

Complete separation, regularization, etc.

Ben Bolker

October 22, 2018



Licensed under the Creative Commons attribution-noncommercial license (<http://creativecommons.org/licenses/by-nc/3.0/>). Please share & remix noncommercially, mentioning its origin.

Complete separation

Some linear combination of predictor variables perfectly separates

- glm goes as far as it can, stops — may or may not warn you!

```
x <- runif(25)
y <- ifelse(x<0.5,0,1)
d <- data.frame(x,y)
g0 <- glm(y~x,family=binomial,data=data.frame(x,y))

## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0
or 1 occurred

coef(g0)

## (Intercept)          x
## -1102.494    2232.735
```

- easy to diagnose for small/low-dimensional data sets, harder for high-dimensional data sets
- GLMMs partly take care of this (by handling completely separated blocks)
- *support vector machines*
- Firth algorithm: bias-reduced logistic regression; modify score function; equivalent to imposing *Jeffreys prior* on the data (`logistf` package). Prior on p is $\text{Beta}(1/2, 1/2)$

```
library(logistf)
g1 <- logistf(y~x,family=binomial,data=d)
coef(g1)

## (Intercept)          x
## -11.87273    23.55905
```

```

confint(g1)

##           Lower 95%  Upper 95%
## (Intercept) -56.993068 -4.506823
## x           9.260545 115.624696

```

```

library(brglm)
g3 <- brglm(y~x, family=binomial, data=d)
coef(g3)

```

```

## (Intercept)      x
## -11.87259      23.55878

```

```

confint(g3)

```

```

## Warning: glm.fit: algorithm did not converge

```

```

## Warning: glm.fit: fitted probabilities numerically 0
or 1 occurred

```

```

## Profiling the ordinary deviance for the corresponding ML fit...

```

```

## Profiling the penalized deviance for the supplied fit...

```

```

## Calculating confidence intervals for the ML fit using deviance profiles...

```

```

## Calculating confidence intervals for the BR fit using penalized likelihood profiles...

```

```

##           2.5 %   97.5 %

```

```

## (Intercept)   -Inf -5.288122

```

```

## x           10.59565      Inf

```

- `bayesglm` function, from the `arm` package; t -distributed priors, default is `df=1` (Cauchy)

```

library(arm)

```

```

## Loading required package: MASS

```

```

## Loading required package: Matrix

```

```

## Loading required package: lme4

```

```

##

```

```

## arm (Version 1.10-1, built: 2018-4-12)

```

```

## Working directory is /media/sf_Documents/classes/stat4c03/notes

```

```

g2 <- bayesglm(y~x, family=binomial, data=d)

```

```

coef(g2)

```

```

## (Intercept)      x

```

```

## -14.09758      28.08939

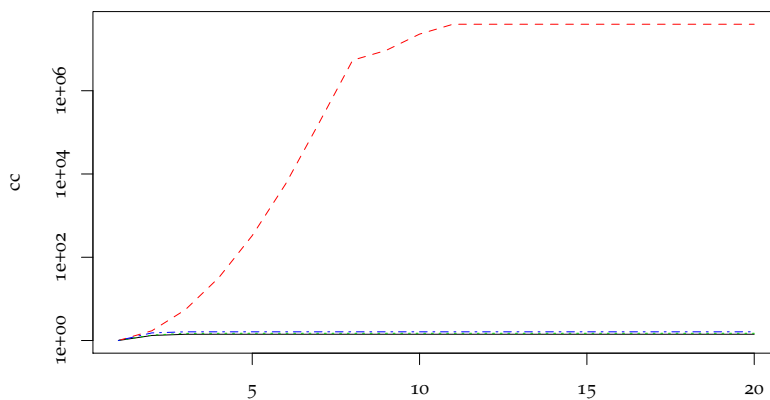
```

We used this in (Pasch et al., 2013); we initially used a GLMM but the variances kept coming out to zero so we decided to fit a bias-reduced logistic instead.

```
library("brglm2")
data("endometrial", package = "brglm2")
modML <- glm(HG ~ NV + PI + EH,
             family = binomial("probit"), data = endometrial)
update(modML, method="detect_separation")

## Separation: TRUE
## Existence of maximum likelihood estimates
## (Intercept)      NV      PI      EH
##           0      Inf      0      0
## 0: finite value, Inf: infinity, -Inf: -infinity

cc <- check_infinite_estimates(modML)
matplot(cc, log="y", type="l")
```



```
update(modML, method="brglmFit")

##
## Call:  glm(formula = HG ~ NV + PI + EH, family = binomial("probit"),
##         data = endometrial, method = "brglmFit")
##
## Coefficients:
## (Intercept)      NV      PI      EH
##    1.9146    1.6589   -0.0152   -1.3799
##
## Degrees of Freedom: 78 Total (i.e. Null);  75 Residual
```

```
## Null Deviance:      104.9  
## Residual Deviance: 57.59 AIC: 65.59
```

References

Pasch, B., B. M. Bolker, and S. M. Phelps (2013, November). Interspecific dominance via vocal interactions mediates altitudinal zonation in neotropical singing mice. *The American Naturalist* 182(5), E161–E173.