$$

And the unconditional survival function is therefore

$$(1 - \alpha_i) * S(t_i), \tag{3}$$

where

$$\alpha_i = \frac{\exp(\mathbf{Z}\gamma)}{1 + \exp(\mathbf{Z}\gamma)}. \tag{4}$$

The likeihood function of the Parametric Zombie Survival Model is defined as

$$L = \prod_{i=1}^{N} [\alpha_i + (1 - \alpha_i) f(t_i | \mathbf{X}, \beta)]^{C_i} [(1 - \alpha_i) S(t_i | \mathbf{X}, \beta)]^{1 - C_i} \tag{5}$$

And the log likelihood is

$$lnL = \sum_{i=1}^{N} \{C_i \ln[\alpha_i + (1 - \alpha_i) f(t_i | \mathbf{X}, \beta)] + (1 - C_i) \ln[(1 - \alpha_i) S(t_i | \mathbf{X}, \beta)]\}. \tag{6}$$

We extend these definitions and notation from the Parametric Zombie Survival Model to the Cox PH framework in order to develop the Cox Overreported Failure (OF) model.

Cox-OF Model

Likelihood function

The survival function of the conventional Cox Proportional Hazard (PH) model is

$$S(t_i|\mathbf{X}_i,\beta) = \exp\left\{-\exp(\mathbf{X}_i,\beta)\int_0^{t_i} h_0(u)du\right\} \tag{7}$$

where $\int_0^{t_i} h_0(u)du$ is the cumulative baseline hazard. The density function of the Cox Proportional Hazard (PH) model is

$$f(t_i|\mathbf{X}_i,\beta) = h(t_i|\mathbf{X}_i)S(t_i|\mathbf{X}_i,\beta) \tag{8}$$

which means that

$$f(t_i|\mathbf{X}_i,\beta) = h_0(t_i)e^{\mathbf{X}_i,\beta}\exp\left\{-\exp(\mathbf{X}_i,\beta)\int_0^{t_i} h_0(u)du\right\} \tag{9}$$

where $h(t_i|\mathbf{X}_i)$ is the conditional hazard function and $h_0(t_i)$ is the unspecified baseline hazard. Hence the likelihood of the Cox-OF model is

$$L = \prod_{i=1}^{N}\left[\alpha_i + (1-\alpha_i)h_0(t_i)e^{\mathbf{X}_i,\beta}\exp\left\{-\exp(\mathbf{X}_i,\beta)\int_0^{t_i} h_0(u)du\right\}\right]^{C_i} \left[(1-\alpha_i)\exp\left\{-\exp(\mathbf{X}_i,\beta)\int_0^{t_i} h_0(u)du\right\}\right]^{1-C_i} \tag{10}$$

The log-likelihood from this expression is

$$\ln L = \sum_{i=1}^{N}\left\{C_i \ln\left[\alpha_i + (1-\alpha_i)h_0(t_i)e^{\mathbf{X}_i,\beta}\exp\left(-\exp(\mathbf{X}_i,\beta)\int_0^{t_i} h_0(u)du\right)\right] + (1-C_i)\ln\left[(1-\alpha_i)\exp\left(-\exp(\mathbf{X}_i,\beta)\int_0^{t_i} h_0(u)du\right)\right]\right\} \tag{11}$$

E-M estimation of Cox-OF model

We need some additional notation to describe the E-M (Expectation-Maximization) algorithm for estimating the Cox-OF model. To this end, let the log of the marginal full likleihood for the Cox-OF model be defined (more briefly) as

$$l_f(\theta,\Lambda_0) = \sum_{i=1}^{N}\{C_i \ln[\alpha_i + (1-\alpha_i)h(t_i|\mathbf{X}_i)S(t_i|\mathbf{X}_i,\beta)] + (1-C_i)\ln[(1-\alpha_i)S(t_i|\mathbf{X}_i,\beta)]\} \quad (12)$$

where $h(t_i|\mathbf{X}_i) = h_0(t_i)e^{\mathbf{X}_i,\beta}$and $S(t_i|\mathbf{X}_i,\beta) = \exp\{-\exp(\mathbf{X}_i,\beta)\Lambda_0(t)\}$ with $\Lambda_0(t) = \int_0^{t_i} h_0(u)du$. Let $\theta = (\gamma,\beta)$. The estimate of (θ,Λ_0) is obtained by maximizing l_f over (θ,Λ_0). This maximization is performed using the E-M algorithm given a suitable starting value $(\theta^{(0)},\Lambda_0^{(0)})$ of (θ,Λ_0). Once we have $\theta^{(0)}$, a suitable $\Lambda_0^{(0)}$ corresponding to $\theta^{(0)}$ can be computed by using the Newton-Raphson method. If we have appropriate starting values for $(\theta^{(0)},\Lambda_0^{(0)})$, then – following Taylor (1995) and Sy and Taylor (2000) who developed the E-M algorithm for the classic Cox cure model – we can formally describe the E-M algorithm for the Cox-OF model as follows. To start, the m-th E-step in the E-M algorithm first transforms l_f into the following form for the complete-data log likelihood

$$l_f^{EM}(\gamma|\mathbf{w}^{(m)}) = \sum_{i=1}^{N}\left[(1-w_i^{(m)}(t_i))\ln\alpha_i + w_i^{(m)}(t_i)\ln(1-\alpha_i)\right] \quad (13)$$

$$l_f^{EM}(\beta,\Lambda_0|\mathbf{w}^{(m)}) = \sum_{i=1}^{N}\left[C_i\ln h_0(t_i) + C_i\mathbf{X}_i\beta + w_i^{(m)}(t_i)\ln f(t_i)\right] \quad (14)$$

for a given $\mathbf{w}^{(m)} = (w_1^{(m)}(t_1), ..., w_N^{(m)}(t_N))(m \geq 1)$, where $\mathbf{w}^{(m)}$ is computed by current parameter values $(\theta^{(m-1)}, \Lambda_0^{(m-1)})$ and

$$w_i(t_i; \theta, \Lambda_0) = \left\{ \begin{array}{ll} \frac{(1-\alpha_i)f(t_i)}{\alpha_i+(1-\alpha_i)f(t_i)} & C_i = 0 \\ 1 & C_i = 1 \end{array} \right\} \tag{15}$$

In the m-th M-step in the E-M algorithm, the Breslow's estimate of Λ_0,

$$\widehat{\Lambda}(t_i|\beta, \mathbf{w}^{(m)}) = \sum_{\{i;T_i \leq t\}} \frac{C_i}{\sum_{j=1}^{N} I(T_j \geq T_i) \exp(\mathbf{X}_j\beta) w_j^{(m)}(T_j)} \tag{16}$$

(where I is an indicator function) can be employed for the maximization of $l_f^{EM}(\beta,\Lambda_0|\mathbf{w}^{(m)})$. Substituting $\widehat{\Lambda}_0(t_i|\beta, \mathbf{w}^{(m)})$ into Λ_0 of $l_f^{EM}(\beta,\Lambda_0|\mathbf{w}^{(m)})$ leads to the following log partial likelihood

$$l_p^{EM}(\beta|\mathbf{w}^{(m)}) = \sum_{i=1}^{N} C_i \left[\mathbf{X}_i\beta - \ln\left(\sum_{j=1}^{N} I(T_j \geq T_i) \exp(\mathbf{X}_j\beta) w_j^{(m)}(T_j) \right) \right] \tag{17}$$

The M-step of $l_f^{EM}(\beta,\Lambda_0|\mathbf{w}^{(m)})$ over (β,Λ_0) is then replaced by maximizing $l_p^{EM}(\beta|\mathbf{w}^{(m)})$ only over β and by treating Λ_0 as a nonparametric function. Hence, the M-step is easily performed by maximizing $l_f^{EM}(\gamma|\mathbf{w}^{(m)})$ and $l_p^{EM}(\beta|\mathbf{w}^{(m)})$ over γ and β via the Newton-Raphson method.

Now suppose that $\theta^{(m)} = (\gamma^{(m)}, \beta^{(m)})$ are the current estimates obtained by the M-step for a given $\mathbf{w}^{(m)}$. This means that the current estimate $\Lambda_0^{(m)}$of Λ_0 can be written as $\widehat{\Lambda}_0(.|\beta^{(m)}, \mathbf{w}^{(m)})$. Further, for the next $(m+1)$-th E-step, $\mathbf{w}^{(m)}$ is updated to $\mathbf{w}^{(m+1)}$ by substituting $(\theta^{(m)},\ \Lambda_0^{(m)})$ into $(\theta,\ \Lambda_0)$ in $\widehat{\Lambda}(t_i|\beta, \mathbf{w}^{(m)})$. Finally, $(\theta^{(\widehat{m})},\ \Lambda_0^{(\widehat{m})})$ which provides $\widehat{m} = \arg\max_{m=0,1...}.l_f(\theta^{(m)},\ \Lambda_0^{(m)})$ becomes an estimate of $(\theta,\ \Lambda_0)$ in a series of the E-M iteration.

Posterior Distribution of Cox-OF model

For the Cox-OF model, we need to assign a prior to each of the following parameters: the baseline hazard h_0 which is labeled as h for convenience, $\beta = \{\beta_1, ..., \beta_{p_1}\}$, and $\gamma = \{\gamma_1, ..., \gamma_{p_2}\}$. For the baseline hazard, we follow standard practice[1] and assume a Gamma prior where both parameters are set equal to 0.001.[2] Hence

$$h\tilde{}\ \text{Gamma}(a_h, b_h) \tag{18}$$

where $a_h = b_h = 0.001$. From the parametric Zombie survival models, we assume the prior of $\beta = \{\beta_1, ..., \beta_{p_1}\}$ as

$$\beta \sim \text{MVN}_{p_1}(\mu_\beta, \Sigma_\beta), \tag{19}$$

Thus the conditional posterior distribution for h_0 and β is respectively given by

$$\pi(h|\mathbf{C}, \alpha, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \beta, \gamma) \propto L(h|\mathbf{C}, \alpha, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \beta, \gamma) \times \pi(h|a_h, b_h) \tag{20}$$

$$\pi(\beta|\mathbf{C}, \alpha, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \gamma) \propto L(\beta|\mathbf{C}, \alpha, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \beta, \gamma) \times \pi(\beta|\mu_\beta, \Sigma_\beta) \tag{21}$$

Further, following the parametric Zombie survival models, we can assign mutivariate Normal prior to $\gamma = \{\gamma_1, ..., \gamma_{p_2}\}$,

$$\gamma \sim \text{MVN}_{p_2}(\mu_\gamma, \Sigma_\gamma), \tag{22}$$

and the corresponding conditional posterior distribution of γ becomes

$$\pi(\gamma|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \beta, \mathbf{h}) \propto L(\gamma|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \beta, \mathbf{h}) \times \pi(\gamma|\mu_\gamma, \Sigma_\gamma). \tag{23}$$

What about the cumulative basline hazard?

MCMC algorithm for Cox-OF model

[1] complete footnote

[2] This implies that the expectation is equal to 1 and the variance is equal to 1000.