

ICPSR 2015 “Advanced Maximum Likelihood”: Survival Analysis

Day Seven

August 11, 2015

Separation

“Separation” = “perfect prediction”

Intuition:

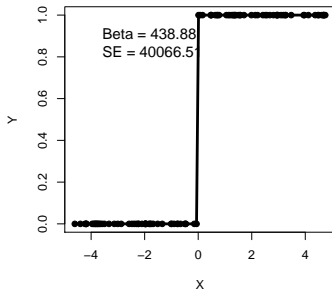
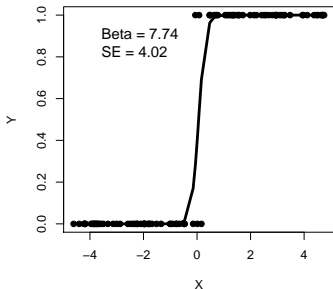
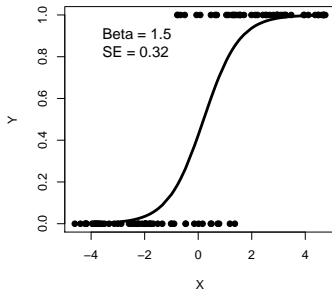
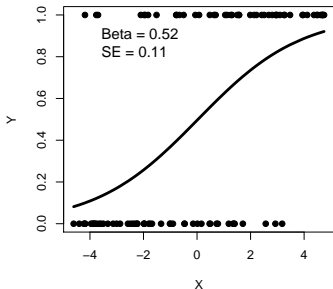
	Dems	
Yeas	0	1
0	178	34
1	0	219

$$\Pr(Y = 1|X = 0) = ?$$

Separation: Effects

- $\hat{\beta}_X = \pm\infty$
- $\widehat{\text{s.e.}}_{\beta} = \infty$
- $\left. \frac{\partial^2 \ln L}{\partial X^2} \right|_{\hat{\beta}} = 0$

Separation Illustrated



Separation: What Happens

```
> summary(glm(Y~W+Z+X,family="binomial"))
```

Call:

```
glm(formula = Y ~ W + Z + X, family = "binomial")
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.363	0.111	-3.26	0.00111	**
W	0.424	0.119	3.57	0.00036	***
Z	-0.412	0.112	-3.67	0.00024	***
X	18.746	541.835	0.03	0.97240	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 684.41 on 499 degrees of freedom
Residual deviance: 464.69 on 496 degrees of freedom
AIC: 472.7

Number of Fisher Scoring iterations: 17

Separation: What Happens (Stata Remix)

```
. logit Y W Z X
```

note: X != 0 predicts success perfectly

X dropped and 136 obs not used

Iteration 0: log likelihood = -245.53269

Iteration 1: log likelihood = -232.41173

Iteration 2: log likelihood = -232.34436

Iteration 3: log likelihood = -232.34436

Logistic regression

Number of obs = 364

LR chi2(2) = 26.38

Prob > chi2 = 0.0000

Pseudo R2 = 0.0537

Log likelihood = -232.34436

Y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
-----+-----							
W	.4242117	.1187869	3.57	0.000	.1913936	.6570298	
Z	-.4120285	.1123154	-3.67	0.000	-.6321628	-.1918943	
X	0	(omitted)					
_cons	-.3626348	.1112033	-3.26	0.001	-.5805892	-.1446803	

Solution (?): Exact Logistic Regression

- Cox (1970, Ch. 4); Hirji et al. (1987 *JASA*); Mehta & Patel (1995 *Stat. Med.*)
- Conditions on permutations of covariate patterns
- Always has finite solutions;
- Computational issues...

Firth's (1993) Correction

Firth proposed:

$$L(\beta|Y)^* = L(\beta|Y) |\mathbf{I}(\beta)|^{\frac{1}{2}}$$

$$\ln L(\beta|Y)^* = \ln L(\beta|Y) + 0.5 \ln |\mathbf{I}(\beta)|$$

“Penalized likelihood”:

- Consistent
- Eliminates small-sample bias
- Exist given separation
- Bayesians: “Jeffreys’ prior”

Potential Drawbacks

- “Profile” (= “concentrated”) likelihood
- $L(\hat{\beta})$ can be asymmetrical...
- \rightarrow inference...

- R
 - `elrm` (exact logistic regression via MCMC)
 - `brlr` (“bias-reduced logistic regression”)
 - `logistf` (“Firth’s logistic regression”)
- Stata
 - `exlogistic` (exact logistic regression)
 - `firthlogit` (Firth corrected logit)

Example: Pets as Family

- CBS/NYT Poll, April 1997
- Standard political/demographics, plus
- “Do you consider your pet to be a member of your family, or not?”
- Yes = 84.4%, No = 15.6%

Pets as Family: Data

```
> summary(Pets)
```

petfamily	female	married
Min. :0.00	Female:403	Divorced/Sep:118
1st Qu.:1.00	Male :321	Married :442
Median :1.00		NBM :118
Mean :0.84		Widowed : 46
3rd Qu.:1.00		
Max. :1.00		

partyid	education
: 58	< HS : 71
Democrat :224	College Grad:131
GOP :228	HS diploma :244
Independent:214	Post-Grad : 96
	Some college:182

Pets as Family: Basic Model

```
> Pets.1<-glm(petfamily~female+as.factor(married)+as.factor(partyid)
+as.factor(education),data=Pets,family=binomial)
```

```
> summary(Pets.1)
```

Call:

```
glm(formula = petfamily ~ female + as.factor(married) + as.factor(partyid) +
    as.factor(education), family = binomial, data = Pets)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.0133	0.5388	3.74	0.00019 ***
femaleMale	-0.6959	0.2142	-3.25	0.00116 **
as.factor(married)Married	-0.0657	0.2911	-0.23	0.82147
as.factor(married)NBM	0.4599	0.3957	1.16	0.24504
as.factor(married)Widowed	-0.1568	0.4921	-0.32	0.75007
as.factor(partyid)Democrat	-0.1241	0.4286	-0.29	0.77213
as.factor(partyid)GOP	-0.0350	0.4321	-0.08	0.93537
as.factor(partyid)Independent	-0.1521	0.4299	-0.35	0.72338
as.factor(education)College Grad	0.2511	0.4121	0.61	0.54228
as.factor(education)HS diploma	0.0595	0.3685	0.16	0.87182
as.factor(education)Post-Grad	0.1946	0.4331	0.45	0.65321
as.factor(education)Some college	0.0587	0.3867	0.15	0.87928

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 627.14 on 723 degrees of freedom
Residual deviance: 612.76 on 712 degrees of freedom
AIC: 636.8

Number of Fisher Scoring iterations: 4



Pets as Family: More Complicated Model

```
> Pets.2<-glm(petfamily~female+as.factor(married)*female+as.factor(partyid)+  
  as.factor(education),data=Pets,family=binomial)
```

```
> summary(Pets.2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.2971	0.6166	3.73	0.0002	***
femaleMale	-1.1833	0.5305	-2.23	0.0257	*
as.factor(married)Married	-0.3218	0.4470	-0.72	0.4716	
as.factor(married)NBM	0.1854	0.6140	0.30	0.7628	
as.factor(married)Widowed	-0.7415	0.5780	-1.28	0.1995	
as.factor(partyid)Democrat	-0.1575	0.4297	-0.37	0.7140	
as.factor(partyid)GOP	-0.0445	0.4334	-0.10	0.9182	
as.factor(partyid)Independent	-0.1757	0.4312	-0.41	0.6837	
as.factor(education)College Grad	0.2332	0.4137	0.56	0.5730	
as.factor(education)HS diploma	0.0558	0.3703	0.15	0.8801	
as.factor(education)Post-Grad	0.2171	0.4342	0.50	0.6171	
as.factor(education)Some college	0.0358	0.3890	0.09	0.9266	
femaleMale:as.factor(married)Married	0.4853	0.5908	0.82	0.4114	
femaleMale:as.factor(married)NBM	0.5260	0.8051	0.65	0.5136	
femaleMale:as.factor(married)Widowed	15.2516	549.3719	0.03	0.9779	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 627.14 on 723 degrees of freedom
Residual deviance: 607.42 on 709 degrees of freedom
AIC: 637.4

Number of Fisher Scoring iterations: 14

What's Going On?

```
> with(Pets, xtabs(~petfamily+as.factor(married)+female))
```

```
, , female = Female
```

```
      as.factor(married)
petfamily Divorced/Sep Married NBM Widowed
      0           7      28    5      7
      1          67     199   58     32
```

```
, , female = Male
```

```
      as.factor(married)
petfamily Divorced/Sep Married NBM Widowed
      0           11      47    8      0
      1          33     168   47      7
```


Pets as Family: Firth Model

```
> Pets.Firth<-logistf(petfamily~female+as.factor(married)*female+as.factor(partyid)+  
  as.factor(education),data=Pets)
```

```
> Pets.Firth
```

Model fitted by Penalized ML

Confidence intervals and p-values by Profile Likelihood

	coef	se(coef)	lower 0.95	upper 0.95	Chisq	p
(Intercept)	2.1589	0.60	1.05	3.40	16.17636	0.000058
femaleMale	-1.1387	0.52	-2.19	-0.14	5.04186	0.024742
as.factor(married)Married	-0.2739	0.43	-1.19	0.53	0.41518	0.519353
as.factor(married)NBM	0.1589	0.59	-0.99	1.37	0.07322	0.786705
as.factor(married)Widowed	-0.7263	0.56	-1.84	0.38	1.67233	0.195947
as.factor(partyid)Democrat	-0.1182	0.42	-0.99	0.66	0.08159	0.775159
as.factor(partyid)GOP	-0.0078	0.42	-0.89	0.78	0.00034	0.985289
as.factor(partyid)Independent	-0.1364	0.42	-1.01	0.65	0.10813	0.742278
as.factor(education)College Grad	0.2390	0.40	-0.57	1.02	0.34480	0.557069
as.factor(education)HS diploma	0.0753	0.36	-0.67	0.76	0.04289	0.835933
as.factor(education)Post-Grad	0.2184	0.43	-0.63	1.05	0.26307	0.608019
as.factor(education)Some college	0.0524	0.38	-0.72	0.78	0.01888	0.890698
femaleMale:as.factor(married)Married	0.4558	0.58	-0.66	1.61	0.63550	0.425347
femaleMale:as.factor(married)NBM	0.5233	0.78	-1.02	2.05	0.45133	0.501702
femaleMale:as.factor(married)Widowed	2.4017	1.68	-0.14	7.37	3.37453	0.066212

Likelihood ratio test=17 on 14 df, p=0.24, n=724

- Separation \nrightarrow dropping covariates!
- Firth's approach $>$ ELR
- Can also be applied to other sparse-data situations...

From the coxph Documentation

Convergence

In certain data cases the actual MLE estimate of a coefficient is infinity, e.g., a dichotomous variable where one of the groups has no events. When this happens the associated coefficient grows at a steady pace and a race condition will exist in the fitting routine: either the log likelihood converges, the information matrix becomes effectively singular, an argument to exp becomes too large for the computer hardware, or the maximum number of interactions is exceeded. (Nearly always the first occurs.) The routine attempts to detect when this has happened, not always successfully. The primary consequence for the user is that the Wald statistic = coefficient/se(coefficient) is not valid in this case and should be ignored; the likelihood ratio and score tests remain valid however.

Separation in Survival Data

```
> set.seed(7222009)
> X<-rep(0:1,times=100)
> T<-abs(rweibull(200,shape=1.2,
                  scale=1/(exp(0+0.2*X))))
> C<-rbinom(200,1,0.2)
> C<-ifelse(X==0,0,C)
```

```
> table(C,X)
```

	X	
C	0	1
0	100	81
1	0	19

Cox Results

```
> cox.fit<-coxph(Surv(T,C)~X,method="efron")
```

Warning message:

In fitter(X, Y, strats, offset, init, control, weights = weights, :
Loglik converged before variable 1 ; beta may be infinite.

```
> summary(cox.fit)
```

Call:

```
coxph(formula = Surv(T, C) ~ X, method = "efron")
```

n= 200, number of events= 19

	coef	exp(coef)	se(coef)	z	Pr(> z)
X	2.038e+01	7.112e+08	5.630e+03	0.004	0.997

	exp(coef)	exp(-coef)	lower	.95	upper	.95
X	711225014	1.406e-09		0		Inf

Concordance= 0.761 (se = 0.064)

Rsquare= 0.137 (max possible= 0.583)

Likelihood ratio test= 29.37 on 1 df, p=5.994e-08

Wald test = 0 on 1 df, p=0.9971

Score (logrank) test = 22.11 on 1 df, p=2.58e-06

Parametric Model = No Help

```
> weib.fit<-survreg(Surv(T,C)~X,dist="weibull")  
> summary(weib.fit)
```

Call:

```
survreg(formula = Surv(T, C) ~ X, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	61.5536	0.337	182.851	0.000
X	-60.1634	0.000	-Inf	0.000
Log(scale)	-0.0184	0.190	-0.097	0.923

Scale= 0.982

Weibull distribution

Loglik(model)= -45.9 Loglik(intercept only)= -60.8

Chisq= 29.94 on 1 degrees of freedom, p= 4.5e-08

Number of Newton-Raphson Iterations: 9

n= 200

Heinze and Schemper (2001):

$$\ln PL(\beta|Y)^* = \ln PL(\beta|Y) + 0.5 \ln |\mathbf{I}(\beta)|$$

with $\mathbf{I}(\beta) = -E \left[\frac{\partial^2}{\partial \theta^2} \ln PL(X, \beta) \middle| \beta \right]$.

Also, software: `coxphf`...

Firth-Corrected Cox

```
> SIM<-cbind(T,C,X)
> SIM<-data.frame(SIM)
> firth.fit<-coxphf(SIM,formula=Surv(T,C)~X)
```

```
> firth.fit
coxphf(formula = Surv(T, C) ~ X, data = SIM)
Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood
```

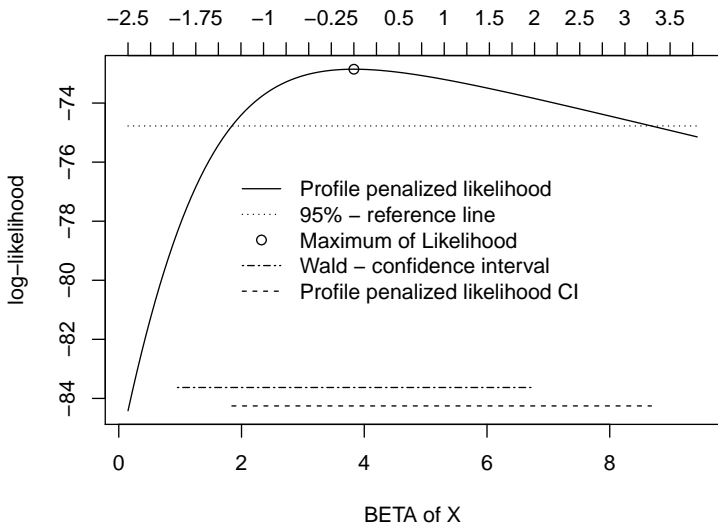
	coef	se(coef)	exp(coef)	lower 0.95	upper 0.95	Chisq	p
X	3.830922	1.472103	46.10502	6.316689	5871.037	26.08802	3.262011e-07

```
Likelihood ratio test=26.08802 on 1 df, p=3.262011e-07, n=200
```


Examining the Profile Likelihood

Profile likelihood

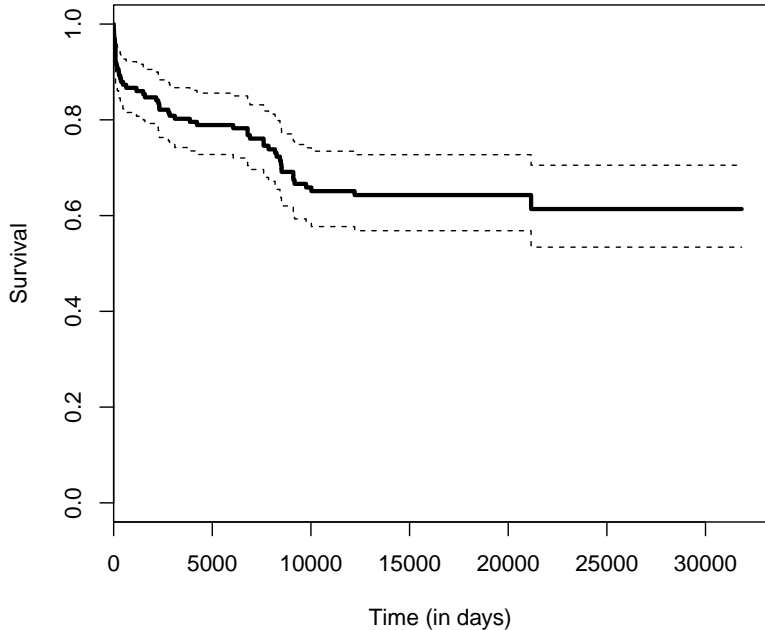
distance from maximum in standard deviations



Example: Lo et al. (2008)

- Outcome: “Cease-Fire Duration”
- Key covariate: *Foreign-Imposed Regime Change* (“FIRC”)
- Data: Annual data on cease-fires, 1914-2001 (expanding Fortna 1998)
- Hypotheses:
 - FIRCs → more durable cease-fires
 - Pacifying influence of FIRCs declines over time

Lo et al.: Kaplan-Meier



From Lo et al. (2008)

<i>Variables</i>	<i>Model 1</i> <i>(ARCHIGOS data)</i>	<i>Model 2</i> <i>(ARCHIGOS data)</i>
<hr/>		
FOREIGN-IMPOSED REGIME CHANGE	-161*** (29.3)	—
FIRC* $\ln(t)$	16.8*** (3.03)	—
PUPPET-FIRC	—	-161*** (29.4)
PUPPET-FIRC* $\ln(t)$	—	16.8*** (3.04)
CHANGE IN CAPABILITIES	.272 (.376)	.274 (.376)
BATTLE CONSISTENCY	-.796** (.336)	-.809** (.342)

Simplified Cox Model

```
> LHR.Cox<-coxph(LHR.S~archigosFIRC+archigosFIRClnt,data=LHR,method="efron",  
                 iter.max=10000)
```

Warning message:

```
In fitter(X, Y, strats, offset, init, control, weights = weights,  :  
  Ran out of iterations and did not converge
```

```
> LHR.Cox
```

Call:

```
coxph(formula = LHR.S ~ archigosFIRC + archigosFIRClnt, data = LHR,  
      method = "efron", iter.max = 10000)
```

	coef	exp(coef)	se(coef)	z	p
archigosFIRC	-44.11	6.99e-20	20.92	-2.11	0.035
archigosFIRClnt	4.69	1.09e+02	2.24	2.09	0.036

```
Likelihood ratio test=20.8 on 2 df, p=3.04e-05 n= 6368, number of events= 54
```

Firth-Corrected Cox Model

```
> LHR.CoxF<-coxphf(LHR.S~archigosFIRC+archigosFIRCInt,data=LHR,maxit=1000)
```

```
> LHR.CoxF
```

```
coxphf(formula = LHR.S ~ archigosFIRC + archigosFIRCInt, data = LHR,  
        maxit = 1000)
```

Model fitted by Penalized ML

Confidence intervals and p-values by Profile Likelihood

	coef	se(coef)	exp(coef)	lower	0.95
archigosFIRC	-55.591223	26.330771	7.19513e-25	4.808432e-65	
archigosFIRCInt	5.848163	2.775927	3.46597e+02		NaN

	upper	0.95	Chisq	p
archigosFIRC		NaN	9.738246	0.001804731
archigosFIRCInt	7320112	8.647141	0.003275750	

Likelihood ratio test=15.09131 on 2 df, p=0.0005284014, n=6368

What Is Going On?

```
> table(LHR$archigosFIRC,LHR$X_d)
```

	0	1
0	5265	52
1	1049	2

Days \rightarrow Years = Little Help

	Cox	Firth-Corrected
FIRC	-53.83 (51.30)	-21.08 (9.99)
FIRC $\times \ln(T)$	14.57 (13.54)	5.85 (2.78)
AIC	494.07	496
Num. events	54	54