

Bayesian Inference on Parametric Overreported Failure (OF) Survival Model

Bomin Kim

May 26, 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 Parametric OF Survival Model

Likelihood function

Recall from Ben's "Parametric Overreported Failure (OF) Survival Model" that the probability of observation i being truly censored (given that the subset of non-censored failure outcomes are being misclassified) is

$$\alpha_i = \Pr(\tilde{C}_i = 0). \quad (1)$$

The unconditional density of $C_i = 1$ is thus given by the combination of an observation's true censoring probability $\Pr(\tilde{C}_i = 0)$ and its probability of survival conditional on being misclassified $\Pr(T_i > t_i, C_i = 1 | \tilde{C}_i = 0)$, plus the observation's true un-censoring probability $\Pr(\tilde{C}_i = 1)$ and its probability of failure conditional on not being misclassified $\Pr(T_i = t_i, C_i = 1 | \tilde{C}_i = 1)$,

$$\alpha_i S(t_i) + (1 - \alpha_i) * f(t_i) \quad (2)$$

And the unconditional survival function for the case of $C_i = 0$ is therefore

$$\alpha_i * S(t_i), \quad (3)$$

where

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

The likelihood function of the Parametric OF Survival Model is defined as

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

And the log likelihood is

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

Posterior Distribution of Parametric OF Survival Models

Exponential

For exponential survival model, the density function and survival function are

$$\begin{aligned} f(t_i|X_i, \boldsymbol{\beta}) &= \exp(X_i\boldsymbol{\beta})\exp(-\exp(X_i\boldsymbol{\beta})t_i) \\ S(t_i|X_i, \boldsymbol{\beta}) &= \exp(-\exp(X_i\boldsymbol{\beta})t_i). \end{aligned} \tag{7}$$

Then, the likelihood function of Exponential OF survival model is

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= \prod_{i=1}^N [\alpha_i \exp(-\exp(X_i\boldsymbol{\beta})t_i) + (1 - \alpha_i) \exp(X_i\boldsymbol{\beta}) \exp(-\exp(X_i\boldsymbol{\beta})t_i)]^{C_i} \\ &\quad \times [\alpha_i \exp(-\exp(X_i\boldsymbol{\beta})t_i)]^{1-C_i} \end{aligned} \tag{8}$$

where X_i is the i^{th} row of the covariate matrix \mathbf{X} .

In the exponential survival model, we assume the prior of $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_{p_1}\}$ as

$$\boldsymbol{\beta} \sim \text{Multivariate Normal}_{p_1}(\boldsymbol{\mu}_\beta, \Sigma_\beta), \tag{9}$$

thus the conditional posterior distribution for $\boldsymbol{\beta}$ parameters is given by

$$\pi(\boldsymbol{\beta}|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\gamma}) \propto L(\boldsymbol{\beta}|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\gamma}) \times \pi(\boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \Sigma_\beta). \tag{10}$$

Moreover, we can also assign mutivariate Normal prior to $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_{p_2}\}$,

$$\boldsymbol{\gamma} \sim \text{Multivariate Normal}_{p_2}(\boldsymbol{\mu}_\gamma, \Sigma_\gamma), \tag{11}$$

and the corresponding conditional posterior distribution of $\boldsymbol{\gamma}$ becomes

$$\pi(\boldsymbol{\gamma}|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}) \propto L(\boldsymbol{\gamma}|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}) \times \pi(\boldsymbol{\gamma}|\boldsymbol{\mu}_\gamma, \Sigma_\gamma). \tag{12}$$

Weibull

If the survival time t has a Weibull distribution of $W(t|\lambda, X_i\boldsymbol{\beta})$, the density function and survival function are

$$\begin{aligned} f(t_i|\lambda, X_i, \boldsymbol{\beta}) &= \lambda(\exp(X_i\boldsymbol{\beta}))(\exp(X_i\boldsymbol{\beta})t_i)^{\lambda-1}\exp(-(\exp(X_i\boldsymbol{\beta})t_i)^\lambda) \\ S(t_i|\lambda, X_i, \boldsymbol{\beta}) &= \exp(-(\exp(X_i\boldsymbol{\beta})t_i)^\lambda), \end{aligned} \quad (13)$$

which shows that $\lambda = 1$ reduces to Exponential survival model, which is a well-known property. The likelihood function of Weibull OF survival model is

$$\begin{aligned} L(\lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \prod_{i=1}^N [\alpha_i \exp(-(\exp(X_i\boldsymbol{\beta})t_i)^\lambda) + (1 - \alpha_i) \lambda(\exp(X_i\boldsymbol{\beta}))(\exp(X_i\boldsymbol{\beta})t_i)^{\lambda-1} \exp(-(\exp(X_i\boldsymbol{\beta})t_i)^\lambda)]^{C_i} \\ &\quad \times [\alpha_i \exp(-(\exp(X_i\boldsymbol{\beta})t_i)^\lambda)]^{1-C_i}. \end{aligned} \quad (14)$$

For the two parameters λ and $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_{p_1}\}$, we assign prior to each parameter as

$$\begin{aligned} \lambda &\sim \text{Gamma}(a_\lambda, b_\lambda) \\ \boldsymbol{\beta} &\sim \text{Multivariate Normal}_{p_1}(\boldsymbol{\mu}_\beta, \Sigma_\beta), \end{aligned} \quad (15)$$

where the conditional distribution for λ and $\boldsymbol{\beta}$ parameters are given by

$$\begin{aligned} \pi(\lambda|\mathbf{C}, \boldsymbol{\alpha}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}, \boldsymbol{\gamma}) &\propto L(\lambda|\mathbf{C}, \boldsymbol{\alpha}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \times \pi(\lambda|a_\lambda, b_\lambda) \\ \pi(\boldsymbol{\beta}|\mathbf{C}, \boldsymbol{\alpha}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\gamma}, \lambda) &\propto L(\boldsymbol{\beta}|\mathbf{C}, \boldsymbol{\alpha}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\gamma}, \lambda) \times \pi(\boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \Sigma_\beta). \end{aligned} \quad (16)$$

Same as Exponential case, we can assign mutivariate Normal prior to $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_{p_2}\}$,

$$\boldsymbol{\gamma} \sim \text{Multivariate Normal}_{p_2}(\boldsymbol{\mu}_\gamma, \Sigma_\gamma), \quad (17)$$

and the corresponding conditional posterior distribution becomes

$$\pi(\boldsymbol{\gamma}|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}, \lambda) \propto L(\boldsymbol{\gamma}|\mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}, \lambda) \times \pi(\boldsymbol{\gamma}|\boldsymbol{\mu}_\gamma, \Sigma_\gamma). \quad (18)$$

Markov chain Monte Carlo (MCMC) algorithm

Prior specifications

Since we do not have any informative prior for parametric OF survival model, we can follow the common approach and specify our hyperparameters as below:

$$\begin{aligned}\boldsymbol{\mu}_\beta &= \mathbf{0}, \Sigma_\beta \sim \text{Inverse-Wishart}(p_1 \mathbf{I}_{p_1}, p_1), \\ \boldsymbol{\mu}_\gamma &= \mathbf{0}, \Sigma_\gamma \sim \text{Inverse-Wishart}(p_2 \mathbf{I}_{p_2}, p_2), \\ a_\lambda &= b_\lambda = 0.001,\end{aligned}\tag{19}$$

where we use hierarchical Bayesian modeling to estimate Σ_β and Σ_γ using Inverse-Wishart distribution. Note that if this step seems to be unnecessary, we can instead simply fix those such as $\Sigma_\beta = \Sigma_\gamma = 10^4 \times \mathbf{I}$ (very slow mixing).

Sampling scheme

In the survival regression setting, closed forms for the posterior distribution of $\boldsymbol{\beta}$ (as well as $\boldsymbol{\gamma}$ in our model) are generally not available (which is also our cases: Equation (10), (12), (16), (18)), and therefore one needs to use numerical integration or Markov chain Monte Carlo (MCMC) methods. Here we will use MCMC methods with the following update scheme:

Step 0. Choose an arbitrary starting point $\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0$, (and λ_0 if Weibull) and set $i = 0$.

Step 1. Update $\Sigma_\beta \sim \pi(\Sigma_\beta | \boldsymbol{\beta}_i)$ and $\Sigma_\gamma \sim \pi(\Sigma_\gamma | \boldsymbol{\gamma}_i)$ using Gibbs sampler.

Step 2. Update $\boldsymbol{\beta} \sim \pi(\boldsymbol{\beta} | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\gamma}, \lambda, \boldsymbol{\mu}_\beta, \Sigma_\beta)$ and $\boldsymbol{\gamma} \sim \pi(\boldsymbol{\gamma} | \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}, \lambda, \boldsymbol{\mu}_\gamma, \Sigma_\gamma)$ using slice sampling.

Step 2'. If Weibull, update $\lambda \sim \pi(\lambda | \mathbf{C}, \boldsymbol{\alpha}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\beta}, \boldsymbol{\gamma}, a_\lambda, b_\lambda)$ using slice sampling.

Step 3. Set $i = i + 1$, and go to Step 1.

Step 4. After N iterations, summarize the parameter estimates using all sampled values (e.g. confidence intervals for coefficient estimation)

Gibbs sampling for Σ_β and Σ_γ

The closed form of full conditional distributions of $\pi(\Sigma_\beta|\beta_i)$ and $\pi(\Sigma_\gamma|\gamma_i)$ in Step 1 are derived as below:

1. Σ_β :

$$\begin{aligned}
\pi(\Sigma_\beta|\beta) &\propto \pi(\beta|\mu_\beta = \mathbf{0}, \Sigma_\beta) \times \pi(\Sigma_\beta) \\
&\propto |\Sigma_\beta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\beta'\Sigma_\beta^{-1}\beta)\right\} \times |\Sigma_\beta|^{-\frac{p_1+p_1+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(p_1\mathbf{I}_{p_1}\Sigma_\beta^{-1})\right\} \\
&= |\Sigma_\beta|^{-\frac{1+p_1+p_1+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\beta\beta'\Sigma_\beta^{-1})\right\} \times \exp\left\{-\frac{1}{2}\text{tr}(p_1\mathbf{I}_{p_1}\Sigma_\beta^{-1})\right\} \\
&= |\Sigma_\beta|^{-\frac{1+p_1+p_1+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}((\beta\beta' + p_1\mathbf{I}_{p_1})\Sigma_\beta^{-1})\right\} \\
&\sim \text{Inverse-Wishart}(\beta\beta' + p_1\mathbf{I}_{p_1}, \quad 1 + p_1)
\end{aligned}$$

2. Σ_γ :

$$\begin{aligned}
\pi(\Sigma_\gamma|\gamma) &\propto \pi(\gamma|\mu_\gamma = \mathbf{0}, \Sigma_\gamma) \times \pi(\Sigma_\gamma) \\
&\propto |\Sigma_\gamma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\gamma'\Sigma_\gamma^{-1}\gamma)\right\} \times |\Sigma_\gamma|^{-\frac{p_2+p_2+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(p_2\mathbf{I}_{p_2}\Sigma_\gamma^{-1})\right\} \\
&= |\Sigma_\gamma|^{-\frac{1+p_2+p_2+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\gamma\gamma'\Sigma_\gamma^{-1})\right\} \times \exp\left\{-\frac{1}{2}\text{tr}(p_2\mathbf{I}_{p_2}\Sigma_\gamma^{-1})\right\} \\
&= |\Sigma_\gamma|^{-\frac{1+p_2+p_2+1}{2}} \exp\left\{-\frac{1}{2}\text{tr}((\gamma\gamma' + p_2\mathbf{I}_{p_2})\Sigma_\gamma^{-1})\right\} \\
&\sim \text{Inverse-Wishart}(\gamma\gamma' + p_2\mathbf{I}_{p_2}, \quad 1 + p_2)
\end{aligned}$$

Slice Sampling for β, γ and λ

In recent decade, slice sampling has been widely used as an alternative to Metropolis-Hastings algorithm. Following the current practice of Bayesian mixture survival model, we use univariate slice sampler with stepout and shrinkage (Neal, 2003) in Step 2 (and Step 2' if Weibull), where closed form of full conditional distribution is intractable. We also follow the modifications made in 'BayesMixSurv' R package (Mahani, Mansour, and Mahani, 2016). Below are the steps to perform slice sampling for β , and slice sampling for γ and λ could be done in the exactly same manner:

For β_p , $p = 1, \dots, P$,

Step 0. Choose an arbitrary starting point β_{p_0} and size of the slice w , and set $i = 0$.

Step 1. Draw y from $\text{Uniform}(0, f(\beta_{p_0}))$ defining the slice $S = \{\beta_p : y < f(\beta_p)\}$, where

$$\begin{aligned} f(\beta_p) &\propto \pi(\beta_p | \boldsymbol{\beta}_{-p}, \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\gamma}) && \text{if Exponential (Eq. (10))} \\ &\propto \pi(\beta_p | \boldsymbol{\beta}_{-p}, \mathbf{C}, \mathbf{X}, \mathbf{Z}, \mathbf{t}, \boldsymbol{\gamma}, \lambda) && \text{if Weibull (Eq. (16))} \end{aligned}$$

Step 2. Find an interval, $I = (L, R)$, around β_{p_0} that contains all, or much, of the slice, where the initial interval is determined as

$$u \sim \text{Uniform}(0, w)$$

$$L = \beta_{p_0} - u$$

$$R = \beta_{p_0} + (w - u)$$

and expand the interval until its ends are outside the slice or until the limit on steps (limit on steps = m) is reached ("stepping-out" procedure), by comparing y and $(f(L), f(R))$.

$$J = \text{Floor}(\text{Uniform}(0, m))$$

$$K = (m - 1) - J$$

Repeat while $J > 0$ and $y < f(L)$:

$$L = L - w, J = J - 1$$

Repeat while $K > 0$ and $y < f(R)$:

$$R = R + w, K = K - 1$$

Step 3. Draw a new point β_{p_1} from the part of the slice within this interval I , and shrink the interval on each rejection ("shrinkage" procedure)

Repeat

$\beta_{p_1} \sim \text{Uniform}(L, R)$

if $y < f(\beta_{p_1})$, accept β_{p_1} then exit loop

if $\beta_{p_1} < \beta_{p_0}$, then $L = \beta_{p_1}$

else $R = \beta_{p_1}$

Step 4. Set $i = i + 1$, $\beta_{p_0} = \beta_{p_1}$, and go to Step 1.

Step 5. After N iterations, summarize the parameter estimates using all sampled values (e.g. confidence intervals for coefficient estimation)