PLSC 504 – Fall 2017
Binary Response Models II

August 31, 2017

- Models for “separation”
- Models for rare events
- Common Focus: Shortage of information on Y

"Separation" = "perfect prediction"

Intuition:

```
      Dems
Yeas    0   1
   0 178  34
   1   0 219
```

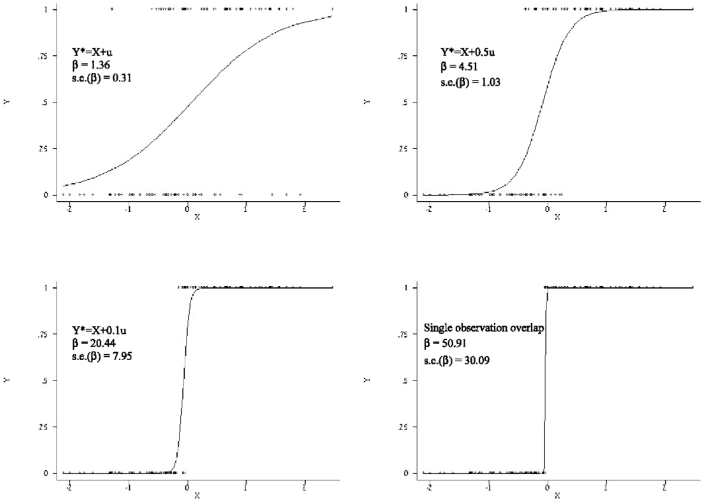
$\Pr(Y = 1 | X = 1) = ?$

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

Figure 1: Actual and Predicted Values, Simulated Logistic Regressions



Separation: What Happens

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

Call:
glm(formula = Y ~ X + W + Z, family = binomial)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.2566     0.1475   -1.74    0.082 .
X            18.3591   744.8791    0.02    0.980
W            -0.4151     0.0511   -8.12  4.7e-16 ***
Z             0.0613     0.0312    1.96    0.050 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 691.10  on 499  degrees of freedom
Residual deviance: 291.67  on 496  degrees of freedom
AIC: 299.7
```

Separation: What Happens (Stata Remix)

```
. logit Y X W Z

note: X != 0 predicts success perfectly
      X dropped and 150 obs not used

Logistic regression                               Number of obs   =        350
                                                  LR chi2(2)      =     152.96
                                                  Prob > chi2     =     0.0000
Log likelihood = -145.83444                       Pseudo R2       =     0.3440
------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           W |  -.4150914   .0511269    -8.12   0.000    -.5152983   -.3148844
           Z |   .0612667   .0312303     1.96   0.050     .0000564     .122477
       _cons |  -.2565799   .1475397    -1.74   0.082    -.5457524    .0325926
------------------------------------------------------------------------------
```

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:

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

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

"Penalized likelihood":

- Consistent
- Eliminates small-sample bias
- Exist given separation
- Bayesians: "Jeffreys' prior"

- "Profile" (= "concentrated") likelihood
- $\hat{\beta}$ 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)

- 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 $= 15.6\%$

```
> summary(Pets)

   petfamily         female                married               partyid               education
 Min.   :0.000   Min.   :0.000   Married     :442   Democrat   :225   < HS          : 71
 1st Qu.:1.000   1st Qu.:0.000   Widowed     : 46   Independent:214   HS diploma   :244
 Median :1.000   Median :1.000   Divorced/Sep:118   GOP        :229   Some college:184
 Mean   :0.844   Mean   :0.556   NBM         :118   NA's       : 58   College Grad:131
 3rd Qu.:1.000   3rd Qu.:1.000   NA's        :  2                     Post-Grad   : 96
 Max.   :1.000   Max.   :1.000
```

Pets as Family: Basic Model

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

Call:
glm(formula = petfamily ~ female + as.factor(married) + as.factor(partyid) +
    as.factor(education), family = binomial)

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                       1.24155    0.39197    3.17   0.0015 **
female                            0.70596    0.22299    3.17   0.0015 **
as.factor(married)Widowed        -0.16901    0.44703   -0.38   0.7054
as.factor(married)Divorced/Sep   -0.00429    0.30159   -0.01   0.9887
as.factor(married)NBM             0.59800    0.35114    1.70   0.0886 .
as.factor(partyid)Independent    -0.02316    0.26549   -0.09   0.9305
as.factor(partyid)GOP             0.09548    0.26973    0.35   0.7234
as.factor(education)HS diploma   -0.13122    0.39564   -0.33   0.7401
as.factor(education)Some college  0.01528    0.41950    0.04   0.9709
as.factor(education)College Grad  0.09920    0.43514    0.23   0.8197
as.factor(education)Post-Grad     0.11758    0.46015    0.26   0.7983
```

Pets as Family: More Complicated Model

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

Coefficients:
                                     Estimate Std. Error z value Pr(>|z|)
(Intercept)                           1.22666    0.39791    3.08   0.0021 **
female                                0.73871    0.27406    2.70   0.0070 **
as.factor(married)Widowed            14.35019  549.43541    0.03   0.9792
as.factor(married)Divorced/Sep       -0.17577    0.40600   -0.43   0.6651
as.factor(married)NBM                 0.61040    0.44798    1.36   0.1730
as.factor(partyid)Independent        -0.01163    0.26669   -0.04   0.9652
as.factor(partyid)GOP                 0.12111    0.27084    0.45   0.6548
as.factor(education)HS diploma       -0.14020    0.39800   -0.35   0.7246
as.factor(education)Some college     -0.00973    0.42218   -0.02   0.9816
as.factor(education)College Grad      0.07601    0.43697    0.17   0.8619
as.factor(education)Post-Grad         0.13015    0.46150    0.28   0.7779
female:as.factor(married)Widowed    -14.89255  549.43562   -0.03   0.9784
female:as.factor(married)Divorced/Sep 0.35197    0.60823    0.58   0.5628
female:as.factor(married)NBM         -0.01729    0.71719   -0.02   0.9808
```

What's Going On?

```
> xtabs(~petfamily+as.factor(married)+female)
, , female = 0

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

, , female = 1

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

The Stata Version

```
. xi: logit petfamily i.married*female i.partyid i.education
i.married         _Imarried_1-4       (naturally coded; _Imarried_1 omitted)
i.marr~d*female   _ImarXfemal_#       (coded as above)
i.partyid         _Ipartyid_0-2       (naturally coded; _Ipartyid_0 omitted)
i.education       _Ieducation_1-5     (naturally coded; _Ieducation_1 omitted)

Iteration 0:   log likelihood = -290.21966
Iteration 1:   log likelihood = -281.02869
Iteration 2:   log likelihood = -280.47066
Iteration 3:   log likelihood = -280.38132
Iteration 4:   log likelihood = -280.34954
.
.
.
```

```
.
.
.
Iteration 14:  log likelihood = -280.33119
Iteration 15:  log likelihood = -280.33119
Iteration 16:  log likelihood = -280.33118
Iteration 17:  log likelihood = -280.33118

Logistic regression                               Number of obs   =        666
                                                  LR chi2(13)     =      19.78
                                                  Prob > chi2     =     0.1009
Log likelihood = -280.33118                       Pseudo R2       =     0.0341
------------------------------------------------------------------------------
   petfamily |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _Imarried_2 |   17.73324   .4741042    37.40   0.000     16.80401    18.66247
 _Imarried_3 |  -.1757744   .4060014    -0.43   0.665    -.9715225    .6199736
 _Imarried_4 |   .6104046     .44798     1.36   0.173      -.26762    1.488429
      female |    .738709   .2740554     2.70   0.007     .2015704    1.275848
_ImarXfema~2 |   -18.2756          .        .       .            .           .
_ImarXfema~3 |   .3519719   .6082335     0.58   0.563    -.8401439    1.544088
_ImarXfema~4 |  -.0172932   .7171902    -0.02   0.981     -1.42296    1.388374
 _Ipartyid_1 |  -.0116295    .266687    -0.04   0.965    -.5343263    .5110673
 _Ipartyid_2 |   .1211075   .2708435     0.45   0.655     -.409736     .651951
_Ieducatio~2 |  -.1402004   .3980019    -0.35   0.725    -.9202698     .639869
_Ieducatio~3 |   -.009729   .4221759    -0.02   0.982    -.8371785    .8177205
_Ieducatio~4 |   .0760091   .4369737     0.17   0.862    -.7804437    .9324618
_Ieducatio~5 |   .1301519   .4615026     0.28   0.778    -.7743767     1.03468
       _cons |   1.226664    .397911     3.08   0.002     .4467729    2.006555
------------------------------------------------------------------------------
Note: 0 failures and 7 successes completely determined.
```

Pets as Family: Firth Model

```
> pets.firth<-logistf(petfamily~female+as.factor(married)*female+as.factor(partyid)+
  as.factor(education))

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)                         1.18226   0.3891     0.4542     1.9885 1.059e+01 0.0011
female                              0.72204   0.2706     0.1988     1.2642 7.369e+00 0.0066
as.factor(married)Widowed           1.49678   1.5587    -0.6443     6.3705 1.587e+00 0.2078
as.factor(married)Divorced/Sep     -0.19676   0.4016    -0.9493     0.6183 2.413e-01 0.6233
as.factor(married)NBM               0.55597   0.4372    -0.2486     1.4772 1.775e+00 0.1827
as.factor(partyid)Independent      -0.01161   0.2616    -0.5280     0.5047 1.953e-03 0.9647
as.factor(partyid)GOP               0.11840   0.2659    -0.4061     0.6440 1.966e-01 0.6575
as.factor(education)HS diploma     -0.11097   0.3885    -0.9137     0.6215 8.273e-02 0.7736
as.factor(education)Some college    0.01203   0.4117    -0.8302     0.7986 8.515e-04 0.9767
as.factor(education)College Grad    0.09122   0.4265    -0.7743     0.9124 4.546e-02 0.8312
as.factor(education)Post-Grad       0.13755   0.4511    -0.7662     1.0154 9.283e-02 0.7606
female:as.factor(married)Widowed   -2.06760   1.6253    -6.9923     0.3025 2.782e+00 0.0953
female:as.factor(married)Divorced/Sep 0.32906 0.5956    -0.8251     1.5156 3.109e-01 0.5772
female:as.factor(married)NBM       -0.05690   0.6883    -1.3841     1.3582 6.837e-03 0.9341

Likelihood ratio test=17.29 on 13 df, p=0.1862, n=666
```

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

- Collect lots of "0s" for a few "1s"
- Classification bias...

Suppose

$$\Pr(Y_i) = \Lambda(0 + 1X_i)$$

Then

$$E(\hat{\beta}_0 - \beta_0) \approx \frac{\bar{\pi} - 0.5}{N\bar{\pi}(1 - \bar{\pi})}$$

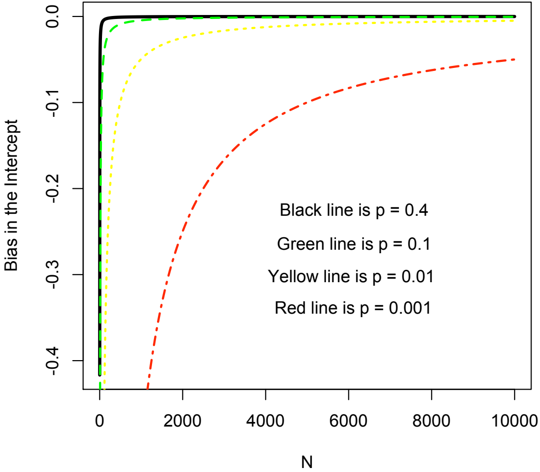
where $\bar{\pi} = \overline{\Pr(Y = 1)}$ is < 0.5.

Bias is:

- always negative,
- worse as $\bar{\pi} \to 0$ (for fixed N),
- disappearing as $N \to \infty$.

Logit/probit are "best" around $\bar{\pi} = 0.5$.

Rare Event Bias, Illustrated



- Calculate $\tau = \frac{N_1 s}{N}$
- Collect data on all "1s"
- Sample from the "0s"
- Estimate a logit*
- *Correct* the estimates ex post...

Sampling...

- $\tau =$ fraction of "1s" in the population
- $\bar{Y} =$ fraction of '1s" in the sample
- K&Z suggest $\bar{Y} \in [0.2, 0.5]$

Weighting...

$$w_1 = \frac{\tau}{\bar{Y}} \text{ (weights for "1s")}$$

$$w_0 = \frac{1-\tau}{1-\bar{Y}} \text{ (weights for "0s")}$$

$$\ln L(\beta|Y) = \sum_{i=1}^{N} w_1 Y_i \ln \Lambda(\mathbf{X}_i\boldsymbol{\beta}) + w_0(1 - Y_i) \ln[1 - \Lambda(\mathbf{X}_i\boldsymbol{\beta})]$$

- Good under (possible) misspecification, but
- Not as efficient as "prior correction," and
- Gets s.e.s wrong...

$$\hat{\beta}_{0\mathsf{pc}} = \hat{\beta}_0 - \ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{Y}}{1-\bar{Y}}\right)\right]$$

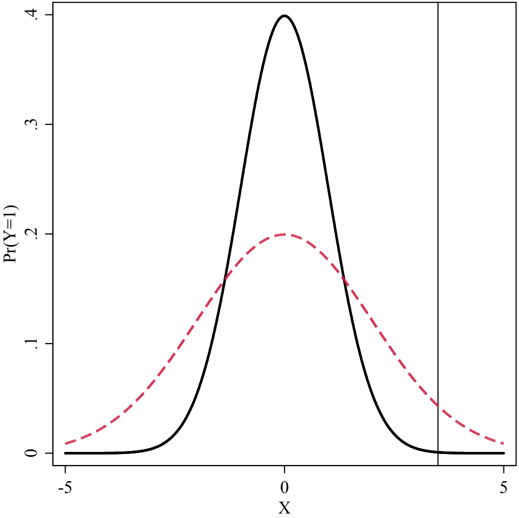
$$\mathrm{bias}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\xi$$

where $\xi = f[w_i, \hat{\pi}_i, \mathbf{X}]$.

Correction is

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \mathrm{bias}(\hat{\boldsymbol{\beta}})$$

- Bias correction introduces additional variability...
- Ignoring it yields underpredictions (again).


.4
.3
.2
.1
0
Pr(Y=1)
-5
0
5
X

Use:

$$\Pr(Y_i = 1) \approx \tilde{\pi}_i + C_i$$

where

$$C_i = (0.5 - \tilde{\pi}_i)\tilde{\pi}_i(1 - \tilde{\pi}_i)\mathbf{X}_i\mathbf{V}(\tilde{\beta})\mathbf{X}_i'$$

- Oneal and Russett 1997; also Beck/Katz/Tucker (1998) etc.
- International disputes
- Politically-relevant dyad-years, 1950-1985
- NT=20448, 405 dyad-years of disputes.

```
> baselogit<-glm(dispute~dembkt+grobkt+allies+contig+capbkt+trade,data=RE,family=binomial)
> summary(baselogit)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.32668    0.11451 -37.785  < 2e-16 ***
dembkt       -0.40120    0.10063  -3.987 6.70e-05 ***
grobkt       -3.42753    1.25181  -2.738  0.00618 **
allies       -0.47969    0.11275  -4.255 2.09e-05 ***
contig        1.35358    0.12091  11.195  < 2e-16 ***
capbkt       -0.19620    0.05011  -3.916 9.01e-05 ***
trade       -21.07611   11.30396  -1.864  0.06225 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    Null deviance: 3978.5  on 20447  degrees of freedom
Residual deviance: 3693.8  on 20441  degrees of freedom
AIC: 3707.8

Number of Fisher Scoring iterations: 9
```

Faking It: Case-Control Sampling

```
> REones<-RE[dispute==1,]
> REzeros<-RE[dispute==0,]
> RSzeros<-REzeros[sample(1:nrow(REzeros),1000,replace=FALSE),]
> REsample<-data.frame(rbind(REones,RSzeros))
> table(REsample$dispute)

   0    1
1000  405

> sample.logit<-glm(dispute~dembkt+grobkt+allies+contig+capbkt+trade,data=REsample,
  family=binomial)
> summary(sample.logit)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.33358    0.13074 -10.200  < 2e-16 ***
dembkt       -0.48623    0.12000  -4.052 5.08e-05 ***
grobkt       -2.96784    1.58954  -1.867   0.0619 .
allies       -0.34848    0.14312  -2.435   0.0149 *
contig        1.20978    0.14520   8.332  < 2e-16 ***
capbkt       -0.22708    0.05544  -4.096 4.21e-05 ***
trade       -16.70921   11.92878  -1.401   0.1613
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    Null deviance: 1687.6  on 1404  degrees of freedom
Residual deviance: 1486.5  on 1398  degrees of freedom
AIC: 1500.5
```

Rare Events Logit, Prior Correction

```
> relogit.pc<-zelig(dispute~dembkt+grobkt+allies+contig+capbkt+trade,
  data=REsample,model="relogit",tau=405/20448,case.control=c("prior"))
> summary(relogit.pc)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.32990    0.13074 -33.119  < 2e-16 ***
dembkt       -0.48217    0.12000  -4.018 5.87e-05 ***
grobkt       -2.99301    1.58954  -1.883   0.0597 .
allies       -0.34345    0.14312  -2.400   0.0164 *
contig        1.20419    0.14520   8.294  < 2e-16 ***
capbkt       -0.21719    0.05544  -3.917 8.96e-05 ***
trade       -14.15039   11.92878  -1.186   0.2355
---
    Null deviance: 1687.6  on 1404  degrees of freedom
Residual deviance: 1486.5  on 1398  degrees of freedom
AIC: 1500.5

Number of Fisher Scoring iterations: 6

Prior correction performed with tau = 0.01980634
Rare events bias correction performed
```

Rare Events Logit, Weighting Correction

```
> relogit.wc<-zelig(dispute~dembkt+grobkt+allies+contig+capbkt+trade,
  data=REsample,model="relogit",tau=405/20448,case.control=c("weighting"))
> summary(relogit.wc)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.22631    0.33221 -12.722  < 2e-16 ***
dembkt      -0.39572    0.17588  -2.250   0.0244 *
grobkt      -4.83260    2.04014  -2.369   0.0178 *
allies      -0.39616    0.30084  -1.317   0.1879
contig       1.22986    0.29886   4.115 3.87e-05 ***
capbkt      -0.08992    0.15045  -0.598   0.5501
trade       20.49682   15.94035   1.286   0.1985
---

    Null deviance: 273.37  on 1404  degrees of freedom
Residual deviance: 253.61  on 1398  degrees of freedom
AIC: 53.664

Number of Fisher Scoring iterations: 9

Weighting performed with tau = 0.01980634
Rare events bias correction performed
Robust standard errors computed using vcovHAC
```

A Warning...

From the R documentation:

> <u>*Differences with Stata Version*</u>
>
> *"The Stata version of ReLogit and the R implementation differ slightly in their coefficient estimates due to differences in the matrix inversion routines implemented in R and Stata. Zelig uses orthogonal-triangular decomposition (through lm.influence()) to compute the bias term, which is more numerically stable than standard matrix calculations."*

Summary

- `relogit` in Stata plays well with $\mathcal{C}$larify
- `Zelig` implements $\mathcal{C}$larify-like functionality
- Key: be able to conduct C-C sampling *in advance*
- In practice: well...