PH 716 Applied Survival Analysis

Part IV: Accelerated Failure Time Model

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2024/Feb/07 11:06:44

Assumptions

- T_i are independent across i
  - NO longer assumed to share the identical distribution
  - i.e., "personalized" or "individualized"
- log-linear model: $\ln T_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \sigma\varepsilon_i$
  - Unknown parameters $\sigma > 0$ and $\beta_j \in \mathbb{R}$
  - Error terms ε_i are iid
- Equiv. $T_i = \exp(\beta_0 + \varepsilon_i)\prod_{j=1}^p \exp(x_{ij}\beta_j)$
  - (Why is called "accelerated failure time model"?) The effect of covariates acts multiplicatively on the survival time and accelerates or decelerates the progress along the time axis.

Parameter interpretation

- β_0 is the baseline of logarithm of survival times. This baseline refers to the scenario where the effect of covariates is neutral (i.e., all β_j, $j > 0$, are all zeros).
- The interpretation of β_j, $j > 0$, is based on controlling covariates associated with other coefficients, i.e., $x_{i1}, \ldots, x_{i,j-1}, x_{i,j+1}, \ldots, x_{ip}$.
- Holding values of other covariates, a unit increase in x_{ij} corresponds to an increase of β_j in the mean of $\ln T_i$. More specifically, it shifts the distribution of $\ln T_i$ to the left by the amount β_j. Or, equivalently, all percentiles of the distribution of $\ln T_i$ are shifted to the left by β_j. Correspondingly, the percentiles of T_i are multiplied by the constant e^{β_j}.

Survival function

- If $\varepsilon_i \overset{iid}{\sim} N(0,1)$,
  - $S_{T_i}(t) = \Pr(\ln T_i > \ln t) = \Pr\{\varepsilon_i > \sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\} = 1 - \Phi\{\sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\}$
    - $\Phi(\cdot)$: the cdf of $N(0,1)$
  - i.e., $T_i \sim \text{log-normal}(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j, \sigma^2)$
- If $\varepsilon_i \overset{iid}{\sim}$ the standard Gumbel distribution for minimum (i.e., $F_{\varepsilon_i}(\epsilon) = 1 - \exp(-\exp\epsilon)$),
  - P.S. $\min(X_1, X_2, \ldots, X_n) - \ln n \xrightarrow{d}$ standard Gumbel distribution (for minimum) as $n \to \infty$ if $X_i \overset{iid}{\sim} \exp(1)$
  - $S_{T_i}(t) = \Pr\{\varepsilon_i > \sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\} = 1 - F_{\varepsilon_i}\{\sigma^{-1}(\ln t - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)\} = \exp[-t^{1/\sigma}\exp\{-(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)/\sigma\}] = \exp[-\{t/\exp(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)\}^{1/\sigma}]$
  - i.e., $T_i \sim$ Weibull with $1/\sigma$ as the "shape" and $\exp(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)$ as the "scale"
    - Specifically, $T \sim$ exponential if $\sigma = 1$

Likelihood principles (for uncensored data)

- Observed $T_1 = t_1, \ldots, T_n = t_n$
- Joint density of $\mathbf{T} = [T_1, \ldots, T_n]^\top$ evaluated at $[t_1, \ldots, t_n]^\top$: $f_{\mathbf{T}}(t_1, \ldots, t_n; \boldsymbol{\theta})$
  - $\boldsymbol{\theta}$: a p-vector of unknown parameters
- Observed-data likelihood $L(\boldsymbol{\theta}) = f_{\mathbf{T}}(t_1, \ldots, t_n; \boldsymbol{\theta})$
  - Taken as a function of $\boldsymbol{\theta}$
  - $L(\boldsymbol{\theta}) = \prod_{i=1}^n f_{T_i}(t_i; \boldsymbol{\theta})$ if T_i is independent across i
- Maximum likelihood (ML) estimator: $\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$
  - $\ell(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta})$
  - A closed-form solution for $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$ usually not available
- Fisher information (the expectation of Hessian matrix of $\ell(\boldsymbol{\theta})$): $I(\boldsymbol{\theta}) = -\mathrm{E}\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \approx -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\mathrm{ML}}}$
- $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{ML}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, nI^{-1}(\boldsymbol{\theta}))$ for iid T_i
  - $\boldsymbol{\theta} \approx N(\hat{\boldsymbol{\theta}}_{\mathrm{ML}}, I^{-1}(\hat{\boldsymbol{\theta}}))$ for iid T_i
- Likelihood ratio test
  - H_0 vs H_1
  - Test statistic: $-2 \ln \frac{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})}$
    - $\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0}$: the (constrained) MLE under H_0
    - $\hat{\boldsymbol{\theta}}_{\mathrm{ML}}$: the MLE under $H_0 \bigcup H_1$
  - Reject H_0 if the value of $-2 \ln \frac{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})}$ is over $\chi^2_{p,1-\alpha}$
    - $\chi^2_{p,1-\alpha}$: the $1-\alpha$ quantile of $\chi^2(p)$
    - Because $-2 \ln \frac{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML},H_0})}{L(\hat{\boldsymbol{\theta}}_{\mathrm{ML}})} \approx \chi^2(p)$
      - p: the difference of free parameters with and without H_0
- Pros and cons
  - Clear pathway
  - Exact inference only available for selected (and really simple) cases, i.e., approximations usually employed
  - MLE considered (approximately) the most efficient in regular cases
  - LRT optimal for simple cases but well accepted even in complex cases

Likelihood principles (for right-censored data)

- Observed $\widetilde{T}_i = \tilde{t}_i$ and $\Delta_i = \delta_i$ (event indicator),
  - $\widetilde{T}_i$: the smaller one between T_i (event time) and C_i (right-cencoring time)
  - Assuming the independence across i
  - Assuming the noninformative censoring, i.e.,
    - $T_i \perp C_i$
    - $S_{T_i}(t \mid \boldsymbol{\theta})$ and $S_{C_i}(t \mid \boldsymbol{\eta})$ have NO common parameter
- Joint density of $\widetilde{T}_i$ and Δ_i: $f_{T_i}(t \mid \boldsymbol{\theta}) S_{C_i}(t \mid \boldsymbol{\eta})$ if
  - $\Pr(\widetilde{T}_i > t, \Delta_i = 1) = \Pr(C_i \geq T_i, T_i > t) = \int_t^\infty \Pr(C_i \geq u, T_i = u) \mathrm{d}u = \int_t^\infty S_{C_i}(u \mid \boldsymbol{\eta}) f_{T_i}(u \mid \boldsymbol{\theta}) \mathrm{d}u \Rightarrow f_{\widetilde{T}_i, \Delta_i}(\tilde{t}_i, \delta_i) =$
    - $f_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta}) S_{C_i}(\tilde{t}_i \mid \boldsymbol{\eta})$ if $\delta_i = 1$
    - $S_{T_i}(\tilde{t}_i \mid \boldsymbol{\theta}) f_{C_i}(\tilde{t}_i \mid \boldsymbol{\eta})$ if $\delta_i = 0$

- Observed-data likelihood: $L(\boldsymbol{\theta},\boldsymbol{\eta}) = \prod_{i=1}^n f_{\widetilde{T}_i,\Delta_i}(\tilde{t}_i,\delta_i) = \prod_{i=1}^n \{f_{T_i}(\tilde{t}\mid\boldsymbol{\theta})S_{C_i}(\tilde{t}\mid\boldsymbol{\eta})\}^{\delta_i}\{S_{T_i}(\tilde{t}\mid\boldsymbol{\theta})f_{C_i}(\tilde{t}\mid\boldsymbol{\eta})\}^{1-\delta_i}$
  - Reducing to $\prod_{i=1}^n f_{T_i}(\tilde{t}_i\mid\boldsymbol{\theta})^{\delta_i}S_{T_i}(\tilde{t}_i\mid\boldsymbol{\theta})^{1-\delta_i}$ if we are only concerned about the MLE of $\boldsymbol{\theta}$

Likelihood principles (for general censored data)

- Assuming the independence across i and noninformative censoring
- Observed-data likelihood:

$$\prod_{i\in\mathfrak{D}} f_{T_i}(\tilde{t}_i)\prod_{i\in\mathfrak{R}} S_{T_i}(\tilde{t}_i)\prod_{i\in\mathfrak{L}}\{1-S_{T_i}(\tilde{t}_i)\}\prod_{i\in\mathfrak{I}}\{S_{T_i}(\tilde{t}_{iL})-S_{T_i}(\tilde{t}_{iR})\}$$

  - $\mathfrak{D}$: the set of **unobserved** subjects
  - $\mathfrak{R}$: the set of **right-censored** subjects
  - $\mathfrak{L}$ the set of **left-censored** subjects
  - $\mathfrak{I}$: the set of **interval-censored** subjects

---

- Ex 4.1 ([DM] pp.147): The purpose of Steinberg et al. (2009) was to evaluate extended duration of a triple-medication combination versus therapy with the nicotine patch alone in smokers with medical illnesses.

```
head(asaur::pharmacoSmoking)
data.ex41 = asaur::pharmacoSmoking
data.ex41 = data.ex41[data.ex41$ttr != 0,] # ttr=0 not allowed in AFT models
is.factor(data.ex41$grp)
aft.ex41 = survival::survreg(
  survival::Surv(ttr, relapse) ~ grp,
  data = data.ex41,
  dist="weibull") # assume weibull for T
summary(aft.ex41) # Confused about "scale" in the output? Check ?survival::survreg.distributions

# prediction for grp='combination'
shape = 1/aft.ex41$scale
scale = unname(exp(aft.ex41$coefficients[1])) # scale
(ET = scale*gamma(1+1/shape)) # expectation of T
(medT = scale*log(2)^(1/shape)) # median of T
surv.fun = function(t){ # survival function
  return(
    1-pweibull(t, shape = shape, scale = scale)
  )
}
curve(surv.fun, from = 0, to = 1e3) # plot the survival curve
```

- Ex. 4.2. (Revisiting the data of Bladder Cancer Recurrences) A dataset on recurrences of bladder cancer. It contains three treatment arms for 118 subjects.

```
data.ex42 = survival::bladder1[
  complete.cases(
    survival::bladder1[,c('id', 'treatment', 'start', 'stop', 'status')]
  ),
  c('id', 'treatment', 'start', 'stop', 'status')
]
data.ex42$status = 1*(data.ex42$status %in% c(1,2,3)) # merging status 1, 2,3
```

```
data.ex42$tte = data.ex42$stop - data.ex42$start
data.ex42 = data.ex42[data.ex42$tte != 0,] # ttr=0 not allowed in AFT models
is.factor(data.ex42$treatment)
aft.ex42 = survival::survreg(
  survival::Surv(tte, status) ~ treatment,
  data = data.ex42,
  dist="lognormal") # assume lognormal for T
summary(aft.ex42)

# prediction for treatment='pyridoxine'
sigma = aft.ex42$scale
mu = sum(aft.ex42$coefficients[1:2]) # scale
(ET = exp(mu+sigma^2/2)) # expectation of T
(medT = exp(mu)) # median of T
surv.fun = function(t){ # survival function
  return(
    1-pnorm((log(t)-mu)/sigma)
  )
}
curve(surv.fun, from = 0, to = 1e2) # plot the survival curve
```

Pros and cons

- Easy to interprete coeffcients: effects on the failure time directly
- Distribution assumptions may be too strong