Competing Risks with R*

Simulated Data

Here is a simple model for competing risks. Time is always discrete in practice.

Roll a die.

1 = Cause of death 1
2 = Cause of death 2
3 = Censored
4,5,6 = Roll again

Time to event is the number of rolls.

Dependence on x variables through an extension of logistic regression (multinomial logit model).

In the simulated data,

X_1 is positively related to the hazard of death, but not to transplant.
X_2 is positively related to the "hazard" of transplant, but not to death.
X_3 is negatively related to both death and transplant, but more strongly for death.

There are individual tendencies toward the outcomes (random effects).

```
> rm(list=ls()); options(scipen=999)
> source('simdata.txt')
>
> head(datta)
  id   x1   x2   x3 Time Outcome
1  1 1.40 1.55 4.39   53       1
2  2 3.16 1.92 6.82  199       1
3  3 3.61 3.47 5.28   26       0
4  4 1.81 2.11 3.66   77       1
5  5 2.27 3.59 5.16    8       1
6  6 3.59 2.66 4.48   13       1
>
```

Outcome:

0 = Censored
1 = Death
2 = Transplant

```
> # Simple one-at-a-time
> library(survival)
> coxph(Surv(Time,Outcome==1) ~ x1+x2+x3, data=datta) # Death
Call:
coxph(formula = Surv(Time, Outcome == 1) ~ x1 + x2 + x3, data = datta)

       coef exp(coef) se(coef)     z        p
x1  0.31403   1.36892  0.09956  3.15   0.0016
x2 -0.00942   0.99062  0.09414 -0.10   0.9203
x3 -0.44342   0.64184  0.10096 -4.39 0.000011

Likelihood ratio test=21.9  on 3 df, p=0.0000691
n= 400, number of events= 174

> coxph(Surv(Time,Outcome==2) ~ x1+x2+x3, data=datta) # Transplant

Call:
coxph(formula = Surv(Time, Outcome == 2) ~ x1 + x2 + x3, data = datta)

     coef exp(coef) se(coef)     z       p
x1  0.147     1.159    0.108  1.36 0.17314
x2  0.342     1.408    0.102  3.37 0.00075
x3 -0.373     0.689    0.110 -3.38 0.00072

Likelihood ratio test=20.4  on 3 df, p=0.000141
n= 400, number of events= 150


> # Stratification method
> head(datta)
  id   x1   x2   x3 Time Outcome
1  1 1.40 1.55 4.39   53       1
2  2 3.16 1.92 6.82  199       1
3  3 3.61 3.47 5.28   26       0
4  4 1.81 2.11 3.66   77       1
5  5 2.27 3.59 5.16    8       1
6  6 3.59 2.66 4.48   13       1

> head(bigdat)
  id   x1   x2   x3 Time Outcome Endpoint delta
1  1 1.40 1.55 4.39   53       1        1     1
2  1 1.40 1.55 4.39   53       1        2     0
3  2 3.16 1.92 6.82  199       1        1     1
4  2 3.16 1.92 6.82  199       1        2     0
5  3 3.61 3.47 5.28   26       0        1     0
6  3 3.61 3.47 5.28   26       0        2     0
```

```
> head(bigdat)
  id   x1   x2   x3 Time Outcome Endpoint delta
1  1 1.40 1.55 4.39   53       1        1     1
2  1 1.40 1.55 4.39   53       1        2     0
3  2 3.16 1.92 6.82  199       1        1     1
4  2 3.16 1.92 6.82  199       1        2     0
5  3 3.61 3.47 5.28   26       0        1     0
6  3 3.61 3.47 5.28   26       0        2     0

> library(coxme)

> # By default, regression coefficients are the same in each stratum.
> coxme(Surv(Time,delta) ~ x1+x2+x3 + strata(Endpoint) + (1|id) , data=bigdat)

Cox mixed-effects model fit by maximum likelihood
  Data: bigdat
  events, n = 324, 800
  Iterations= 5 26
                    NULL Integrated    Fitted
Log-likelihood -1620.737  -1603.543 -1554.832

                   Chisq    df               p   AIC     BIC
Integrated loglik  34.39  4.00 0.0000006202000 26.39   11.27
 Penalized loglik 131.81 48.56 0.0000000012847 34.69 -148.89

Model:  Surv(Time, delta) ~ x1 + x2 + x3 + strata(Endpoint) + (1 | id)
Fixed coefficients
         coef exp(coef)   se(coef)     z           p
x1  0.2703742  1.310455 0.07993475  3.38 0.000720000
x2  0.1712809  1.186824 0.07575827  2.26 0.024000000
x3 -0.4486556  0.638486 0.08180684 -5.48 0.000000042

Random effects
 Group Variable  Std Dev   Variance
 id    Intercept 0.4232343 0.1791273
```

```
>
> # Now set up product terms to make them different.
> # Note endpoint=1 means death, endpoint=2 means transplant.
>
> x1a = bigdat$x1*(bigdat$Endpoint==1); x1b = bigdat$x1*(bigdat$Endpoint==2)
> x2a = bigdat$x2*(bigdat$Endpoint==1); x2b = bigdat$x2*(bigdat$Endpoint==2)
> x3a = bigdat$x3*(bigdat$Endpoint==1); x3b = bigdat$x3*(bigdat$Endpoint==2)
>
> layers = coxme(Surv(Time,delta) ~ x1a+x2a+x3a + x1b+x2b+x3b
+                                 + strata(Endpoint) + (1|id) , data=bigdat)
> summary(layers)

Cox mixed-effects model fit by maximum likelihood
  Data: bigdat
  events, n = 324, 800
  Iterations= 5 26
                    NULL Integrated    Fitted
Log-likelihood -1620.737  -1599.337 -1550.403

                   Chisq    df                p   AIC     BIC
Integrated loglik  42.80  7.00 0.00000036453000 28.80    2.34
 Penalized loglik 140.67 51.75 0.00000000037284 37.17 -158.47

Model:  Surv(Time, delta) ~ x1a + x2a + x3a + x1b + x2b + x3b + strata(Endpoint) +
(1 | id)
Fixed coefficients
           coef exp(coef)   se(coef)     z         p
x1a  0.34395638 1.4105171 0.10429229  3.30 0.0009700
x2a  0.00873347 1.0087717 0.09886632  0.09 0.9300000
x3a -0.48608167 0.6150316 0.10657829 -4.56 0.0000051
x1b  0.18393111 1.2019330 0.11264958  1.63 0.1000000
x2b  0.36022253 1.4336484 0.10695328  3.37 0.0007600
x3b -0.41316783 0.6615512 0.11543913 -3.58 0.0003400

Random effects
 Group Variable  Std Dev   Variance
 id    Intercept 0.4243833 0.1801012

>
> # Very nice. Now test whether x3 is really stronger for death than transplant.
>
> source("http://www.utstat.toronto.edu/~brunner/Rfunctions/Wtest.txt")
> beta_hat = layers$coefficients; Vn_hat = vcov(layers)
> LL = cbind(0,0,1,0,0,-1)
> Wtest(LL,beta_hat,Vn_hat)

        W        df   p-value
0.2380867 1.0000000 0.6255915

>
> # With n=4,000 it was significant.
>
```

```
> # Sub-distribution method
>
> # install.packages("cmprsk",dependencies=TRUE) # Only need to do this once
> library(cmprsk)
>
> # Recalling 0=censored, 1=death, 2=transplant
>
> submod1 = crr(ftime=Time, fstatus=Outcome, cov1=xmat, failcode=1, cencode=0)
> summary(submod1)

Competing Risks Regression

Call:
crr(ftime = Time, fstatus = Outcome, cov1 = xmat, failcode = 1)

        coef exp(coef) se(coef)     z p-value
xmat1  0.158     1.172   0.0914  1.73   0.083
xmat2 -0.127     0.881   0.0909 -1.40   0.160
xmat3 -0.199     0.820   0.0947 -2.10   0.036

      exp(coef) exp(-coef)  2.5% 97.5%
xmat1     1.172      0.853 0.980 1.402
xmat2     0.881      1.136 0.737 1.052
xmat3     0.820      1.220 0.681 0.987

Num. cases = 400
Pseudo Log-likelihood = -957
Pseudo likelihood ratio test = 8.48  on 3 df,

>
> # For comparison again,
>
> coxph(Surv(Time,Outcome==1) ~ x1+x2+x3, data=datta) # Death
Call:
coxph(formula = Surv(Time, Outcome == 1) ~ x1 + x2 + x3, data = datta)

       coef exp(coef) se(coef)     z        p
x1  0.31403   1.36892  0.09956  3.15   0.0016
x2 -0.00942   0.99062  0.09414 -0.10   0.9203
x3 -0.44342   0.64184  0.10096 -4.39 0.000011

Likelihood ratio test=21.9  on 3 df, p=0.0000691
n= 400, number of events= 174


> # cov2 will be multiplied by a function of time.
> # Do failcode one at a time for the outcomes.

> submod1$coef
     xmat1      xmat2      xmat3
 0.1584792 -0.1270744 -0.1986958
> submod1$var
              [,1]         [,2]         [,3]
[1,]  0.008354918 -0.003602872 -0.002606424
[2,] -0.003602872  0.008256387 -0.001939874
[3,] -0.002606424 -0.001939874  0.008962973
```

Prostate cancer

```
> rm(list=ls())
> library(asaur); library(survival)
>
> # help(prostateSurvival)
> head(prostateSurvival)
  grade stage ageGroup survTime status
1  mode   T1c      80+       18      0
2  mode  T1ab    75-79       23      0
3  poor   T1c    75-79       37      0
4  mode    T2    70-74       27      0
5  mode   T1c    70-74       42      0
6  poor    T2    75-79       38      2
> summary(prostateSurvival)
  grade         stage       ageGroup       survTime          status
 mode:10988   T1ab:3881   66-69:1423   Min.   :  0.00   Min.   :0.0000
 poor: 3306   T1c :4493   70-74:2952   1st Qu.: 13.00   1st Qu.:0.0000
              T2  :5920   75-79:4313   Median : 30.00   Median :0.0000
                          80+  :5606   Mean   : 38.96   Mean   :0.5092
                                       3rd Qu.: 60.00   3rd Qu.:1.0000
                                       Max.   :119.00   Max.   :2.0000
> # Status:
>     # 1 = Death from prostate cancer
>     # 2 = Death from other causes
>     # 0 = Censored
>
> # First, simple one-at-a-time
>
> coxph(Surv(survTime,status==1) ~ grade + stage + ageGroup, data =
prostateSurvival) # D from PC
Call:
coxph(formula = Surv(survTime, status == 1) ~ grade + stage +
    ageGroup, data = prostateSurvival)

                coef exp(coef) se(coef)      z       p
gradepoor      1.4222    4.1462   0.0725 19.62 < 2e-16
stageT1c      -0.2800    0.7558   0.1018 -2.75   0.006
stageT2        0.1283    1.1369   0.0890  1.44   0.150
ageGroup70-74  0.1818    1.1993   0.2023  0.90   0.369
ageGroup75-79  0.8221    2.2754   0.1836  4.48 7.6e-06
ageGroup80+    1.2193    3.3849   0.1786  6.83 8.7e-12

Likelihood ratio test=610  on 6 df, p=0
n= 14294, number of events= 799

> coxph(Surv(survTime,status==2) ~ grade + stage + ageGroup, data =
prostateSurvival) # D from Other

Call:
coxph(formula = Surv(survTime, status == 2) ~ grade + stage +
    ageGroup, data = prostateSurvival)

                coef exp(coef) se(coef)      z       p
gradepoor      0.1873    1.2060   0.0411   4.55 5.2e-06
stageT1c      -0.4854    0.6155   0.0463 -10.48 < 2e-16
stageT2       -0.2202    0.8023   0.0415  -5.30 1.1e-07
ageGroup70-74  0.2040    1.2263   0.0839   2.43   0.015
ageGroup75-79  0.5014    1.6510   0.0785   6.38 1.7e-10
ageGroup80+    0.9928    2.6986   0.0756  13.13 < 2e-16

Likelihood ratio test=551  on 6 df, p=0
n= 14294, number of events= 3240
```

```
> # install.packages("cmprsk",dependencies=TRUE) # Only need to do this once
> library(cmprsk)
> attach(prostateSurvival)
>
> # Explanatory variables must be in a model matrix
> X = model.matrix(status ~ grade + stage + ageGroup)
> head(X)
  (Intercept) gradepoor stageT1c stageT2 ageGroup70-74 ageGroup75-79 ageGroup80+
1           1         0        1       0             0             0           1
2           1         0        0       0             0             1           0
3           1         1        1       0             0             1           0
4           1         0        0       1             1             0           0
5           1         0        1       0             1             0           0
6           1         1        0       1             0             1           0
> X = X[,-1]; head(X)
  gradepoor stageT1c stageT2 ageGroup70-74 ageGroup75-79 ageGroup80+
1         0        1       0             0             0           1
2         0        0       0             0             1           0
3         1        1       0             0             1           0
4         0        0       1             1             0           0
5         0        1       0             1             0           0
6         1        0       1             0             1           0
>
> submod = crr(ftime=survTime, fstatus=status, cov1=X, failcode=1, cencode=0)
> summary(submod)
Competing Risks Regression

Call:
crr(ftime = survTime, fstatus = status, cov1 = X, failcode = 1)

                coef exp(coef) se(coef)      z p-value
gradepoor      1.363     3.907   0.0751 18.157 0.0e+00
stageT1c      -0.137     0.872   0.1023 -1.342 1.8e-01
stageT2        0.187     1.206   0.0910  2.053 4.0e-02
ageGroup70-74  0.181     1.198   0.1981  0.913 3.6e-01
ageGroup75-79  0.765     2.148   0.1793  4.265 2.0e-05
ageGroup80+    1.037     2.820   0.1741  5.954 2.6e-09

              exp(coef) exp(-coef)  2.5% 97.5%
gradepoor         3.907      0.256 3.373  4.53
stageT1c          0.872      1.147 0.713  1.07
stageT2           1.206      0.830 1.009  1.44
ageGroup70-74     1.198      0.835 0.813  1.77
ageGroup75-79     2.148      0.466 1.512  3.05
ageGroup80+       2.820      0.355 2.005  3.97

Num. cases = 14294
Pseudo Log-likelihood = -6828
Pseudo likelihood ratio test = 519  on 6 df,

> # submod$coef; submod$var
```
