Applied Survival Analysis Using R
Chapter 12: Additional Topics

Qi Guo

Department of Mathematical Sciences
The University of Texas at Dallas

April, 20 2019

What is "`piecswise`"?

- The *piecewise exponential model* is a *generalization* of the exponential which can offer *considerable flexibility* for modeling.
- "`pieces`" means *the survival time axis* was divided into multiple intervals and the *hazard is constant*.
- *The piecewise exponential is that the likelihood is equivalent to a Poisson likelihood*.
- If it is a single piece, that means it's just an ordinary exponential distribution with rate parameter λ.

$$L_e(\lambda) = \prod_{i=1}^{n} h(t_i)^{\delta_i} S(t_i) = \prod_{i=1}^{n} \lambda^{\delta_i} e^{-\lambda t_i} = \lambda^d e^{-\lambda V} \tag{1}$$

where t_i is the failure time of the ith subject, $V = \sum t_i$ is the total time at risk.

Poisson Distribution

- In Chapter 2, the m.l.e is given by $\hat{\lambda} = d/V$, now we suppose the d(the number of death) has a *Poisson distribution* with mean μ, so $\mu = V\lambda$.
- The likelihood function for a Poisson distribution is:

$$L_p(\lambda) = (\lambda V)^d e^{-\lambda V} = \lambda^d e^{-\lambda V} \cdot V^d \tag{2}$$

Clearly the Poisson likelihood L_p is *proportional* to the exponential likelihood L_e, the constant multiple being V^d

Notations

- Suppose that we divide the time axis into contiguous intervals using cut points 0, c_1, c_2,..., c_k.
- For each subject, say i, we denote the time spent in each interval by t_{i1}, t_{i2},...,$t_{ik'}$, where k' denotes the time interval subject i dies, or the *largest* time interval subject i is still known to be alive.
- Define δ_{ij} to be 0 for each interval j in which the ith subject is known to be alive, and 1 for interval k if the subject died in that interval.
- Then for patient i, the survival time of that patient is $t_i = \sum_{j=1}^{k} t_{ij}$ and the censoring indicator is $\delta_i = \sum_{j=1}^{k} \delta_{ij}$.

Formulas

- Assume a proportional hazards model

$$\lambda_i(t_i, \beta) = \lambda_0(t_i) e^{z_i \beta} \tag{3}$$

where now the *baseline hazard* is a *piecewise* exponential,$\lambda_0(\mu) = \lambda_j$, j being jth interval, the one in which μ falls, so the full likelihood is:

$$L_{pe}(\lambda_1, \lambda_2, ..., \lambda_k, \beta) = \prod_{i=1}^{n} \prod_{j=1}^{k'(i)} {\lambda_{ij}}^{\delta_{ij}} e^{-\lambda_{ij} t_{ij}} \tag{4}$$

where $\lambda_{ij} = \lambda_j e^{z_i \beta}$, which is the proportional hazards assumption for the piecewise exponential

- With the single exponential, treat the censoring indicators δ_{ij} as a Poisson distribution with mean $\lambda_{ij} t_{ij}$

m.l.e

- so the maximum likelihood estimates obtained from the Poisson model with

$$\log(\lambda_{ij}) = \log(\lambda_j) + z_i\beta = \alpha_j + z_i\beta \tag{5}$$

- so for the piecewise exponential proportional hazards model

$$\log(\lambda_{ij}t_{ij}) = \log(\lambda_j) + z_i\beta + \log(t_{ij}) = \alpha_j + z_i\beta + \log(t_{ij}) \tag{6}$$

where the α_j are the logs of the baseline hazard coefficients and β,is the log of the hazard ratio for z_i, The constant log.tij/ is called an *offset*.

Example

- Consider a synthetic example in Chapter 4.

```
> tt <- c(6, 7, 10, 15, 19, 25)
> delta <- c(1, 0, 1, 1, 0, 1)
> trt <- c(0, 0, 1, 0, 1, 1)
> id <- 1:6
> simple <- data.frame(id, tt, delta, trt)
> simple
   id   tt   delta   trt
1   1    6     1      0
2   2    7     0      0
3   3   10     1      1
4   4   15     1      0
5   5   19     0      1
6   6   25     1      1
```

- Define three intervals that we will use to define a piecewise exponential distribution, and partition the data into events and person-weeks per interval:

Example

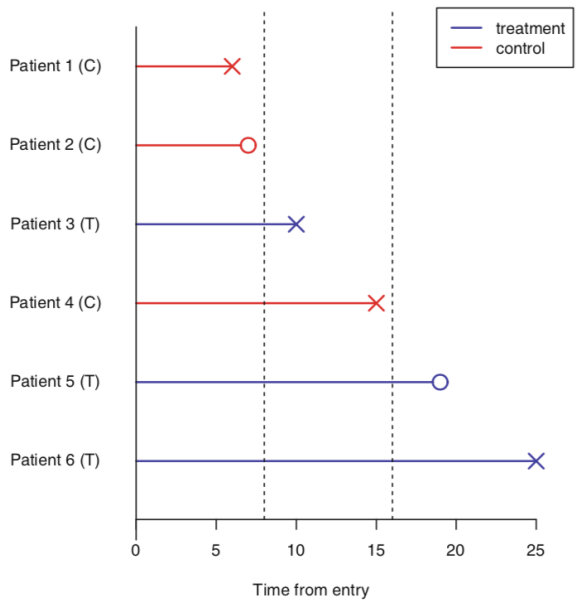

Fig. 12.1 Survival times of a synthetic data set. The intervals defined by the cut points tau are shown using *vertical dotted lines*

```
> tau.s <- c(0, 8, 16, 30)
> simple.split.s <- survSplit(data=simple, cut=tau.s, end="tt",
start="t0", event="delta", episode="diagGrp")
> simple.split.s$expo <- simple.split.s$tt - simple.split.s$t0
> ord <- order(simple.split.s$id)
> simple.split.ord <- simple.split.s[ord,]
> simple.split.ord
```

Example

	id	tt	delta	trt	t0	diagGrp	expo
7	1	6	1	0	0	1	6
8	2	7	0	0	0	1	7
9	3	8	0	1	0	1	8
15	3	10	1	1	8	2	2
10	4	8	0	0	0	1	8
16	4	15	1	0	8	2	7
11	5	8	0	1	0	1	8
17	5	16	0	1	8	2	8
23	5	19	0	1	16	3	3
12	6	8	0	1	0	1	8
18	6	16	0	1	8	2	8
24	6	25	1	1	16	3	9

- To obtain parameter estimates for the piecewise exponential distribution in (6)

```
> result.simple.poisson <- glm(delta ~ -1 + factor(diagGrp)+trt+
offset(log(expo)), family=poisson, data=simple.split.ord)
> summary(result.simple.poisson)
```

Example

```
Coefficients:                Estimate  Std. Error  z value   Pr
(>|z|)
factor(diagGrp)1   -3.2942   1.0370    -3.177      0.00149 **
factor(diagGrp)2   -1.7463   0.8569    -2.038      0.04156 *
factor(diagGrp)3   -1.0912   1.5949    -0.684      0.49389
trt                -1.3937   1.2425    -1.122      0.26199
---
Signif. codes:  0*** 0.001 ** 0.01 * 0.05 . 0.1    1
```

- The "`-1`" in the model definition tells R to fit a separate term for each of the three interval factors (rather than to consider the first level as a reference level), more details see book!

Right Censoring and Interval Censoring

- Right censoring occurs naturally in clinical trials in Chapter 1, and the general likelihood function:

$$L(\beta) = \prod_{i \in D} f(t_i) \cdot \prod_{i \in C} S(t_i) \tag{7}$$

where D represents the set of all subjects who fail and R the set of all who are right-censored.

- Suppose that some of the observations are *interval-censored*, so that,patient k has an event that lies between times L_k and R_k.

$$L(\beta) = \prod_{i \in D} f(t_i) \cdot \prod_{i \in C} [S(t_i) - S(R_i)] \tag{8}$$

Example

- If an event is only known to have occurred before a particular time R_i, this is known as *left censoring*, and accommodated by setting $L_i = 0$
- A special optimization technique known as the *Expectation -Maximization algorithm* to get m.l.e.

Breast Cosmesis Study

Fit a proportional hazards model to interval-censored data, and illustrated it using data from a breast cosmesis study. In this study, 94 breast cancer patients treated with radiation therapy with or without adjuvant chemotherapy were followed to determine the time until cosmetic deterioration (specifically, the appearance breast retraction) of the treated breast. Since patients were evaluated at office visits separated by a number of months, the data were interval-censored.

Example

```
> library(interval)
> data(bcos)
> bcos[c(1,33, 47, 62, 90),]
       left    right    treatment
1        45      Inf          Rad
33        0        5          Rad
47        8       12      RadChem
62       14       17      RadChem
90       16       60      RadChem
```

- Patient 47 received radiation and adjuvant chemotherapy, had not had the event at an office visit at 8 months, but the breast retraction had been observed at the next visit four months later. Thus, the event took place sometime between 8 and 12 months

Example

- obtain m.l.e of the survival distributions and plot:

```
> icout <- icfit(Surv(left,right,type="interval2")~treatment,
data=bcos, conf.int=F)
> plot(icout, XLAB="Time in months", YLAB="Survival probability",
COL=c("lightblue", "pink"), LEGEND=F, estpar=list(col=
c("blue", "red"), lwd=2, lty=1))
> legend("bottomleft",legend=c("Radiation alone", "Radiation and
chemo"), col=c("blue","red"), lwd=2)
```

- fit a Weibull proportional hazards model to the interval-censored data, first we must define modified left and right endpoints of the intervals and set a maximum possible time.

Example

```
> bcos <- within(bcos, {
left.alt <- left
left.alt[left == 0] <- 0.1
right.alt <- right
right.alt[is.infinite(right)] <- 65})
> bcos[c(1,33, 47, 62, 90),]
      left    right    treatment    right.alt    left.alt
1       45      Inf          Rad           65        45.0
33       0        5          Rad            5         0.1
47       8       12      RadChem           12         8.0
62      14       17      RadChem           17        14.0
90      16       60      RadChem           60        16.0
```

- fit the Weibull model and add the fitted survival curves as follows:

Example

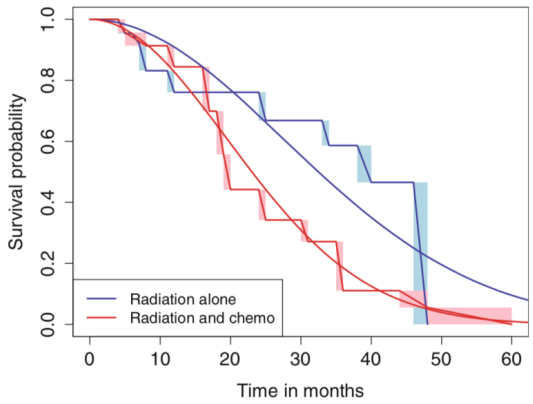

Fig. 12.6 Survival times (times to appearance of breast retractions) for breast cancer patients treated with radiation alone (*blue*) or radiation with adjuvant chemotherapy (*red*). The *smooth lines* are from a proportional hazards (also accelerated failure time) Weibull fit to the interval-censored data, whereas the step functions are non-parametric proportional hazards estimates. The *shaded areas* represent intervals on which the form of the survival curve estimate is ambiguous

```
> bcos.survreg <-survreg(Surv(left.alt, right.alt, type="interval2") ~
treatment, dist="weibull", data=bcos)
> pct <- 1:999/1000
> ptime <- predict(bcos.survreg, type='quantile',
newdata=data.frame(treatment=c("Rad", "RadChem")),p=pct, se=F)
> ines(ptime[1,], 1-pct, xlab="Hours", ylab="Survival", type='l',
lty=c(1,2,2), lwd=c(2,1,1), xlim=c(0,20), col="blue")
> lines(ptime[2,], 1-pct, xlab="Hours", ylab="Survival", type='l',
lty=c(1,2,2), lwd=c(2,1,1), xlim=c(0,20), col="red")
```

select covariates

- In some applications, by contrast, our interest focuses on the predictive ability of a set of covariates. the biomarker focus on using those measurements to *predict* a patient's survival prospects.
- An important such method is the "`lasso`" procedure in the R package "`penalized`"
- This approach *maximizes* the partial likelihood function $\ell(\beta) = \log L(\beta)$

$$L(\beta) = \prod_{j=1}^{D} \frac{h_0(t_j)\psi_j}{\sum_{k \in R_j} h_0(t_j)\psi_k} = \prod_{j=1}^{D} \frac{\psi_j}{\sum_{k \in R_j} \psi_k} \tag{9}$$

lasso procedure

- The parameter satisfies the *constraint* $\sum_{j=1}^{p} |\beta_j| \leq t$ for a constant t, where p is the number of parameters,this may be shown to be equivalent to maximizing the penalized likelihood which, for a pre-specified value of λ.

$$\ell_{pen}(\beta) = \ell(\beta) - \lambda \sum_{j=1}^{p} |\beta_j| \tag{10}$$

- A sufficiently large value of λ will result in *no covariates* at all in the model, and smaller λ values will cause non-zero coefficient estimates.
- How do we select λ?

Cross Validation

- The goal of the procedure is *accurate prediction*, to add the predictive accuracy, we have a procedure called *cross validation*.
- Cross Validation:
  - Start with an initial value of λ and randomly divide the data set into *five* subsets of approximately equal size.
  - Select one of the subsets to be what we shall call the "`validation`" set(20%), and combine the remaining subsets into what we shall call the "`training`" set(80%).
  - Use the training set to construct the lasso model based on (10), and use this model to predict the survivals of patients in the validation set.

Cross Validation

- Use a partial-likelihood-based measure of goodness-of-fit to these data.
- *Repeat* this four more times, with each of the remaining four subsets in turn playing the role of the 20% validation set and the others being the training set.
- Derive an *average* partial-likelihood goodness of fit, and repeated for a range of values of λ, and we select that value that produces the *optimum* goodness of fit.

Example

Example 2

The data contains 17 clinical and biomarker measurements on 227 patients, as well as overall survival and time to recurrence, both recorded in months. Of the 227 patients, 117 have levels of a variety of chemokines and other markers, some representing levels in the tumor itself and some outside the tumor, use 26 of these measurements (columns 23 to 48) as potential predictors of overall survival using a lasso model

- Selecting the 117 patients with complete data:

```
hepatoCellularNoMissing<-hepatoCellular[
complete.cases(hepatoCellular),]
```

Example

- "OS" (overall survival), "Death"(censoring) and a few of the cytokine measurements

```
> hepatoCellularNoMissing[c(1,5,12),c(16,17, 23:27)]
     OS Death       CD4T       CD4N       CD8T     CD8N       CD20T
1    83     0  2.600000   0.000000   190.6000   126.80   20.950000
76   20     1 14.450000   2.758621     2.1500    38.95   26.100000
131  35     1  2.821133   8.294828     8.0064    62.64    2.821133
```

- Then we fit a simple lasso model using the 26 predictors (columns 23 to 48), and we fix the penalty at $\lambda = 10$.

```
> attach(hepatoCellularNoMissing)
> library(penalized)
> hepato.pen <- penalized(Surv(OS,Death),
penalized=hepatoCellularNoMissing[,23:48],
standardize=T, lambda1=10) # nonzero coefficients: 7
```

Example

- List their values using the "`coef`" function

```
> round(coef(hepato.pen, standardize=T), 3)
 CD8N  CD68T   CD4TR   CD8TR  CD68TR    Ki67    CD34
0.104  0.258  -0.035  -0.096   0.111   0.285  -0.013
```

- Use cross- validation to select a value that optimizes the predictive ability of the lasso model, as defined by maximizing the cross-validated partial log-likelihood(CVL).

```
> set.seed(34)
> hepato.prof <- profL1(Surv(OS, Death),
penalized=hepatoCellularNoMissing[,23:48],
standardize=T, fold=10, minlambda1=2, maxlambda1=12)
> plot(hepato.prof$cvl ~ hepato.prof$lambda, type="l", log="x",
xlab="lambda", ylab="Cross-validated log partial likelihood")
```

- "`set.seed`" is to set the random number seed so that we can reproduce this model fit exactly.

Example

- To find the optimal value, we use "`OptL1`" with the same starting seed.

```
> set.seed(34)
> hepato.opt <- optL1(Surv(OS, Death),
penalized=hepatoCellularNoMissing[,23:48], standardize=T,
fold=10)
> hepato.opt$lambda
[1] 8.242321
> abline(v=hepato.opt$lambda, col="gray")
```

- The optimal value is 8.24.

Example

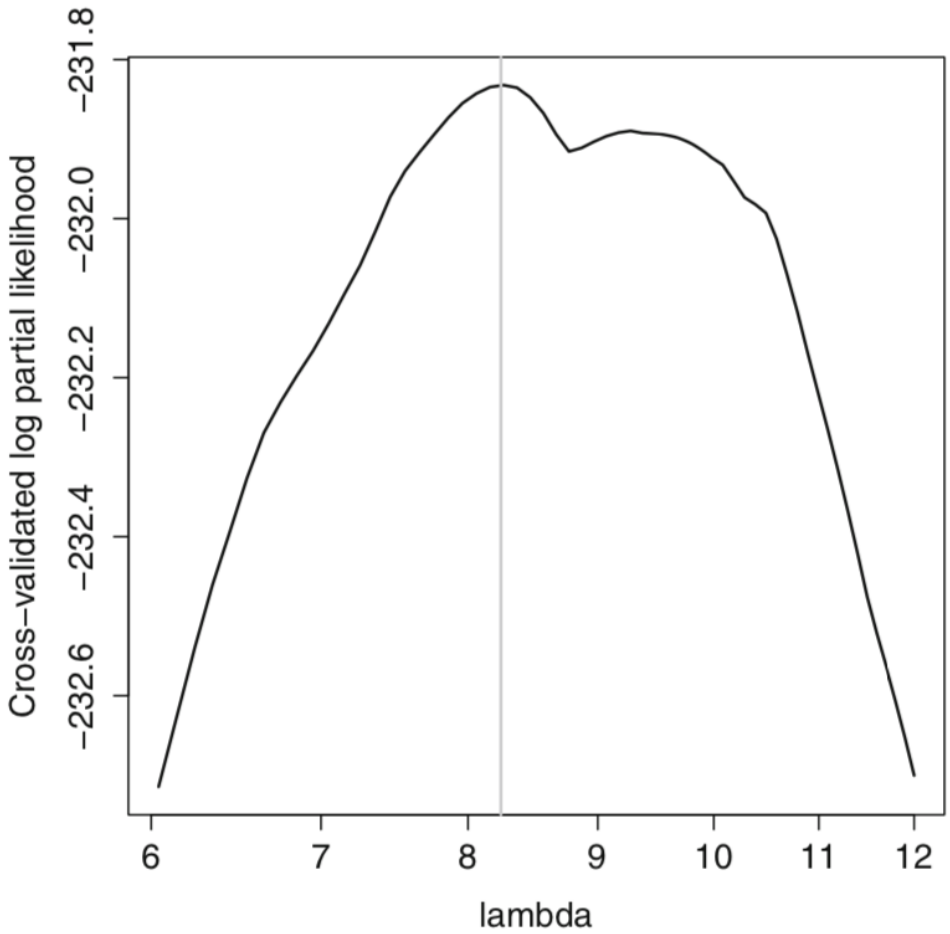

Fig. 12.7 Cross-validated log partial likelihood for a range of values of lambda for the hepatocellular data. The *vertical gray line* shows the global maximum at $\lambda = 8.24$, which was obtained using the "OptL1" function in the "penalized" package