Cox-Poisson Survival Analysis Program

Edward Hughes

Edward Hughes Consulting

Collaborators:

Renjun Ma, Richard Burnett, Michael Jerrett, Yuanli Shi, Daniel Krewski

Problems:

- Cox regression for survival data with random effects
- Case-Cohort designs possible
- Binary-Logistic and Poisson GLM’s coming soon

References

- Therneau, Terry M. and Grambsch, Patricia M.: *Modelling Survival Data*; New York: Springer-Verlag, 2000.
- Ma, Renjun, Krewski, Daniel, and Burnett, Richard T.: "Random effects Cox models: A Poisson modelling approach". *Biometrika* v. 90 (2003), No. 1, 157-169.
- P. K Andersen and R. D. Gill; "Cox's regression model for counting processes: a large sample study*"; Annals of Statistics* v. 10, 1100-1120.
- T. R. Fleming and D. P. Harrington; *Counting Processes and Survival Analysis*; Wiley, 1991.

Features:

- Large data files
- Complicated random effects structures
- Designed to resemble the Therneau survival package supplied with R and S-plus
- Runs with R interface
- Runs under Unix, or Windows with Cygwin

Cox Survival Model

Let h = hazard function, R =matrix of covariates, R^e = covariate row vector for individual e, β = the vector of regression coefficients. Then

$$h^e(t) = h_0(t) U^{r(e)} \exp(R^e(t)\beta)$$

where *r(e)* is the finest-level cluster that individual e belongs to, and U^r is the random effect for cluster r.

Data

Format: anything you can read into an R data frame.

Observations (rows, records) and variables (columns): can have variable-names as first record.

If Covariates are *not* time-dependent: one data record per subject.

Stratification

- Strata: a variable in the data set, say stratumVar, defines stratification: it must be either integer-valued or a factor.

Stratification is specified by

```
... <- CoxPoiss( ... , strata = stratumVar)
```

Each stratum has its own baseline hazard function.

Regression Model Features:

Most of the R facilities for defining regression models are available, such as transformations, interactions, and combinations of variables.

Transformations:

```
ratmod <- survProps(endtime=SurvTime,
                    event=Event) ~ log(Treatment)
```

Interactions:

```
... ~ X + Z*W      # X, Z, W, and interaction of Z and W
```

Combinations:

```
... ~ I(X + Z*W)   # Arithmetic combination X + Z*W
```

Restriction: All variables so used must be in the primary table, not in secondary.

Time-dependent Covariates

Covariates:

NTD: Values depend on individual, not time: R(k)

TD: Values depend on individual and time: R(k,t)

For TD covariates, two ways to handle:

- Record Repetition (like Therneau code)
- Secondary File

TD covariates are *piecewise-constant* in time: best if all TD covariates have the same breakpoints.

Record Repetition

Each data record is repeated for each value of the TD covariates: must use *interval notation* in defining the regression model. A new record is needed for every change in *any* covariate.

Id	Start	End	Death	X	Z	W	…
37	0	10	0	3.2	0	18.8	…
37	10	20	0	4.3	1	18.8	…
37	20	30	0	6.7	3	18.8	…
	. . .						
37	50	57	1	7.4	2	18.8	…
38	0	10	0	1.6	1	22.3	…
38	10	20	0	0.9	0	22.3	…
38	20	. . . . .					

Record Repetition 2

With this data set, we define the regression model

```
TDmod <- survProps(starttime=Start, endtime=End,
                   event=Death) ~ X + Y + W

datSrc <-   . . .
reffs    <- randEffects( . . .
stratVar <-   . . .
TDobj <- CoxPoiss(model = TDmod,
        primary = datSrc, RandomEffects = reffs,
        strata = stratVar)
```

If most of the covariates are TD, record repetition is the best choice.

Secondary Data File

If only a few of the covariates are TD, we can put them into a *secondary table:*

key	start	end	cov1	cov2	. . .
city	StTime	EndTime	X	Z	. . .
17	0	4	3.2	4	. . .
17	4	8	2.8	5	. . .
. . .					

Example: monthly-average readings of air pollution by *city*. Then *city* would be the *key* variable, appearing in both primary and secondary files.

Secondary Data File 2

Primary:

Id	City	Strt	End	Dth	P	W
.	.	.	.	.		
37	14	0	57	1	2.6	8.6
38	23	12	68	0	3.8	17.3
39	61	0	71	1	6.2	4.8
40	14	16	36	0	4.1	12.7
.	.	.	.			

Key Variable must have same name in both tables

Secondary:

City	Starttm	Endtm	X	Z
.	.	.	.	
14	0	8	3.7	0
14	8	12	3.2	0
14	12	16	4.3	1
	.	.	.	
14	65	73	5.2	2
15	0	8	2.2	1
	.	.	.	
15	.	.		